Title Here

Andrei Cuceu

Supervisor: Dr. Andrew Pontzen

University College London

March 23, 2018

Abstract

Eisenstein et al. 2007

Contents

Chapter 1

Introduction

1.1 Cosmological Context

Introduce the context of Modern Cosmology

1.2 The Cosmic Microwave Background

Introduce the CMB and talk about the starting point. E.g. very smooth initial field with some anisotropies that will be locked in the matter distribution.

Introduce BAOs.

1.3 Large Scale Structure and Galaxy Surveys

Introduce the Structure of the Universe today and the tools used to study it.

Talk about the detection of the BAO in the galaxy distribution and its smearing due to collapse.

1.4 The Missing Link (Reconstruction)

Motivate our desire to link the two and talk about the problems we have (Dark Ages)

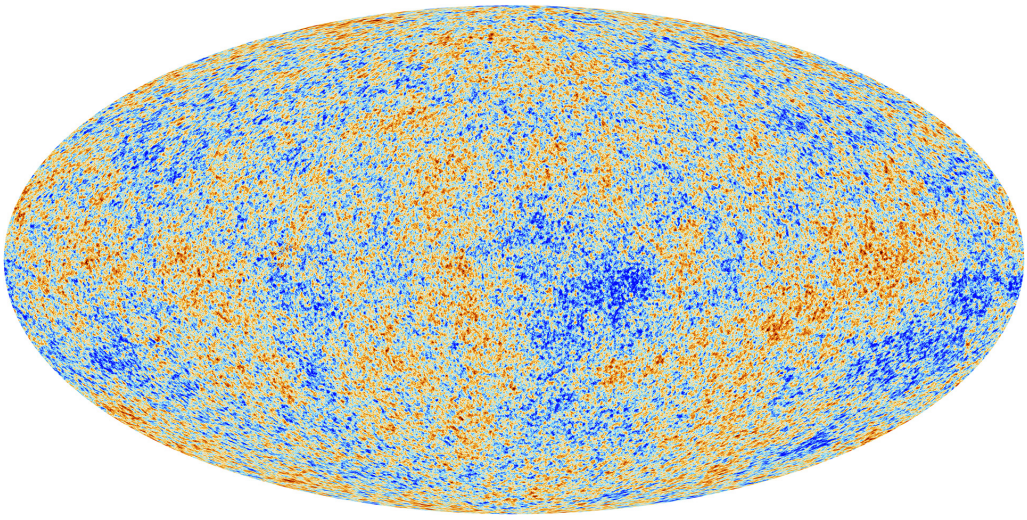

Figure 1.1: Map of the Cosmic Microwave Background aquired by the Planck Space Telescope (ref).

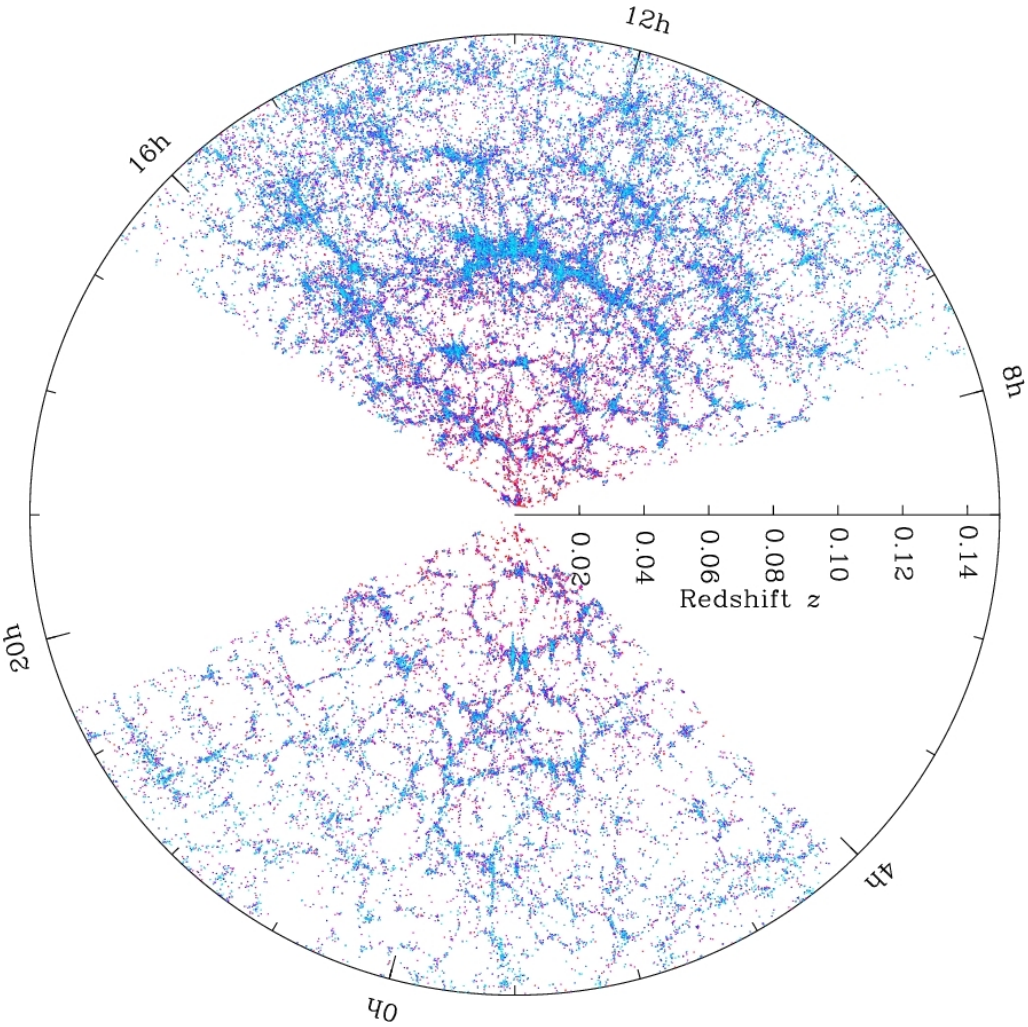

Figure 1.2: Galaxy map from the Sloan Digital Sky Survey.

Motivate the desire to reconstruct the BAO feature

Chapter 2

The Growth of Structure

Introduce cosmic variance somewhere

2.1 Perturbation Theory

Give a brief introduction to the use of Perturbation theory to study the evolution of structure. Present advatages and shortcomings

This could be merged with the next section.

2.2 Linear vs Non-Linear Collapse

Talk about the Linear regime of collapse versus the non-linear regime. Present the difficulty of constructing analytical models of non-linear collapse. Motivate our use of simulations as well as our desire to get back to the linear regime for reconstruction.

Add images of the velocity field here

2.3 The Zeldovich Approximation

Introduce the theory of the Zeldovich Approximation and motivate its use (+ background).

2.4 Reconstruction (BAO)

This can be understood by thinking about the difference between looking at the large scales versus the small scales. For example, a very large (by scale) primordial underdensity will probably still be an underdensity at late times (as a huge void). However, on small scales, information tends to be destroyed. A small scale primordial underdensity might be caught in the larger bulk flows and collapse in a larger halo. In this case, the information about the small underdensity is lost.

Finally link everything with an overview of Reconstruction techniques and how our work fits into the modern context.

this paragraph could go in chapter 2

Showcase the BAO reconstruction.

Chapter 3

The Perfect Reconstruction

The first step in understanding the evolution of the universe is to look at the theoretical limits we encounter when trying to reconstruct it. 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 primordial positions of all particles (which obviously is not available to observers). However, due to the reasons outlined above, not even this perfect reconstruction succeeds in completely recovering the primordial matter distribution.

3.1 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, Slosar, et al. 2016 (also re-

ferred 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.

I should probably give a bit more details about the simulations

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

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 we want the particles to carry the density field when we move them. After that, all the particles were moved to their starting positions (taken from the initial snapshot at $z = 99$).

3.2 From Images to Statistics

3.2.1 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

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

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

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, we see the effect that we are after. The reconstruction spreads out the matter from the collapsed filaments onto a more uniform field. 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. The total mass of the simulation should be conserved.

3.2.2 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

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

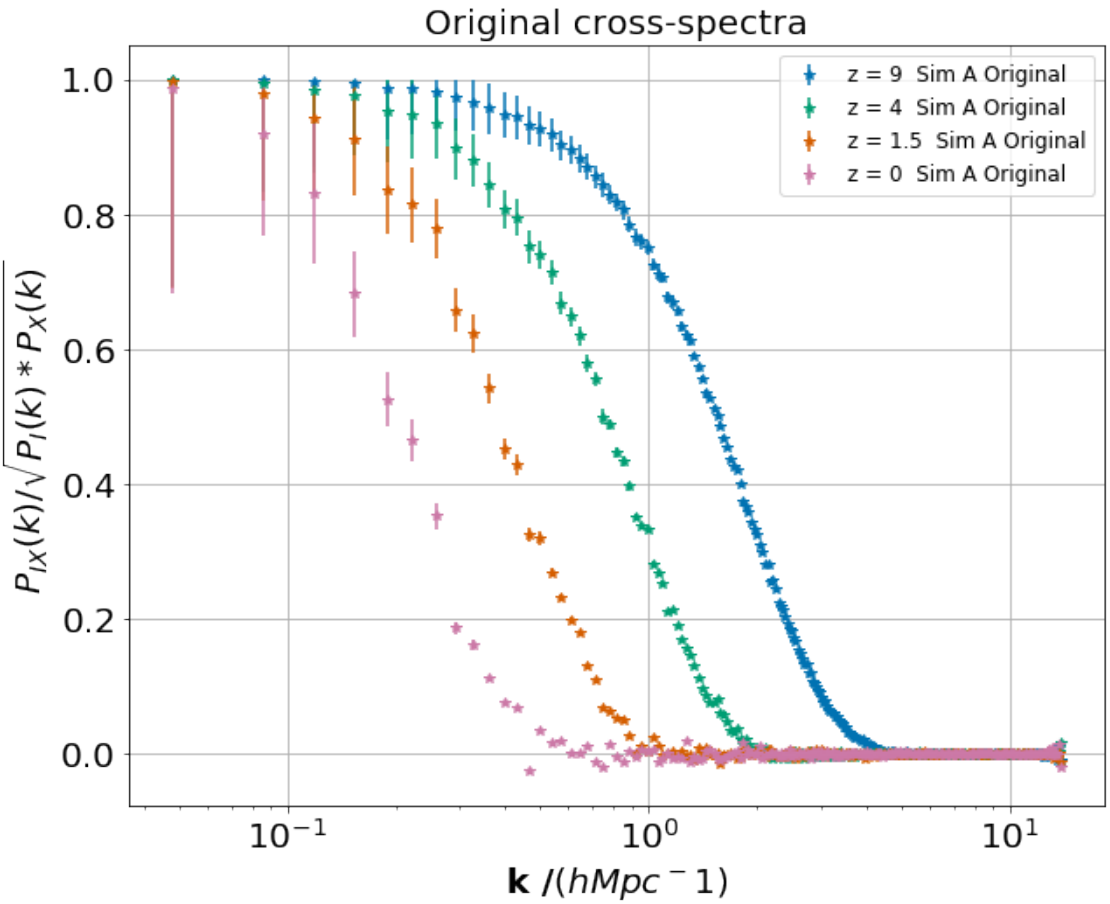

Figure 3.1: Normalized cross-spectra between the initial and final density fields as a function of scale. At small wavenumbers k (large scales) the two fields are perfectly correlated because the universe is not affected by gravitational collapse over such scales (they are both uniform). On the other 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.

initial and the final fields (from Sim A) can be seen in Figure 3.1. 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 gravitational collapse does not have a large impact over such large scales. Because of this, both the initial and the final density fields tend to be very uniform, which preserves the correlation on these scales. However, over small scales, gravity has

a large impact. This results in 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 breakdown in correlation.

Figure 3.1 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 3.2, where we compare it with the original correlation at different redshifts.

3.3 Analysis

The cross-spectra presented in Figure 3.2 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 3.3 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 the decisive role in the per-

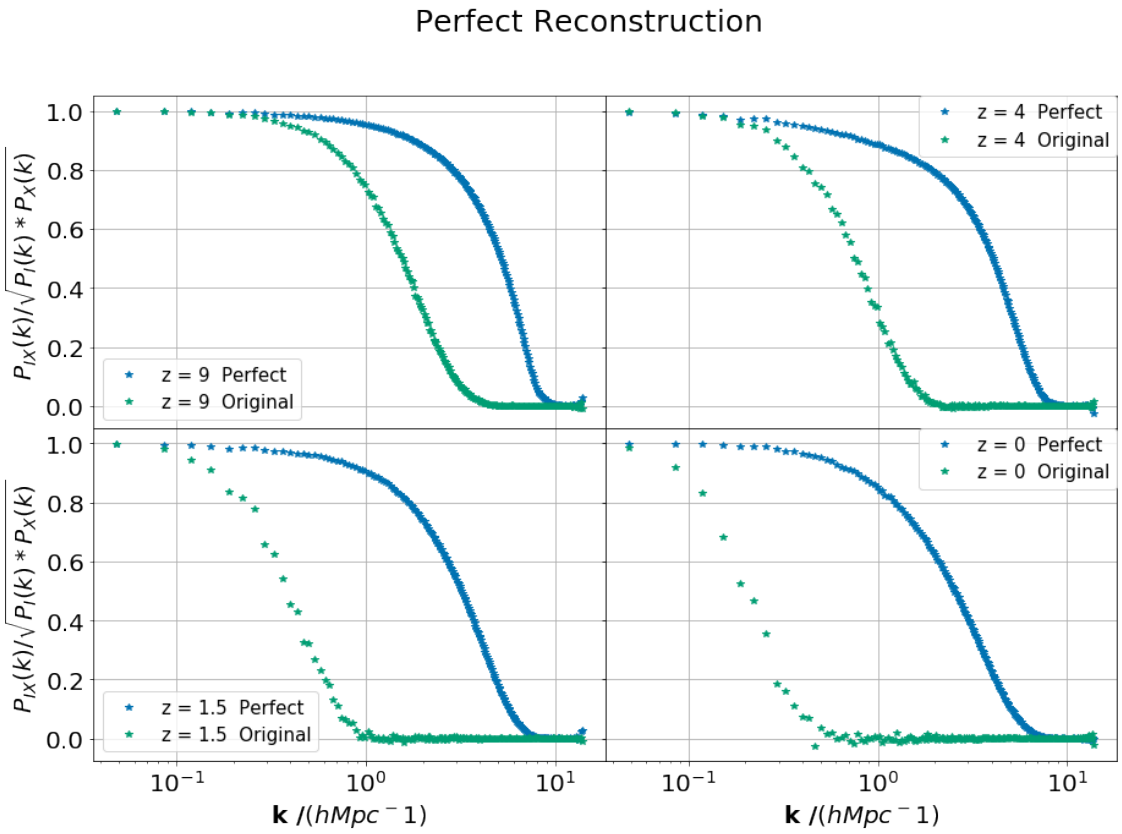

Figure 3.2: Normalized cross-spectra between the initial and the reconstructed fields compared to the original correlation. A large improvement in the correlation was achieved with the perfect reconstruction. This shows up as a shift of the decorrelation scale towards larger k (smaller scales). However, 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 more information being recovered from smaller redshifts.

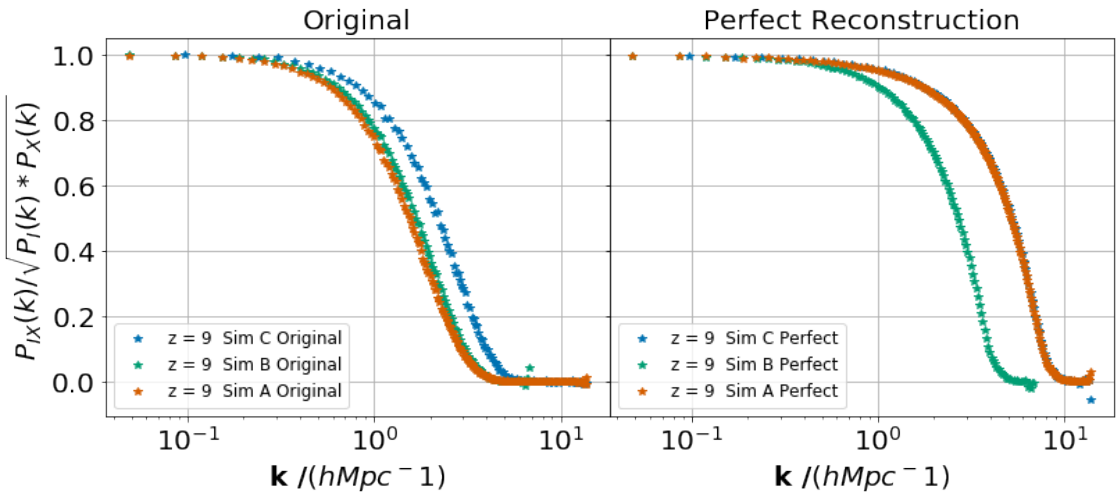

Figure 3.3: Normalized cross-spectra across the three simulations. The left plot shows the original correlation between the initial and the final fields, and the right panel shows the correlation with the reconstructed field. For the original correlations, the size of the simulation plays an important 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.

fect reconstruction. However, this was exactly the starting point of this chapter. The limiting factor for this reconstruction is the resolution used to measure the density field. The right panel in Figure 3.3 shows the lowest scale that can be reconstructed depending on the resolution of our simulation.

The perfect reconstructions in Figure 3.2 show that a perfect correlation cannot be achieved even in the ideal case of perfect knowledge of all the starting particle positions. 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 4

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 Chapter 2 we outlined many 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).

hopefully

4.1 The Zel'dovich reconstruction

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 (see Chapter 2). *Pynbody* has tools for calculating the Zel'dovich offset using the particle velocities in a snapshot. This offset is given by:

ref???

$$\Psi_z(\mathbf{q}) = (1+z) \times \mathbf{v}(\mathbf{q}) \times \frac{D(z)}{f(z)} \tag{4.1}$$

Where $D(z)$ is the linear growth factor, $f(z)$ is the rate of linear growth and $\mathbf{v}(\mathbf{q})$ is the velocity field.

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 4.1 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 after is then given by:

$$\Psi(\mathbf{q}) = \Psi_z(\mathbf{q}) - \Psi_{99}(\mathbf{q}) \tag{4.2}$$

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 4.1. 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.

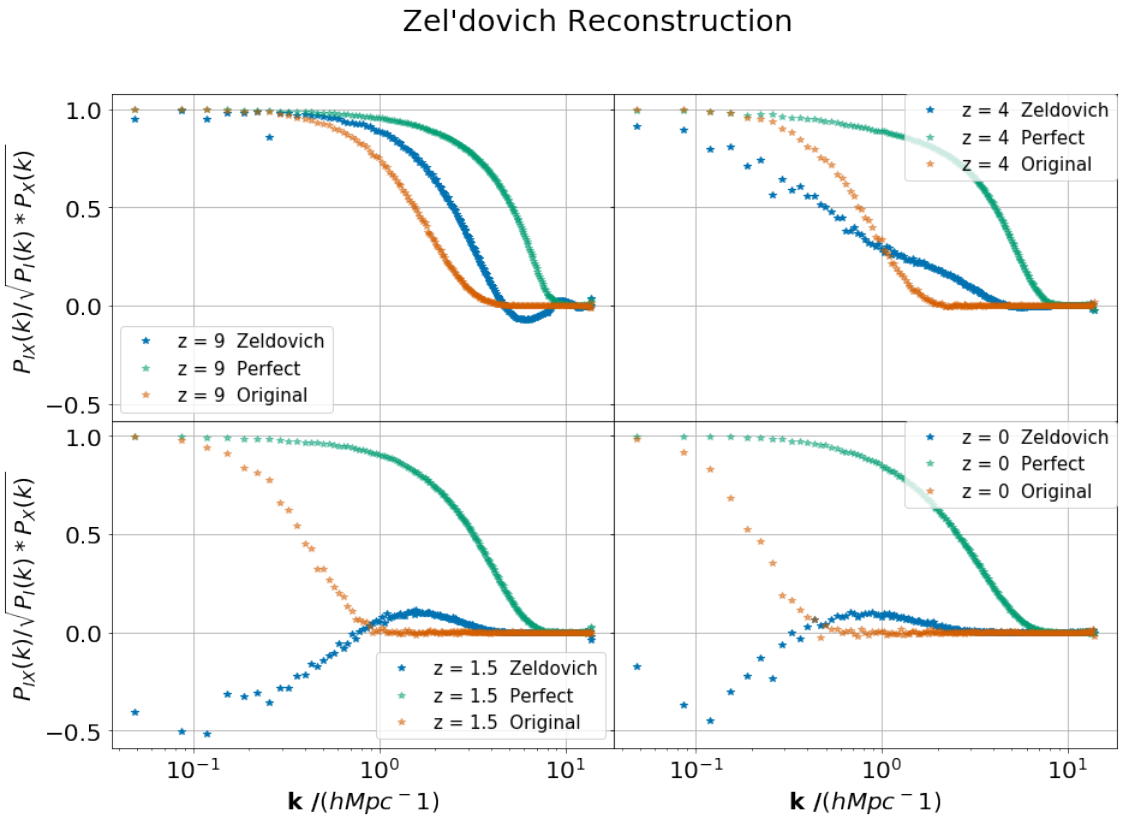

Figure 4.1: Normalized cross-spectra between the Zel'dovich reconstruction and the initial field. This reconstruction was performed by linearly moving the particles back in time (using the Zel'dovich approximation). As we apply the linear approximation directly to the particle velocities, this gives 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.

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 4.1 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 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.

find ref

4.2 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. The other effect of this smoothing is that it makes this method more realistic. With the technology we currently have, observers can maybe detect a few galaxies in a 10 Mpc bin, so ref by smoothing our velocity field over that scale, we simulate a more realistic scenario. The reason for also 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.

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

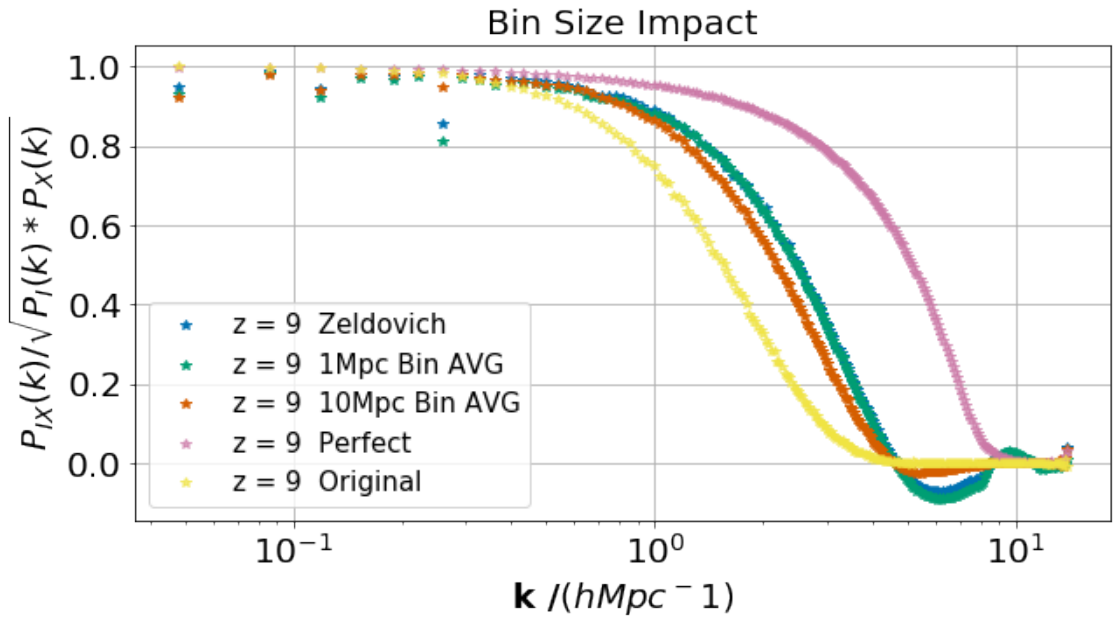

Figure 4.2: Normalized cross-spectra of realistic reconstructions with the initial field. This figure shows the impact of smoothing velocities over certain scales. Because we lose information by smoothing, the 10 Mpc reconstruction recovers less information compared to the 1 Mpc reconstruction. The Zel'dovich reconstruction is also present for comparison. It's proximity to the 1 Mpc reconstruction might indicate that smoothing velocities over this scale is not enough to get back into the linear regime. However, as we are looking at reconstructions from $z = 9$, they all work quite well because most particles are still in the quasi-linear regime.

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.

4.3 Results

The first step is to investigate the impact of the smoothing scale.

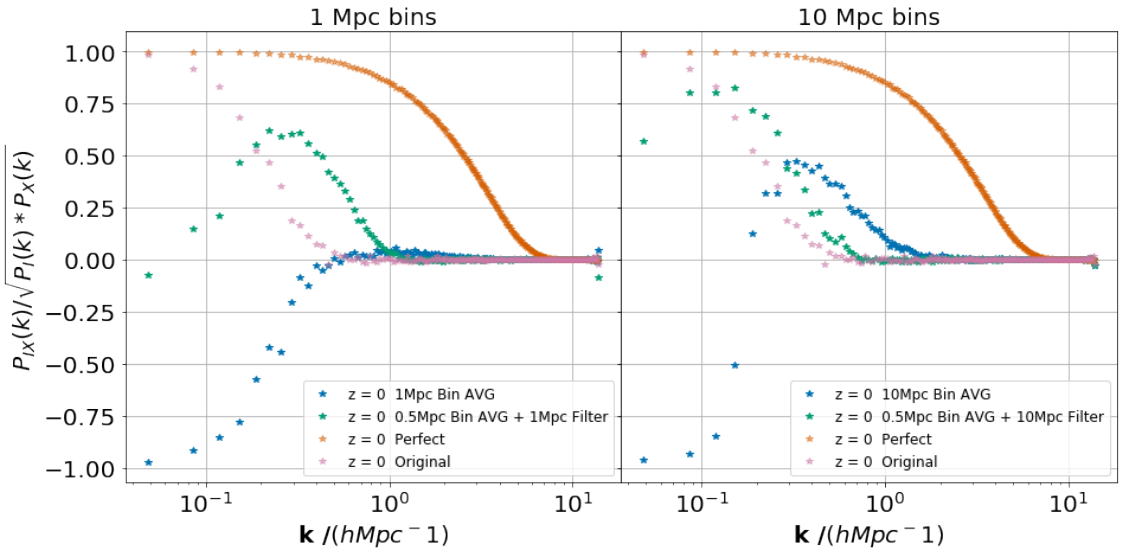

Figure 4.3: Cross spectra of realistic reconstructions from redshift 9 and 0, using 1 Mpc and 10 Mpc bins to smooth velocities. Filter -no filter comparison

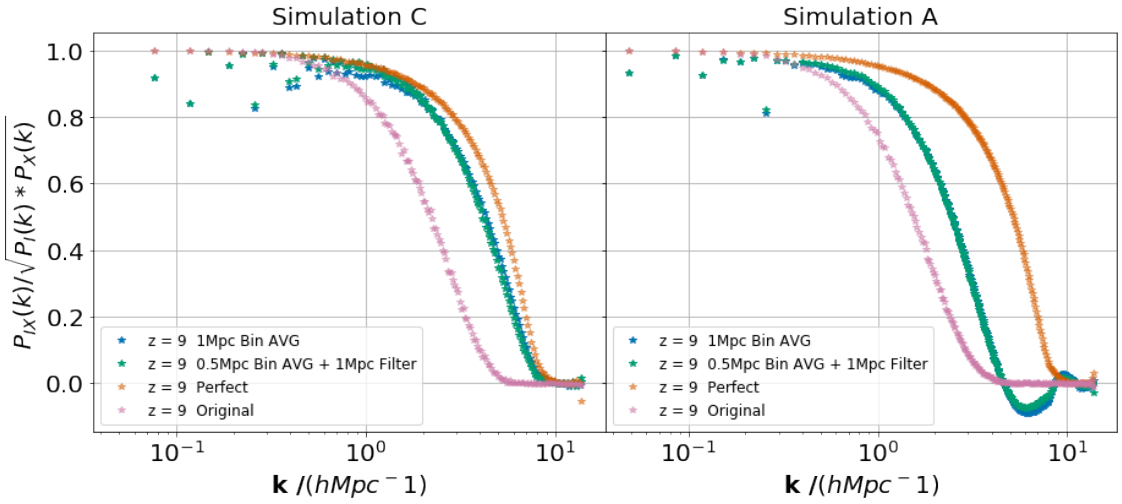

Figure 4.4: Z9 comparison

4.4 Analysis

Talk about the results for different averaging methods and Bin Sizes.

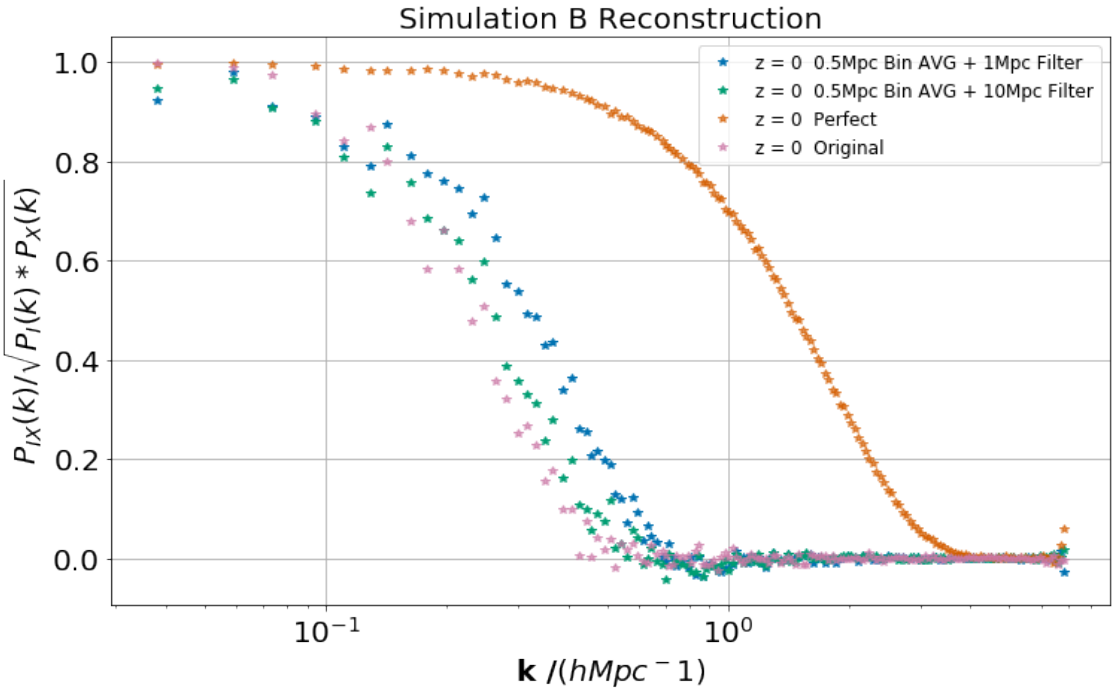

Figure 4.5: Simulation B reconstruction

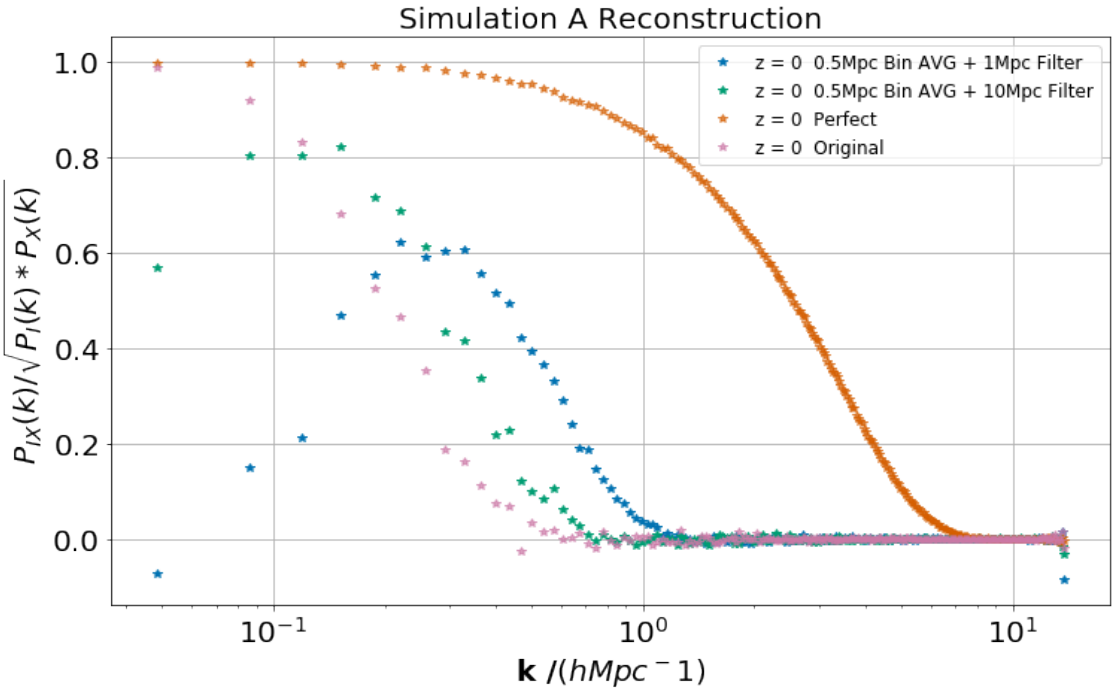

Figure 4.6: Simulation A reconstruction

Chapter 5

Conclusions

5.1 Information loss

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

5.2 Future Work

Talk about the problems encountered and Future Work.

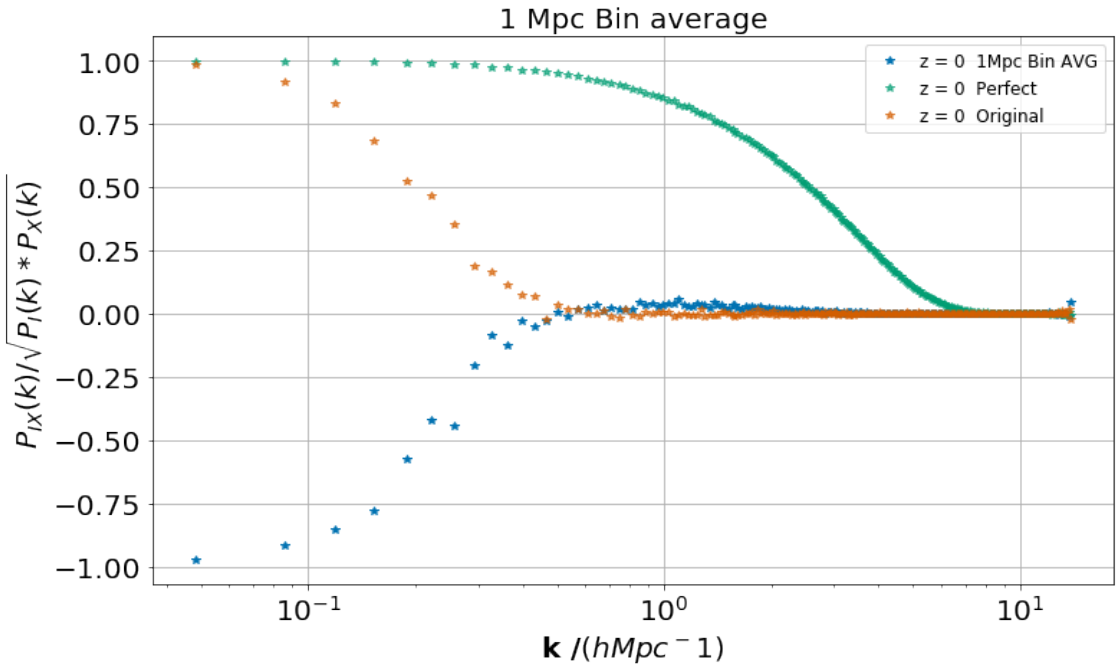

Figure 5.1: Simulation A reconstruction

Bibliography

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

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.

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

Pontzen, A., A. Slosar, et al. (2016). "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.