Galaxy Survey Cosmology

Hannu Kurki-Suonio

24.2.2025

Contents

Preface

These are the lecture notes of the first part (GSC1) of my Galaxy Survey Cosmology course lectured at the University of Helsinki in spring 2017. The second part of the course discussed gravitational lensing with the focus on weak lensing and cosmic shear. The lecture notes for the second part (GSC2) are based on the lectures by Schneider in the textbook Schneider, Kochanek, and Wambsganss (Gravitational Lensing: Strong, Weak and Micro; Springer 2006), and they are available in hand-written form only. This was the first time I lectured this course, and consequently the lecture notes are a bit raw. In the future I hope to add more material about the practical aspects and cosmological results from galaxy surveys.

– Hannu Kurki-Suonio, May 2017

Preface for 2019

The current version of these lecture notes contains no introduction, but jumps directly to the mathematical formulation of correlation functions and power spectra, the main tools in galaxy survey cosmology. For a 4-page introduction to the field, read Sec. 2.7 of [2]. I also gave a (different) introduction during the first lecture. The amount of calculus in these lecture notes may seem formidable to some students. I have aimed for completeness so these notes can be used as a reference for results that may be needed, but the student need not absorb all of the mathematical results. This year I have added new material including recent observational results, and correspondingly some of the older material (Sec. 3 and the latter half of Sec. 5) in these notes were not covered in the course.

– Hannu Kurki-Suonio, February 2019

Preface for 2021

These lecture notes will be updated as the course progresses. The current version is essentially as it was at the end of the 2019 course, except some typos and errors have been fixed. I thank Elina Keihänen for finding some of them.

– Hannu Kurki-Suonio, January 2021

Preface for 2023

For 2023 the course was divided in two. The first part became the new shorter course Galaxy Survey Cosmology, for which these are the lecture notes. The second part became the course Gravitational Lensing, lectured by Elina Keihänen.

– Hannu Kurki-Suonio, January 2023

1 Statistical measures of a density field

We begin by discussing statistical measures of a density field $\rho(\mathbf{x})$ in Euclidian d-dimensional space. We begin with a general treatment where we do not specify in more detail what density we are talking about. It may refer to number density of objects such as galaxies (which will be the main application) or just mass density, but we treat $\rho(\mathbf{x})$ as a continuous quantity for now.

In d dimensions the volume element corresponding to radial distance between r and $r + dr$ is[1]

$$dV = C_d r^{d-1} dr\,, \qquad \text{where} \qquad C_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}\,, \tag{1.1}$$

and the volume within distance R is

$$V(R) = \frac{C_d}{d} R^d\,. \tag{1.2}$$

We will have applications for $d = 1$, 2, and 3, for which

$$C_1 = 2\,,\ C_2 = 2\pi\,,\ C_3 = 4\pi\,. \tag{1.3}$$

The main application is $d = 3$ (3D), but $d = 1$ (1D) corresponds to, e.g., a pencil-beam survey with a very long exposure of a small field on the sky with distance (redshift) determinations for a large number of galaxies along this line of sight. The main 2D application is the distribution of galaxies on the sky, without distance determinations, when the sky is approximated as a flat plane, but it also corresponds to a redshift survey along a great circle on the sky (e.g., the equator, see Fig. 1).

We assume that the density variations originate from a *statistically isotropic and homogeneous ergodic random process*, and we are really interested in the statistics of this random process rather than in that of a particular realization of $\rho(\mathbf{x})$.

It is currently thought that initial density perturbations in the Universe[2] were produced during inflation in the very early universe by quantum fluctuations of the inflaton field, which is a random process that in standard models of inflation satisfies these properties of isotropy, homogeneity, and ergodicity. The density field then evolved until today through deterministic physics, which modified its various statistical measures, but maintained these fundamental properties.

We follow [1] and [2]. Sec. 2.7 of [2] gives a 4-page introduction to the field. I recommend reading it at this point.

1.1 Ergodicity and statistical homogeneity and isotropy

Statistical properties are typically defined as averages of some quantities. We will deal with two kind of averages: *volume average* and *ensemble average*.[3]

The volume average applies to a particular realization (and to some volume V in it). We denote the volume average of a quantity $f(\mathbf{x})$ with the overbar, $\bar{f}$, and it is defined as

$$\bar{f} \equiv \frac{1}{V} \int_V d^d x f(\mathbf{x})\,. \tag{1.4}$$

[1] Here $\Gamma(x) = (x-1)!$ is the gamma function, with values $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, $\Gamma(1) = 1$, $\Gamma(\frac{3}{2}) = \sqrt{\pi}/2$, etc. You get easily other values using the recursion formula $\Gamma(x+1) = x\Gamma(x)$.

[2] 'Universe' with a capital U refers to the universe we live in; whereas 'universe' refers to the theoretical concept, or any hypothetical universe we may consider.

[3] We are redoing here material from Cosmology II, Section 8.1 (2018 version), but with a different approach. In Cosmo II the approach was theoretical, so we assumed the volume V was very large so that $\bar{\rho} = \langle\rho\rangle$. Now the volume V is related to the volume of a galaxy survey, and for accurate treatment we need to take into account that $\bar{\rho} \neq \langle\rho\rangle$.

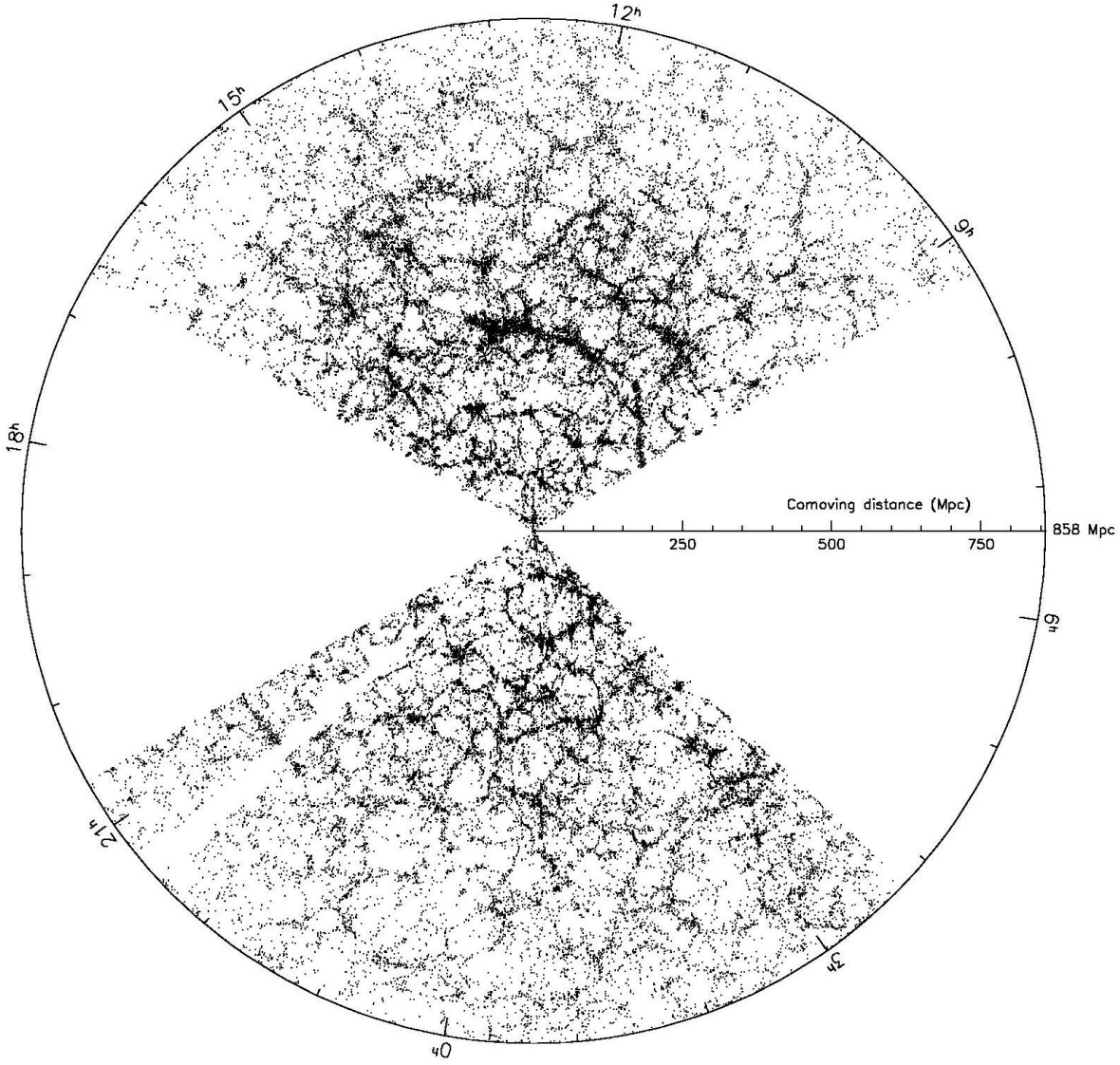


Figure 1: Distribution of galaxies according to the Sloan Digital Sky Survey (SDSS). This figure shows galaxies that are within 2° of the equator and closer than 858 Mpc (assuming $H_0 = 71$ km/s/Mpc). Figure from astro-ph/0310571[12].

The ensemble average refers to the random process. We assume that the observed density field is just one of an *ensemble* of an infinite number of possible *realizations* that could have resulted from the random process. To know the random process, means to know the probability distribution Prob(γ) of the quantities γ produced by it. (At this stage we use the abstract notation of γ to denote the infinite number of these quantities. They could be the values of the density field $\rho(\mathbf{x})$ at every location, or its Fourier coefficients $\rho_{\mathbf{k}}$.) The ensemble average of a quantity f depending on these quantities γ as $f(\gamma)$ is denoted by $\langle f\rangle$ and defined as the (possibly infinite-dimensional) integral

$$\langle f\rangle \equiv \int d\gamma \mathrm{Prob}(\gamma) f(\gamma)\,. \tag{1.5}$$

Here f could be, e.g., the value of $\rho(\mathbf{x})$ at some location $\mathbf{x}$. The ensemble average is also called the *expectation value*. Thus the ensemble represents a probability distribution. And the properties of the density field we will discuss (e.g., statistical homogeneity and isotropy, and ergodicity, see below) will be properties of this ensemble.

Statistical homogeneity means that the expectation value $\langle f(\mathbf{x})\rangle$ must be the same at all $\mathbf{x}$, and thus we can write it as $\langle f\rangle$. Statistical isotropy means that for quantities which involve a direction, the statistical properties are independent of the direction. For example, for vector quantities $\mathbf{v}$, all directions must be equally probable. This implies that $\langle \mathbf{v}\rangle = 0$.

If theoretical properties are those of an ensemble, and we can only observe one realization (the Universe) from that ensemble, how can we compare theory and observation? It seems reasonable that the statistics we get by comparing different parts of a large volume should be similar to the statistics of a given part over different realizations, i.e., that they provide a fair sample of the probability distribution. This is called *ergodicity*. Fields $f(\mathbf{x})$ that satisfy

$$\bar{f} \to \langle f\rangle \qquad \text{as} \qquad V\to\infty \tag{1.6}$$

are called *ergodic*. We assume that the density field is ergodic. It can be shown that a statistically homogeneous and isotropic Gaussian random process is ergodic[4] (but we do not here make the assumption of Gaussianity).

Because of the ergodicity assumption, the concepts of volume and ensemble average are not always kept clearly separate in literature, so that the notation $\langle\cdot\rangle$ is used without specifying which one it refers to, but we shall distinguish between these concepts.[5] The equality of $\bar{f}$ with $\langle f\rangle$ does not hold for a finite volume V; the difference is called *sample variance* or *cosmic variance*. The larger the volume, the smaller is the difference. Since cosmological theory predicts $\langle f\rangle$, whereas observations probe $\bar{f}$ for a limited volume, cosmic variance limits how accurately we can compare theory with observations.

1.2 Density 2-point autocorrelation function

We define the *density perturbation field* as

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - \langle\rho\rangle}{\langle\rho\rangle}. \tag{1.7}$$

Since $\rho \geq 0$, necessarily $\delta \geq -1$.

From statistical homogeneity,

$$\langle\rho(\mathbf{x})\rangle = \langle\rho\rangle \quad \Rightarrow \quad \langle\delta\rangle = 0\,. \tag{1.8}$$

[4] Liddle & Lyth [7] make this statement on p. 73, but do not give a reference to an actual proof.

[5] The $\langle\cdot\rangle$ notation is more convenient than $\bar{\cdot}$ for complicated expressions, so we may sometimes use $\langle\cdot\rangle_V$ for volume average.

Thus we cannot use $\langle\delta\rangle$ as a measure of the inhomogeneity. Instead we can use the square of δ, which is necessarily nonnegative everywhere, so it cannot average out like δ did. Its expectation value $\langle\delta^2\rangle$ is the *variance* of the density perturbation, and the square root of the variance,

$$\delta_{\rm rms} \equiv \sqrt{\langle\delta^2\rangle} \tag{1.9}$$

the *root-mean-square* (rms) density perturbation, is a typical expected absolute value of δ at an arbitrary location.[6] It tells us about how strong the inhomogeneity is, but nothing about the shapes or sizes of the inhomogeneities. To get more information, we introduce the correlation function ξ.

We define the *density 2-point autocorrelation function* (often called just *correlation function*) as

$$\xi(\mathbf{x}_1, \mathbf{x}_2) \equiv \langle\delta(\mathbf{x}_1)\delta(\mathbf{x}_2)\rangle\,. \tag{1.10}$$

It is positive if the density perturbation is expected to have the same sign at both $\mathbf{x}_1$ and $\mathbf{x}_2$, and negative for an overdensity at one and underdensity at the other. Thus it probes how density perturbations at different locations are correlated with each other. Due to statistical homogeneity, $\xi(\mathbf{x}_1, \mathbf{x}_2)$ can only depend on the difference (*separation*) $\mathbf{r} \equiv \mathbf{x}_2 - \mathbf{x}_1$, so we redefine ξ as

$$\xi(\mathbf{r}) \equiv \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle\,. \tag{1.11}$$

From statistical isotropy, $\xi(\mathbf{r})$ is independent of direction, i.e., spherically symmetric (we use this as a generic term for arbitrary d – we might also say 'isotropic' – i.e., for $d = 2$, read 'circularly symmetric', and for $d = 1$, read 'even'),

$$\xi(\mathbf{r}) = \xi(r)\,. \tag{1.12}$$

The correlation function is large and positive for r smaller than the size of a typical over- or underdense region, and becomes small for larger separations.

The correlation function at zero separation gives the variance of the density perturbation,

$$\langle\delta^2\rangle \equiv \langle\delta(\mathbf{x})\delta(\mathbf{x})\rangle \equiv \xi(0)\,. \tag{1.13}$$

We define the volume average of ξ up to a distance R as

$$\bar{\xi}(R) \equiv \frac{1}{V(R)}\int_0^R \xi(r) C_d r^{d-1} dr\,. \tag{1.14}$$

For $d = 3$ this becomes

$$\bar{\xi}(R) \;\equiv\; \frac{3}{R^3}\int_0^R \xi(r) r^2 dr \;\equiv\; \frac{3}{R^3} J_3(R)\,, \tag{1.15}$$

where

$$J_3(R) \equiv \int_0^R \xi(r) r^2 dr \tag{1.16}$$

is called the "J_3 integral" (not a Bessel function; see Sec. 7.3.1 for more on J_ℓ and K_ℓ integrals).

[6] In other words, $\delta_{\rm rms}$ is the standard deviation of $\rho/\langle\rho\rangle$.

1.3 Finite surveys and integral constraint

Overdensities must be balanced by underdensities and positive correlations are (usually) balanced by negative correlations. We show in Sec. 1.6 that under certain assumptions the *integral constraint*

$$\int d^d r\, \xi(\mathbf{r}) = 0\,, \tag{1.17}$$

where the integral goes to infinity, holds. At small separations $\mathbf{r}$ the correlations are positive, so if the integral is extended to only some moderate finite maximum separation, the integral typically is positive, as more of the negative correlations, which occur at large separations, are missed.

Consider then a single realization and a finite volume V, so that (probably) $\bar{\rho} \neq \langle\rho\rangle$. We define

$$\widehat{\delta}(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}} \qquad \text{and} \qquad \widehat{\xi}(\mathbf{r}) \equiv \frac{1}{V(\mathbf{r})} \int_{V(\mathbf{r})} d^d x\, \widehat{\delta}(\mathbf{x})\widehat{\delta}(\mathbf{x}+\mathbf{r})\,, \tag{1.18}$$

where $V(\mathbf{r}) \subset V$ is that subvolume of V where also $\mathbf{x}+\mathbf{r}$ is in $V(\mathbf{r})$. If we were able to determine the density field $\rho(\mathbf{r})$ in V, we could calculate $\widehat{\delta}(\mathbf{x})$ and $\widehat{\xi}(\mathbf{r})$, and $\widehat{\xi}(\mathbf{r})$ would serve as an estimate of $\xi(\mathbf{r})$ from survey volume V. We would get an estimate for all values of $\mathbf{r}$, where the separation $\mathbf{r}$ fits in V. Denote this set of $\mathbf{r}$ values $V_{\mathbf{r}}$. One can easily show (**exercise**) that

$$\int_{V_{\mathbf{r}}} d^d r\, \widehat{\xi}(\mathbf{r}) = 0\,, \tag{1.19}$$

the *integral constraint for finite survey*. The key to showing this is to note that the double integral over $\mathbf{r}$ and $\mathbf{x}$ goes exactly over all pairs $(\mathbf{x}, \mathbf{x}')$ in V (where $\mathbf{x}' = \mathbf{x} + \mathbf{r}$), i.e.,

$$\int_{V_{\mathbf{r}}} d^d r \int_{V(\mathbf{r})} d^d x \;=\; \int_V d^d x \int_V d^d x'\,. \tag{1.20}$$

We just argued that if we had the true correlation function $\xi(\mathbf{r})$ in (1.19) instead of the *observed correlation function* $\widehat{\xi}(\mathbf{r})$, we would expect the integral to be positive. Thus $\langle\widehat{\xi}(\mathbf{r})\rangle \neq \xi(\mathbf{r})$, meaning that our estimator is *biased*. This is not unexpected, since, if $\bar{\rho} \neq \langle\rho\rangle$, then $\widehat{\delta}(\mathbf{x}) \neq \delta(\mathbf{x})$; there is both a scaling and an offset:

$$\widehat{\delta}(\mathbf{x}) \;=\; \frac{\langle\rho\rangle}{\bar{\rho}}\delta(\mathbf{x}) + \frac{\langle\rho\rangle - \bar{\rho}}{\bar{\rho}}\,. \tag{1.21}$$

Because of the offset, for small perturbations $\widehat{\delta}(\mathbf{x})$ and $\delta(\mathbf{x})$ may even have opposite signs.

We would like to figure out the bias, i.e, the difference between $\langle\widehat{\xi}(\mathbf{r})\rangle$ and $\xi(\mathbf{r})$. Consider a *subensemble with a fixed* $\bar{\rho}$, whose expectation values we denote by $\langle\cdot\rangle'$: $\rho(\mathbf{x}) = \bar{\rho}\left[1 + \widehat{\delta}(\mathbf{x})\right]$ and

$$\begin{aligned}
\langle\rho(\mathbf{x})\rho(\mathbf{x}+\mathbf{r})\rangle' &= \bar{\rho}^2 \left\langle \left[1 + \widehat{\delta}(\mathbf{x})\right]\left[1 + \widehat{\delta}(\mathbf{x}+\mathbf{r})\right] \right\rangle' \\
&= \bar{\rho}^2 + \langle\widehat{\delta}(\mathbf{x})\rangle' + \langle\widehat{\delta}(\mathbf{x}+\mathbf{r})\rangle' + \bar{\rho}^2\langle\widehat{\delta}(\mathbf{x})\widehat{\delta}(\mathbf{x}+\mathbf{r})\rangle' \\
&\equiv \bar{\rho}^2\left[1 + 0 + 0 + \tilde{\xi}(\mathbf{r})\right] \;\equiv\; \bar{\rho}^2 C_V\left[1 + \xi(\mathbf{r})\right]\,,
\end{aligned} \tag{1.22}$$

where

$$\tilde{\xi}(\mathbf{r}) \equiv \langle\widehat{\delta}(\mathbf{x})\widehat{\delta}(\mathbf{x}+\mathbf{r})\rangle' \tag{1.23}$$

so that $\tilde{\xi}(\mathbf{r}) = \langle\widehat{\xi}(\mathbf{r})\rangle'$and we have made the *assumption* that $1 + \tilde{\xi}(\mathbf{r})$ and $1 + \xi(\mathbf{r})$ are related by a constant bias factor that we denoted by C_V.

We can now calculate the double integral of $\langle\rho(\mathbf{x})\rho(\mathbf{x}+\mathbf{r})\rangle'$ in two different ways:

$$\begin{aligned}\int_{V_\mathbf{r}} d^dr \int_{V(\mathbf{r})} d^dx\langle\rho(\mathbf{x})\rho(\mathbf{x}+\mathbf{r})\rangle' &= \int_V d^dx \int_V d^dx' \,\langle\rho(\mathbf{x})\rho(\mathbf{x}')\rangle' \\ &= \left\langle \int_V d^dx\,\rho(\mathbf{x}) \int_V d^dx'\,\rho(\mathbf{x}')\right\rangle' = \langle V\bar\rho V\bar\rho\rangle' = V^2\bar\rho^2\end{aligned} \tag{1.24}$$

and

$$\begin{aligned}\int_{V_\mathbf{r}} d^dr \int_{V(\mathbf{r})} d^dx\,\langle\rho(\mathbf{x})\rho(\mathbf{x}+\mathbf{r})\rangle' &= \bar\rho^2 C_V \int_{V_\mathbf{r}} d^dr \int_{V(\mathbf{r})} d^dx\,[1+\xi(\mathbf{r})] \\ &\equiv \bar\rho^2 C_V\, V^2(1+\xi_V)\,,\end{aligned} \tag{1.25}$$

where we defined

$$1+\xi_V \equiv \frac{1}{V^2}\int_{V_\mathbf{r}} d^dr \int_{V(\mathbf{r})} d^dx\,[1+\xi(\mathbf{r})] = \frac{1}{V^2}\int_V d^dx \int_V d^dx'\,\left[1+\xi(\mathbf{x}'-\mathbf{x})\right]\,, \tag{1.26}$$

i.e., ξ_V is the *average true correlation over all pairs of points in the survey volume.* Comparing (1.24) to (1.25) we find $C_V = 1/(1+\xi_V)$, i.e.,

$$1+\langle\widehat{\xi}(\mathbf{r})\rangle = \frac{1+\xi(\mathbf{r})}{1+\xi_V}\,. \tag{1.27}$$

The result is independent of $\bar\rho$, so therefore I wrote $\langle\cdot\rangle$ instead of $\langle\cdot\rangle'$.

Usually ξ_V is positive, since the finite survey volume is likely to miss more negative than positive correlations, since the former occur at large separations. For sufficiently large survey volumes $\xi_V \ll 1$, so we can approximate

$$1+\langle\widehat{\xi}(\mathbf{r})\rangle \approx (1-\xi_V)[1+\xi(\mathbf{r})] = 1+\xi(\mathbf{r})-\xi_V-\xi_V\xi(\mathbf{r})\,. \tag{1.28}$$

For large separations, where $|\xi(\mathbf{r})| \ll 1$, we can further approximate

$$\langle\widehat{\xi}(\mathbf{r})\rangle \approx \xi(\mathbf{r})-\xi_V\,, \tag{1.29}$$

i.e., we expect a constant shift downwards.

This bias is known as the *integral constraint* effect. It comes because the integral constraint forces the estimated correlation factor to average to zero within the survey volume, although the true correlation probably does not yet average to zero in this volume. The effect appears to be related to defining the density perturbation with respect to the survey mean density $\bar\rho$ instead of the "true" mean density $\langle\rho\rangle$, i.e., the ensemble average, which by ergodicity is the mean density for an infinite volume. However, the situation is more subtle, since the derived result, $C_V = 1/(1+\xi_V)$ is independent of $\bar\rho$, i.e., we get the same bias whether $\bar\rho < \langle\rho\rangle$ or $\bar\rho > \langle\rho\rangle$, and even if, by accident, $\bar\rho = \langle\rho\rangle$.

Note that $\langle\bar\rho\rangle = \langle\rho\rangle$, but $\langle\bar\rho^2\rangle \neq \langle\rho\rangle^2$:

$$\begin{aligned}\bar\rho^2 &= \left[\frac{1}{V}\int d^dx\,\rho(\mathbf{x})\right]^2 = \frac{\langle\rho\rangle^2}{V^2}\int d^dx\,[1+\delta(\mathbf{x})]\int d^dx'\,\left[1+\delta(\mathbf{x}')\right] \\ &= \langle\rho\rangle^2 + \frac{\langle\rho\rangle^2}{V}\left[\int d^dx\delta(\mathbf{x}) + \int d^dx\delta(\mathbf{x}')\right] + \frac{\langle\rho\rangle^2}{V^2}\int d^dx d^dx'\,\delta(\mathbf{x})\delta(\mathbf{x}')\end{aligned} \tag{1.30}$$

so that

$$\langle\bar\rho^2\rangle = \langle\rho\rangle^2 + \frac{\langle\rho\rangle^2}{V^2}\int d^dx d^dx'\,\xi(\mathbf{x}'-\mathbf{x}) = \langle\rho\rangle^2\,[1+\xi_V]\,. \tag{1.31}$$

This bias can be seen in simulations, where many realizations are produced (from a single realization the result may be too noisy to see it well). If **either** all realizations are produced with the same $\bar{\rho}$, **or** $\bar{\rho}$ is allowed to vary between the realizations but the $\widehat{\delta}(\mathbf{x})$ is defined for each realization with respect to the $\bar{\rho}$ of that realization, the $1/(1+\xi_V)$ bias is seen. However, if $\bar{\rho}$ is allowed to vary between the realizations and the $\delta(\mathbf{x})$ is defined for each realization with respect to the average $\bar{\rho}$ of all realization, the bias is not seen, since this case models the ensemble of universes and $\delta(\mathbf{x})$ the deviation from the ensemble average. (The actual simulations I refer to here are for a discrete set of points (representing galaxies) rather than a continuous density field. We introduce the discrete point (galaxy) distribution in Sec. 2 and the integral constraint in that case is discussed in Sec. 8.)

Another derivation of the integral constant bias. The above derivation parallels the standard derivation in the literature [39, 15]. However, I was not satisfied with it, since no justification for the assumption of a constant (independent of $\mathbf{r}$) bias factor C_V was given. Therefore I present another derivation:

Consider the full ensemble, i.e., $\bar{\rho}$ not restricted to a fixed value. Now

$$\langle\rho(\mathbf{x})\rho(\mathbf{x}+\mathbf{r})\rangle \;=\; \langle\rho\rangle^2\,\langle[1+\delta(\mathbf{x})]\,[1+\delta(\mathbf{x}+\mathbf{r})]\rangle \;=\; \langle\rho\rangle^2\,[1+\xi(\mathbf{r})]\;. \tag{1.32}$$

But, defining $\widehat{\delta}(\mathbf{x}) \equiv [\rho(\mathbf{x})-\bar{\rho}]/\bar{\rho}$ for each realization, this is also

$$\begin{aligned}\langle\rho(\mathbf{x})\rho(\mathbf{x}+\mathbf{r})\rangle &= \left\langle\bar{\rho}^2\left[1+\widehat{\delta}(\mathbf{x})\right]\left[1+\widehat{\delta}(\mathbf{x}+\mathbf{r})\right]\right\rangle \\ &= \langle\bar{\rho}^2\rangle + \left\langle\bar{\rho}^2\widehat{\delta}(\mathbf{x})\right\rangle + \left\langle\bar{\rho}^2\widehat{\delta}(\mathbf{x}+\mathbf{r})\right\rangle + \left\langle\bar{\rho}^2\widehat{\delta}(\mathbf{x})\widehat{\delta}(\mathbf{x}+\mathbf{r})\right\rangle\;. \end{aligned}\tag{1.33}$$

The two middle terms vanish, since for a subensemble with fixed $\bar{\rho}$, $\langle\bar{\rho}^2\widehat{\delta}(\mathbf{x})\rangle' = \bar{\rho}^2\langle\widehat{\delta}(\mathbf{x})\rangle' = 0$, so averaging these zeros over the $\bar{\rho}$ distribution gives also zero. If we now *assume* that $\bar{\rho}^2$ and $\widehat{\delta}(\mathbf{x})\widehat{\delta}(\mathbf{x}+\mathbf{r})$ are *uncorrelated* so that

$$\left\langle\bar{\rho}^2\widehat{\delta}(\mathbf{x})\widehat{\delta}(\mathbf{x}+\mathbf{r})\right\rangle \;=\; \langle\bar{\rho}^2\rangle\langle\widehat{\delta}(\mathbf{x})\widehat{\delta}(\mathbf{x}+\mathbf{r})\rangle \;=\; \langle\rho\rangle^2\,[1+\xi_V]\,\langle\widehat{\xi}(\mathbf{r})\rangle\,, \tag{1.34}$$

we get

$$\langle\rho(\mathbf{x})\rho(\mathbf{x}+\mathbf{r})\rangle \;=\; \langle\rho\rangle^2\,[1+\xi(\mathbf{r})] \;=\; \langle\rho\rangle^2\,[1+\xi_V]\left[1+\langle\widehat{\xi}(\mathbf{r})\rangle\right]\;. \tag{1.35}$$

Thus we get the same result when the assumption is replaced by another assumption.

Exercise: Integral constraint for a single realization. For a single realization and finite volume (so that $\bar{\rho}\neq\langle\rho\rangle$) we can define

$$\widehat{\delta}(\mathbf{x}) \equiv \frac{\rho(\mathbf{x})-\bar{\rho}}{\bar{\rho}} \qquad \text{and} \qquad \widehat{\xi}(\mathbf{r}) \equiv \frac{1}{V}\int_V d^dx\,\widehat{\delta}(\mathbf{x})\widehat{\delta}(\mathbf{x}+\mathbf{r})\,. \tag{1.36}$$

1. **Theoretical approach:** assume periodic boundary conditions. This makes also $\widehat{\xi}(\mathbf{r})$ periodic. All integrals, also (1.37), are to be taken over the volume V. Show that

$$\int_V d^dr\,\widehat{\xi}(\mathbf{r}) = 0 \tag{1.37}$$

   (the *integral constraint*). Thus the positive values of $\widehat{\xi}$ at small separations must be compensated by negative values at larger r. Note that here we do not need any statistical assumptions (like statistical homogeneity or ergodicity). If $\widehat{\xi}(\mathbf{r})\to 0$ for large r fast enough, for large volumes the boundary conditions do not matter.

2. **Practical approach:** To avoid using boundary conditions and going outside the volume, redefine

$$\widehat{\xi}(\mathbf{r}) \equiv \frac{1}{V(\mathbf{r})}\int_{V(\mathbf{r})} d^dx\,\widehat{\delta}(\mathbf{x})\widehat{\delta}(\mathbf{x}+\mathbf{r}) \tag{1.38}$$

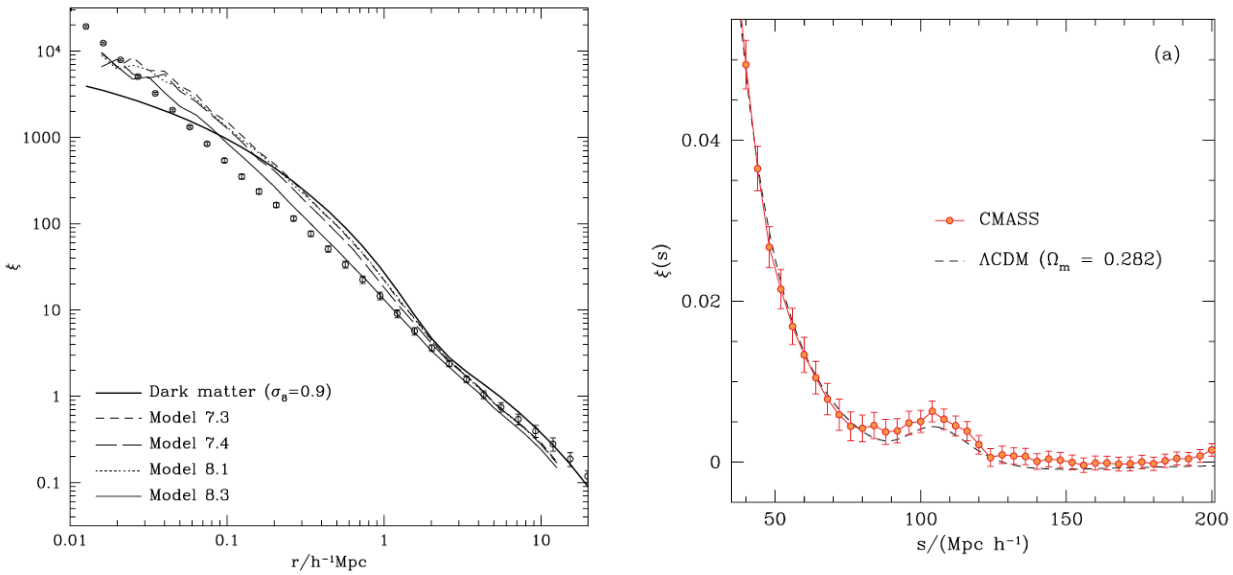


Figure 2: The 2-point correlation function $\xi(r)$ from galaxy surveys. Left: Small scales shown in a log-log plot. The circles with error bars show the observational determination from the APM galaxy survey [4]. The different lines are theoretical predictions by [5] (this is Fig. 9 from [5]). Right: Large scales shown in a linear plot. Red circles with error bars show the observational determination from the CMASS Data Release 9 (DR9) sample of the Baryonic Oscillation Spectroscopic Survey (BOSS). The dashed line is a theoretical prediction from the ΛCDM model. The bump near $100\,h^{-1}$Mpc is the baryon acoustic oscillation (BAO) peak that will be discussed in Sec. 10. This is Fig. 2a from [6].

so that the integral for each $\mathbf{r}$ goes over only those values of $\mathbf{x}$, for which both $\mathbf{x}$ and $\mathbf{x}+\mathbf{v}$ are within the volume. This is what one does with real galaxy surveys. Show that

$$\int_{V_\mathbf{r}} d^d r\, \widehat{\xi}(\mathbf{r}) = 0\,, \tag{1.39}$$

where the integral goes over those values of $\mathbf{r}$, for which $\widehat{\xi}(\mathbf{r})$ is defined by (1.38), i.e., $\mathbf{r}$ separates two points inside the volume.

1.4 Fourier expansion

Fourier analysis is a method for separating out the different distance scales, so that the dependence of the physics on distance scale becomes clear and easy to handle.

For a Fourier analysis of the density field we consider a cubic volume $V = L^d$ and assume periodic boundary conditions. This does not imply that the density field should be periodic in reality; we are just interested only in the density field within the volume, and so we can replace the part outside the volume with a periodic replication of the volume. This will introduce discontinuities at the volume boundary. The expansion (1.40) itself does not assume anything about $f(\mathbf{x})$ outside V (the expansion will be correct inside V and outside V it will represent such a periodic extension; the discontinuities imply that the expansion will contain high-$\mathbf{k}$ modes); but some of the discussion below, like convolution with a window function (Sec. 1.10), makes use of this periodicity. Near the boundary of the volume the window function will extend outside the volume; and thus in reality this will introduce edge effects as the real universe is not periodic. This will have to be treated (Sec. 9.2) together with the fact that an actual survey does not cover exactly a cubic volume. For theoretical work one can also take V to be much larger than the observable universe so that the boundaries are so far away that they do not matter.

We can now expand any function of space $f(\mathbf{x})$ as a Fourier series

$$f(\mathbf{x}) = \sum_{\mathbf{k}} f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}\,, \tag{1.40}$$

where the wave vectors $\mathbf{k} = (k_1, \ldots, k_d)$ take values

$$k_i = n_i \frac{2\pi}{L}\,, \qquad n_i = 0, \pm 1, \pm 2, \ldots \tag{1.41}$$

(Note that we use the Fourier conventions of [2], not those of [1].)

The Fourier coefficients $f_{\mathbf{k}}$ are obtained as

$$f_{\mathbf{k}} = \frac{1}{V}\int_V f(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d^d x\,. \tag{1.42}$$

The term $\mathbf{k} = 0$ gives the mean value,

$$f_0 = \bar{f}\,. \tag{1.43}$$

The Fourier coefficients are complex numbers even though we are dealing with real quantities[7] $f(\mathbf{x})$. From the reality $f(\mathbf{x})^* = f(\mathbf{x})$ follows that

$$f_{-\mathbf{k}} = f_{\mathbf{k}}^*\,. \tag{1.44}$$

Thus Fourier modes comes in pairs

$$f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} + f_{\mathbf{k}}^* e^{-i\mathbf{k}\cdot\mathbf{x}} \;=\; 2 * \mathrm{Re} f_{\mathbf{k}} \cos\mathbf{k}\cdot\mathbf{x} - 2 * \mathrm{Im} f_{\mathbf{k}} \sin\mathbf{k}\cdot\mathbf{x}\,, \tag{1.45}$$

and only the real part of each, $\mathrm{Re} f_{\mathbf{k}} \cos\mathbf{k}\cdot\mathbf{x} - \mathrm{Im} f_{\mathbf{k}} \sin\mathbf{k}\cdot\mathbf{x}$, survives; so to visualize a Fourier mode, just visualize this real part, which is a sinusoidal plane wave. The size of the Fourier coefficients depends on the volume V – increasing V tends to make the $f_{\mathbf{k}}$ smaller to compensate for the denser sampling of $\mathbf{k}$ in Fourier space.

The Fourier expansion is an expansion in terms of plane waves $e^{i\mathbf{k}\cdot\mathbf{x}}$, which form an *orthogonal* and *complete* (closed) set of functions in the Euclidean volume V. (We will later encounter other such expansions in terms of other functions.) Thus they satisfy the orthogonality relation

$$\int dV \left(e^{i\mathbf{k}\cdot\mathbf{x}}\right)^* \left(e^{i\mathbf{k}'\cdot\mathbf{x}}\right) \;=\; \int dV e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{x}} \;=\; V\delta_{\mathbf{k}\mathbf{k}'}\,, \tag{1.46}$$

where $\delta_{\mathbf{k}\mathbf{k}'}$ is the Kronecker delta ($\delta_{\mathbf{k}\mathbf{k}'} = 1$ for $\mathbf{k} = \mathbf{k}'$, and $\delta_{\mathbf{k}\mathbf{k}'} = 0$ otherwise), and the *closure* (completeness) relation

$$\frac{1}{V}\sum_{\mathbf{k}} \left(e^{i\mathbf{k}\cdot\mathbf{x}}\right)^* \left(e^{i\mathbf{k}\cdot\mathbf{x}'}\right) \;=\; \frac{1}{V}\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{x}'-\mathbf{x})} \;=\; \delta_D^d(\mathbf{x}'-\mathbf{x})\,, \tag{1.47}$$

where $\delta_D^d(\mathbf{x}'-\mathbf{x})$ is the d-dimensional Dirac delta function.[8] Do not confuse the Dirac and Kronecker deltas with the density perturbation $\delta(\mathbf{x})$ or its Fourier coefficient $\delta_{\mathbf{k}}$! Thus the functions

$$\left\{\frac{1}{\sqrt{V}} e^{i\mathbf{k}\cdot\mathbf{x}}\right\} \tag{1.49}$$

[7] In Gravitational Lensing, one introduces the complex shear, so these reality conditions do not apply to its Fourier coefficients/transform.

[8] The Dirac delta function is not a true function but rather an operator (the correct mathematical term is 'distribution') defined by its action on a function $f(\mathbf{x})$ under an integral:

$$\int_V \delta_D^d(\mathbf{x}'-\mathbf{x}) f(\mathbf{x}') d^d x \;\equiv\; f(\mathbf{x})\,. \tag{1.48}$$

It can be thought of as a limit of a set of functions that have large values very near 0, are close to zero elsewhere, and are normalized so that their integral gives 1.

form an *orthonormal* set.

The point of a completeness relation[9] for orthogonal functions is that any function can indeed be expanded in them. Here

$$\sum_{\mathbf{k}} f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} \;=\; \frac{1}{V}\sum_{k}\int_V d^dx' f(\mathbf{x}')e^{-i\mathbf{k}\cdot\mathbf{x}'}e^{i\mathbf{k}\cdot\mathbf{x}} \;=\; \int d^dx' f(\mathbf{x}')\delta_D^d(\mathbf{x}-\mathbf{x}') \;=\; f(\mathbf{x})\,. \tag{1.50}$$

The *convolution theorem* states that convolution in coordinate space becomes just multiplication in Fourier space (**exercise**):

$$(f*g)(\mathbf{x}) \;\equiv\; \int_V d^dx' f(\mathbf{x}')g(\mathbf{x}-\mathbf{x}') \;=\; \int_V d^dx' f(\mathbf{x}-\mathbf{x}')g(\mathbf{x}') \;=\; V\sum_{\mathbf{k}} f_{\mathbf{k}}g_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}}\,, \tag{1.51}$$

and multiplication in coordinate space becomes convolution in Fourier space (**exercise**):

$$\frac{1}{V}\int_V f(\mathbf{x})g(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}d^dx \;=\; \sum_{\mathbf{q}} f_{\mathbf{q}}g_{\mathbf{k}-\mathbf{q}}\,. \tag{1.52}$$

The *Plancherel formula* states (**exercise**):

$$\frac{1}{V}\int_V d^dx f(\mathbf{x})g(\mathbf{x}) \;=\; \sum_{\mathbf{k}} f^*_{\mathbf{k}}g_{\mathbf{k}}\,. \tag{1.53}$$

With $g=f$ this becomes the *Parseval formula*:

$$\frac{1}{V}\int_V d^dx f(\mathbf{x})^2 \;=\; \sum_{\mathbf{k}} |f_{\mathbf{k}}|^2\,. \tag{1.54}$$

A great benefit of Fourier analysis is that derivation is replaced by multiplication:

$$\mathbf{g}(\mathbf{x}) \equiv \nabla f(\mathbf{x}) = \nabla\sum_{\mathbf{k}} f_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}} = \sum_{\mathbf{k}} i\mathbf{k}f_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}} \quad\Rightarrow\quad \mathbf{g}_{\mathbf{k}} = i\mathbf{k}f_{\mathbf{k}}\,. \tag{1.55}$$

1.5 Fourier transform

The separation of neighboring k_i values is $\Delta k_i = 2\pi/L$, so we can write

$$f(\mathbf{x}) \;=\; \sum_{\mathbf{k}} f_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}}\left(\frac{L}{2\pi}\right)^d \Delta k_1\ldots\Delta k_d \;\approx\; \frac{1}{(2\pi)^d}\int f(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}d^dk\,, \tag{1.56}$$

where

$$f(\mathbf{k}) \equiv L^d f_{\mathbf{k}}\,. \tag{1.57}$$

replacing the Fourier series with the Fourier integral.

In the limit $V\to\infty$, the approximation in (1.56) becomes exact, and we have the *Fourier transform* pair

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{(2\pi)^d}\int f(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}d^dk \\ f(\mathbf{k}) &= \int f(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}d^dx\,. \end{aligned} \tag{1.58}$$

[9]For some reason these closure relations are rarely given in standard sources for mathematical methods. (Mathematicians do not like the Dirac delta function?) For example, I could not find Eq. (1.47) anywhere.

Note that this assumes that the integrals converge, which requires that $f(\mathbf{x}) \to 0$ for $|\mathbf{x}| \to \infty$.[10] Thus we don't use this for, e.g., $\delta(\mathbf{x})$, but for, e.g., the correlation function $\xi(\mathbf{x})$ the Fourier transform is appropriate.

A special case is $f = 1$ (which does not satisfy $f(\mathbf{x}) \to 0$, so it does not lead to a true function), whose Fourier transform is the Dirac delta function:

$$\int d^d x e^{-i\mathbf{k}\cdot\mathbf{x}} \;=\; (2\pi)^d \delta_D^d(\mathbf{k})\,. \tag{1.59}$$

Writing $\mathbf{k} - \mathbf{k}'$ in place of $\mathbf{k}$ we get

$$\int d^d x e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{x}} \;=\; (2\pi)^d \delta_D^d(\mathbf{k}-\mathbf{k}')\,, \tag{1.60}$$

the *orthogonality relation* of plane waves for the infinite volume. The orthonormal set is thus

$$\left\{ \frac{1}{(2\pi)^{d/2}} e^{i\mathbf{k}\cdot\mathbf{x}} \right\}. \tag{1.61}$$

The *closure relation* is the same except $\mathbf{x}$ and $\mathbf{k}$ change places:

$$\int d^d k e^{i\mathbf{k}\cdot(\mathbf{x}'-\mathbf{x})} \;=\; (2\pi)^d \delta_D^d(\mathbf{x}'-\mathbf{x})\,. \tag{1.62}$$

The *convolution theorem* becomes (**exercise**):

$$\begin{aligned} (f*g)(\mathbf{x}) \;&\equiv\; \int d^d x' f(\mathbf{x}') g(\mathbf{x}-\mathbf{x}') \;=\; \frac{1}{(2\pi)^d} \int d^d k\, f(\mathbf{k}) g(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \\ (f*g)(\mathbf{k}) \;&\equiv\; \int d^d k' f(\mathbf{k}') g(\mathbf{k}-\mathbf{k}') \;=\; (2\pi)^d \int d^d x\, f(\mathbf{x}) g(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}\,, \end{aligned} \tag{1.63}$$

so that the Fourier transform of $(f*g)(\mathbf{x})$ is $f(\mathbf{k})g(\mathbf{k})$ and the Fourier transform of $(f*g)(\mathbf{k})$ is $(2\pi)^d f(\mathbf{x})g(\mathbf{x})$.

The *Plancherel theorem* (**exercise**) is

$$\int d^d x\, f(\mathbf{x}) g(\mathbf{x}) \;=\; \frac{1}{(2\pi)^d} \int d^d k\, f^*(\mathbf{k}) g(\mathbf{k}) \tag{1.64}$$

and the *Parseval theorem* is

$$\int d^d x\, f(\mathbf{x})^2 \;=\; \frac{1}{(2\pi)^d} \int d^d k\, |f(\mathbf{k})|^2\,. \tag{1.65}$$

Even with a finite V we can use the Fourier integral as an approximation. Often it is conceptually simpler to work first with the Fourier series (so that one can, e.g., use the Kronecker delta $\delta_{\mathbf{k}\mathbf{k}'}$ instead of the Dirac delta function $\delta_D^d(\mathbf{k}-\mathbf{k}')$), replacing it with the integral in the end, when it needs to be calculated. The recipe for going from the series to the integral is

$$\begin{aligned} \left(\frac{2\pi}{L}\right)^d \sum_{\mathbf{k}} \quad &\to \quad \int d^d k \\ L^d f_{\mathbf{k}} \quad &\to \quad f(\mathbf{k}) \\ \left(\frac{L}{2\pi}\right)^d \delta_{\mathbf{k}\mathbf{k}'} \quad &\to \quad \delta_D^d(\mathbf{k}-\mathbf{k}') \end{aligned} \tag{1.66}$$

[10] The condition is tighter than this, but the condition that $\int |f(\mathbf{x})| d^d x$ over the infinite volume is finite, assumed sometimes in literature (for 1D), seems too tight, as it is not satisfied by any power law, and yet we transform power laws successfully in Sec. 1.9.

Exercise: CMB lensing. For a small part of the sky we can use the flat-sky approximation, treating it as a 2D plane. As is common in this context, denote the 2D coordinate by $\boldsymbol{\theta}$ and the corresponding 2D wave vector by $\mathbf{l}$ (the letter ell). Gravitational lensing deflects the CMB photons so that a photon originating from $\boldsymbol{\theta}$ is seen coming from $\boldsymbol{\theta}+\nabla\psi(\boldsymbol{\theta})$, where $\psi(\boldsymbol{\theta})$ is the *lensing potential*. Thus the observed ($T_{\rm obs}$) and unlensed (T) CMB temperatures are related

$$T_{\rm obs}(\boldsymbol{\theta}) \;=\; T[\boldsymbol{\theta}+\nabla\psi(\boldsymbol{\theta})] \;\approx\; T(\boldsymbol{\theta})+\nabla\psi(\boldsymbol{\theta})\cdot\nabla T(\boldsymbol{\theta})\,. \tag{1.67}$$

Express $T_{\rm obs}(\mathbf{l})$ in terms of $T(\mathbf{l})$ and $\psi(\mathbf{l})$.

1.6 Power spectrum

We now expand the density perturbation as a Fourier series (assuming a large cubic box $V=L^d$ with periodic boundary conditions)

$$\delta(\mathbf{x}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}\,, \tag{1.68}$$

with

$$\delta_{\mathbf{k}} = \frac{1}{V}\int_V \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d^dx \tag{1.69}$$

and $\delta_{-\mathbf{k}}=\delta^*_{\mathbf{k}}$. Note that

$$\langle\delta(\mathbf{x})\rangle = 0 \quad\Rightarrow\quad \langle\delta_{\mathbf{k}}\rangle = 0\,. \tag{1.70}$$

The Fourier coefficients of the density field $\rho(\mathbf{x})$ and the density perturbation $\delta(\mathbf{x})$ are related by

$$\rho_{\mathbf{k}} = \langle\rho\rangle\delta_{\mathbf{k}} \qquad \text{for } \mathbf{k}\neq 0\,, \tag{1.71}$$

since the $\mathbf{k}\neq 0$ coefficients vanish for the homogeneous part, and

$$\rho_0 = \bar\rho = \langle\rho\rangle(1+\delta_0) = \langle\rho\rangle(1+\bar\delta)\,, \tag{1.72}$$

where

$$\bar\delta = \frac{\bar\rho-\langle\rho\rangle}{\langle\rho\rangle} \tag{1.73}$$

(see Eq. 1.7) is the mean density perturbation within the volume V indicating whether the volume is over- or underdense.

In analogy with the correlation function $\xi(\mathbf{x},\mathbf{x}') = \langle\delta(\mathbf{x})\delta(\mathbf{x}')\rangle$, we may ask what is the corresponding correlation in Fourier space, $\langle\delta^*_{\mathbf{k}}\delta_{\mathbf{k}'}\rangle$. Note that due to the mathematics of complex numbers, correlations of Fourier coefficients are defined with the complex conjugate *. This way the correlation of $\delta_{\mathbf{k}}$ with itself, $\langle\delta^*_{\mathbf{k}}\delta_{\mathbf{k}}\rangle = \langle|\delta_{\mathbf{k}}|^2\rangle$ is a real (and nonnegative) quantity, the expectation value of the absolute value (modulus) of $\delta_{\mathbf{k}}$ squared, i.e., the variance of $\delta_{\mathbf{k}}$. Calculating

$$\begin{aligned}
\langle\delta^*_{\mathbf{k}}\delta_{\mathbf{k}'}\rangle &= \frac{1}{V^2}\int d^dx e^{i\mathbf{k}\cdot\mathbf{x}}\int d^dx' e^{-i\mathbf{k}'\cdot\mathbf{x}'}\langle\delta(\mathbf{x})\delta(\mathbf{x}')\rangle \\
&= \frac{1}{V^2}\int d^dx e^{i\mathbf{k}\cdot\mathbf{x}}\int d^dr e^{-i\mathbf{k}'\cdot(\mathbf{x}+\mathbf{r})}\langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle \\
&= \frac{1}{V^2}\int d^dr e^{-i\mathbf{k}'\cdot\mathbf{r}}\xi(\mathbf{r})\int d^dx e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} \\
&= \frac{1}{V}\delta_{\mathbf{k}\mathbf{k}'}\int d^dr e^{-i\mathbf{k}\cdot\mathbf{r}}\xi(\mathbf{r}) \;\equiv\; \frac{1}{V}\delta_{\mathbf{k}\mathbf{k}'}P(\mathbf{k})\,,
\end{aligned} \tag{1.74}$$

where we used $\langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle = \xi(\mathbf{r})$, which results from statistical homogeneity. Note that here $\delta_{\mathbf{k}\mathbf{k}'}$ is the Kronecker delta, not a density perturbation! Thus, *from statistical homogeneity follows that the Fourier coefficients $\delta_{\mathbf{k}}$ are uncorrelated*. The quantity

$$P(\mathbf{k}) \;\equiv\; V\langle|\delta_{\mathbf{k}}|^2\rangle \;=\; \int d^dr\, e^{-i\mathbf{k}\cdot\mathbf{r}}\xi(\mathbf{r})\,, \tag{1.75}$$

which gives the *variance* of $\delta_{\mathbf{k}}$, is called the *power spectrum* of $\delta(\mathbf{x})$. For a finite volume V, the values of $\xi(\mathbf{r})$ for large $\mathbf{r}$ are missed in (1.75). So we define the true, theoretical, power spectrum $P(\mathbf{k})$ to be the limit of (1.75) as $V\to\infty$. In this limit also $\bar\rho = \langle\rho\rangle$, from the ergodicity assumption. Thus the power spectrum and correlation function form a *d-dimensional Fourier transform pair*, so that

$$\xi(\mathbf{r}) \;=\; \frac{1}{(2\pi)^d}\int d^dk\, e^{i\mathbf{k}\cdot\mathbf{r}}P(\mathbf{k})\,. \tag{1.76}$$

(The correlation function $\to 0$ for large separations, so Fourier transform is appropriate for it.) Unlike the correlation function, the power spectrum $P(\mathbf{k})$ is positive everywhere. We discuss estimation of $P(\mathbf{k})$ from a finite survey in Sec. 9.

The correlation function is a dimensionless quantity, whereas the power spectrum $P(\mathbf{k})$ has the dimension of volume ($\delta(\mathbf{x})$ and $\delta_{\mathbf{k}}$ are dimensionless). We noted earlier that the magnitude of Fourier coefficients depends on the volume V. From (1.75) we see that the typical magnitude of $\delta_{\mathbf{k}}$ goes down with volume as $\propto\sqrt{V}$. Although the density of $\mathbf{k}$-modes increases $\propto V$, neighboring $\delta_{\mathbf{k}}$ are uncorrelated, so they add up incoherently, so that, e.g., 4 times as many $\mathbf{k}$ modes bring only a factor of 2 increase in $\sum_{\mathbf{k}}\delta_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}}$, to be compensated by the $\delta_{\mathbf{k}}$ being a factor of 2 smaller.

From statistical isotropy

$$\xi(\mathbf{r}) = \xi(r) \quad\Rightarrow\quad P(\mathbf{k}) = P(k) \tag{1.77}$$

(the Fourier transform of a spherically symmetric function is also spherically symmetric), so that the variance of $\delta_{\mathbf{k}}$ depends only on the magnitude k of the wave vector $\mathbf{k}$, i.e., on the corresponding distance scale. Since small distance scales correspond to large k and vice versa, to avoid confusion it is better to use the words *high* and *low* instead of "large" and "small" for k, i.e., small scales correspond to high k, and large scales to low k.

Using the recipe (1.66) for going from Fourier coefficients to Fourier transform, (1.74) gives

$$\boxed{\langle\delta(\mathbf{k})^*\delta(\mathbf{k}')\rangle \;\equiv\; (2\pi)^d\delta_D^d(\mathbf{k}-\mathbf{k}')P(\mathbf{k})\,.} \tag{1.78}$$

Notice that with $\delta_{\mathbf{k}}$ we can write $P(\mathbf{k})\equiv V\langle\delta_{\mathbf{k}}^*\delta_{\mathbf{k}}\rangle$ (without having to use $\delta_{\mathbf{k}\mathbf{k}'}$ in the equation), but with $\delta(\mathbf{k})$ we need to use the δ_D-function in the definition of $P(\mathbf{k})$.

The correlation function is more closely connected to observations, whereas theoretical predictions come more naturally in terms of $P(k)$, especially at large distance scales, where the density perturbations are small and closer to their primordial state. In principle, when we have determined one of $\xi(r)$ and $P(k)$ from observations, we get the other by Fourier transform. In practice, observational errors make this inaccurate, and it is better to determine each one separately with a method optimized for it. Especially for large separations, where $\xi(r)$ is small, it is difficult to determine it accurately, if at all. For these reasons, density perturbations at large distance scales (low k) are more commonly discussed in terms of $P(k)$ and for small distance scales (small r) in terms of $\xi(r)$.

For the density variance we get

$$\begin{aligned}\langle\delta^2\rangle\equiv\xi(0) \;&=\; \frac{1}{(2\pi)^d}\int d^dk\,P(k) \;=\; \frac{C_d}{(2\pi)^d}\int_0^\infty P(k)k^{d-1}dk \\ &=\; \frac{C_d}{(2\pi)^2}\int_0^\infty k^dP(k)\frac{dk}{k} \;\equiv\; \int_{-\infty}^\infty \mathcal{P}(k)d\ln k\,.\end{aligned} \tag{1.79}$$

where we have defined

$$\mathcal{P}(k) \equiv \frac{C_d k^d}{(2\pi)^d} P(k) \;=\; \frac{k}{\pi}P(k)\,,\; \frac{k^2}{2\pi}P(k)\,,\; \frac{k^3}{2\pi^2}P(k) \quad \text{for } d=1,2,3\,. \tag{1.80}$$

Another common notation for $\mathcal{P}(k)$ is $\Delta^2(k)$.[11] The word "power spectrum" is used to refer to both $P(k)$ and $\mathcal{P}(k)$. Of these two, $\mathcal{P}(k)$ has the more obvious physical meaning: it gives the contribution of a logarithmic interval of scales, i.e., from k to ek, to the density variance. $\mathcal{P}(k)$ is dimensionless, whereas $P(k)$ has the dimension of (d-dimensional) volume.

See Fig. 3 for the observed power spectrum from the Sloan Digital Sky Survey.

The pair of (1.79) is

$$P(0) \;=\; \lim_{k\to 0} P(k) \;=\; \int d^d r\, \xi(\mathbf{r})\,. \tag{1.81}$$

If[12] $P(k)\to 0$ as $k\to 0$ we get the *integral constraint*

$$\int d^d r\, \xi(\mathbf{r}) \;=\; 0\,. \tag{1.82}$$

Therefore $\xi(r)$ must become negative for some r, so that at such a separation from an overdense region we are more likely to find an underdense region. (Going to ever larger separations, ξ as a function of r may oscillate around zero, the oscillation becoming ever smaller in amplitude. Most of the interest in $\xi(r)$ is for the smaller r within the initial positive region.) The present understanding is that indeed $P(k)\to 0$ as $k\to 0$, meaning that structure gets weaker at larger scales at a sufficient rate. Of course, we have information only about a finite part of the universe, so this represents an extrapolation from the observable universe. In principle, there could be weird behavior outside the horizon, but since the observable universe does not contain evidence of such, it is fine to work with such an extrapolation.

For isotropic $\xi(r)$ and $P(k)$ we can switch to polar, spherical, etc. coordinates and do the angular integrals to rewrite (1.75) and (1.76) as 1-dimensional integrals (**exercise**):

$$\begin{aligned}
P(k) &= \int_0^\infty \xi(r)\cos kr\, 2dr \qquad \text{or} \quad \mathcal{P}(k) \;=\; \frac{2k}{\pi}\int_0^\infty \xi(r)\cos kr dr \qquad \text{(1D)}\\
P(k) &= \int_0^\infty \xi(r) J_0(kr)\, 2\pi r dr \quad \text{or} \quad \mathcal{P}(k) \;=\; k^2\int_0^\infty \xi(r) J_0(kr) r dr \qquad \text{(2D)}\\
P(k) &= \int_0^\infty \xi(r)\frac{\sin kr}{kr} 4\pi r^2 dr \quad \text{or} \quad \mathcal{P}(k) \;=\; \frac{2k^3}{\pi}\int_0^\infty \xi(r)\frac{\sin kr}{kr} r^2 dr \quad \text{(3D)}
\end{aligned} \tag{1.83}$$

and

$$\begin{aligned}
\xi(r) &= \frac{1}{\pi}\int_0^\infty P(k)\cos kr dk \qquad\qquad = \int_0^\infty \mathcal{P}(k)\cos kr\frac{dk}{k} \qquad \text{(1D)}\\
\xi(r) &= \frac{1}{2\pi}\int_0^\infty P(k) J_0(kr) k dk \qquad\qquad = \int_0^\infty \mathcal{P}(k) J_0(kr)\frac{dk}{k} \qquad \text{(2D)}\\
\xi(r) &= \frac{1}{(2\pi)^3}\int_0^\infty P(k)\frac{\sin kr}{kr} 4\pi k^2 dk \;=\; \int_0^\infty \mathcal{P}(k)\frac{\sin kr}{kr}\frac{dk}{k} \qquad \text{(3D)}\,,
\end{aligned} \tag{1.84}$$

where J_0 is a Bessel function and the $\sin kr/kr = j_0(kr)$ is a spherical Bessel function. In 3D, for the volume-averaged $\bar\xi(R)$ defined in (1.15) we get (**exercise**):

$$\bar\xi(R) \;=\; \int_0^\infty \mathcal{P}(k)\left[\frac{3(\sin kR - kR\cos kR)}{(kR)^3}\right]\frac{dk}{k} \;=\; \int_0^\infty \mathcal{P}(k)\left[\frac{3}{kR} j_1(kR)\right]\frac{dk}{k}\,. \tag{1.85}$$

[11] The notation $\mathcal{P}(k)$ and calling it "power spectrum" is common among cosmologists. Astronomers seem to use the notation $\Delta^2(k)$ for it, and reserve the word "power spectrum" for $P(k)$.

[12] From (1.75), $P(0) = V\langle(\delta_0)^2\rangle$, where $\delta_0 = \bar\delta = (\bar\rho - \langle\rho\rangle)/\langle\rho\rangle$. While $\langle\bar\delta\rangle = 0$, $\langle\bar\delta^2\rangle = (\langle\bar\rho^2\rangle - \langle\rho\rangle^2)/\langle\rho\rangle^2 = \xi_V \neq 0$ for a finite V.

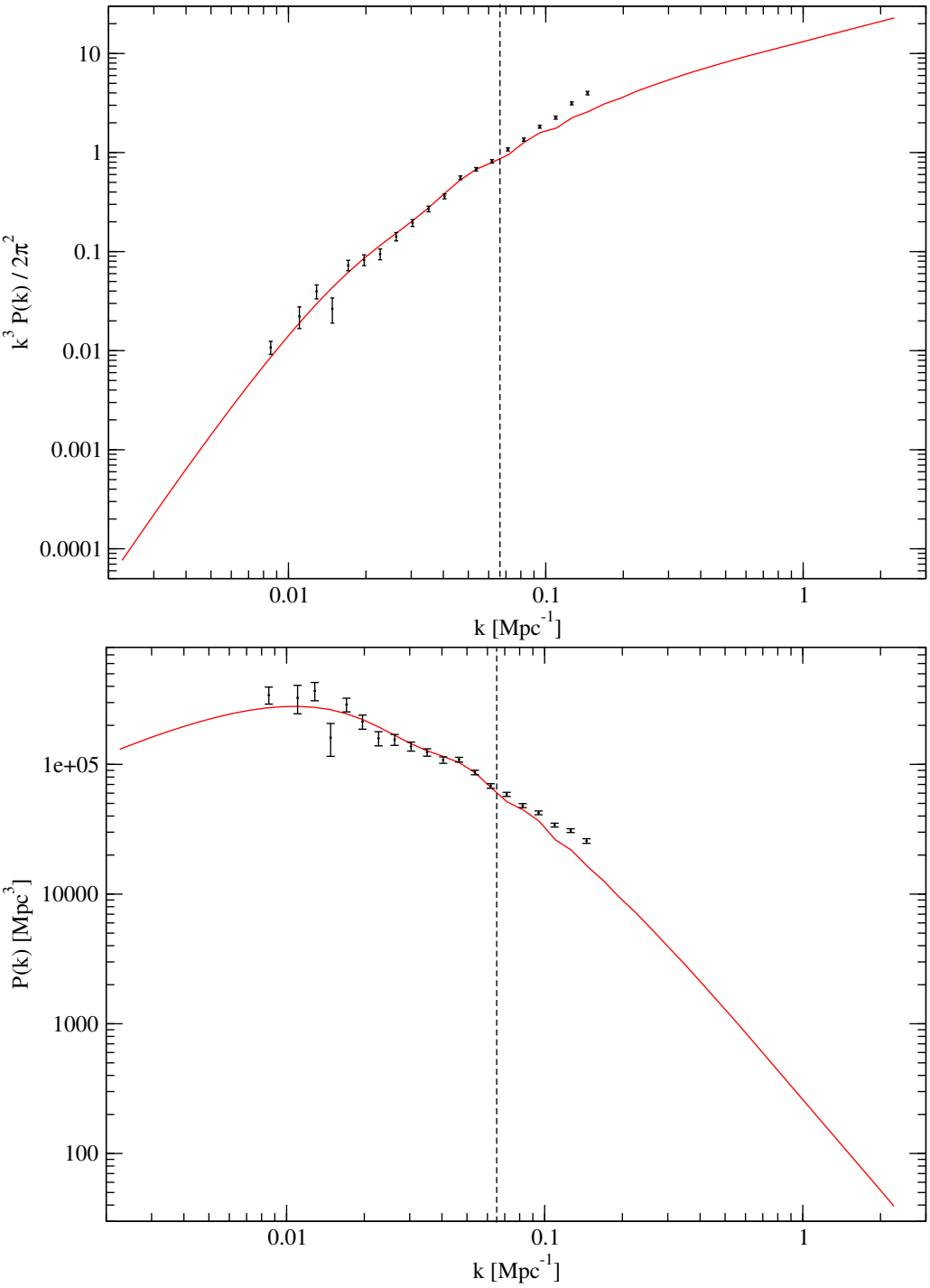


Figure 3: The matter power spectrum from the SDSS obtained using luminous red galaxies [13]. The top figure shows $\mathcal{P}(k)$ and the bottom figure $P(k)$. A Hubble constant value $H_0 = 71.4$ km/s/Mpc has been assumed for this figure. (These galaxy surveys only obtain the scales up to the Hubble constant, and therefore the observed $P(k)$ is usually shown in units of h Mpc^{-1} for k and h^{-3} Mpc^3 for $P(k)$, so that no value for H_0 need to be assumed.) The black bars are the observations and the red curve is a theoretical fit, from linear perturbation theory, to the data. The bend in $P(k)$ at $k_{\rm eq} \sim 0.01$ Mpc^{-1} is clearly visible in the bottom figure. Linear perturbation theory fails when $\mathcal{P}(k) \gtrsim 1$, and therefore the data points do not follow the theoretical curve to the right of the dashed line (representing an estimate on how far linear theory can be trusted). Figure by R. Keskitalo.

"The factor in brackets dies off faster with increasing k than the $(\sin kr/kr)$ in (1.84), so $\bar{\xi}(R)$ gives a cleaner measure of the power spectrum at $k \sim 1/R$ than does $\xi(R)$." (MBW[2], p. 263.) (This is a comparison between $j_0(x)$ and $(3/x)j_1(x)$, where the $j_n(x)$ are spherical Bessel functions. These are elementary functions, so we do not discuss them more at this stage, but we will meet them later.)

Example: We do the 2D case of (1.83), since it involves the non-elementary function J_0:

$$P(k) \;=\; \int d^2r\, e^{-i\mathbf{k}\cdot\mathbf{r}}\xi(r) \;=\; \int_0^\infty r dr\, \xi(r) \int_0^{2\pi} d\varphi\, e^{-ikr\cos\varphi}\,, \tag{1.86}$$

where the angular integral gives

$$\int_0^{2\pi} d\varphi\, e^{-ikr\cos\varphi} \;=\; 2\int_0^{\pi} d\varphi\, \cos(kr\cos\varphi) \;=\; 2\pi J_0(kr)\,, \tag{1.87}$$

where we used the integral representation of the Bessel function ([8], p. 680)

$$J_0(x) \;=\; \frac{2}{\pi}\int_0^{\pi/2} \cos(x\sin\varphi)d\varphi \;=\; \frac{2}{\pi}\int_0^{\pi/2} \cos(x\cos\varphi)d\varphi\,. \tag{1.88}$$

These two integrals are equal since $\cos\varphi$ and $\sin\varphi$ go over the same values at the same rate. Since (the outer) cos is an even function, the integrals give the same result over each quadrant, i.e, we could as well integrate from $\pi/2$ to π.

1.6.1 Power spectrum from a finite survey

Continuing from Sec. 1.3, we define

$$\widehat{\delta}_{\mathbf{k}} \equiv \frac{1}{V}\int_V \widehat{\delta}(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d^dx\,, \tag{1.89}$$

where

$$\widehat{\delta}(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}}$$

according to (1.36). The constant term affects only the $\mathbf{k} = 0$ coefficient and we easily see that

$$\widehat{\delta}_0 = 0 \tag{1.90}$$

and for $\mathbf{k} \neq 0$

$$\widehat{\delta}_{\mathbf{k}} = \frac{\rho_{\mathbf{k}}}{\bar{\rho}} = \frac{\langle\rho\rangle}{\bar{\rho}}\delta_{\mathbf{k}}\,. \tag{1.91}$$

Defining an "estimate for the power spectrum from the survey volume V" (real power spectrum estimation is discussed in Sec. 9),

$$\widehat{P}(\mathbf{k}) \equiv V|\widehat{\delta}_{\mathbf{k}}|^2\,, \tag{1.92}$$

we have

$$\widehat{P}(0) = 0 \tag{1.93}$$

and for $\mathbf{k} \neq 0$

$$\widehat{P}(\mathbf{k}) \;=\; V\left(\frac{\langle\rho\rangle}{\bar{\rho}}\right)^2 |\delta_{\mathbf{k}}|^2 \quad\Rightarrow\quad \langle\widehat{P}(\mathbf{k})\rangle \;=\; \left(\frac{\langle\rho\rangle}{\bar{\rho}}\right)^2 P(\mathbf{k})\,. \tag{1.94}$$

Thus the finite volume biases power spectrum estimation just by a constant factor, instead of the more complicated integral constraint effect for correlation function estimation.

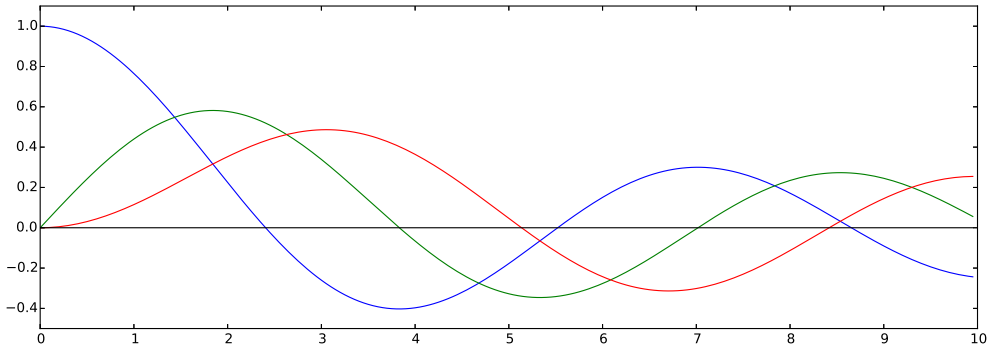


Figure 4: The first three Bessel functions: J_0 (blue), J_1 (green), and J_2 (red).

Exercise: Fourier expand the "observed" $\widehat{\delta}(\mathbf{x})$ of Eq. (1.36) to get its Fourier coefficients $\widehat{\delta}_\mathbf{k}$. Show that

$$\widehat{\xi}(\mathbf{r}) \;=\; \frac{V}{(2\pi)^d}\int d^dk\,|\widehat{\delta}_\mathbf{k}|^2 e^{i\mathbf{k}\cdot\mathbf{r}}\,, \tag{1.95}$$

for this single realization. Note that here we do not need any statistical assumptions (like statistical homogeneity or ergodicity). Contrast this result with (1.76).

1.7 Bessel functions

The 2D Fourier transform brings in Bessel functions $J_n(x)$, with $n = 0, 1, 2, \ldots$ They are mostly used on the positive real axis, where they are oscillating functions, whose amplitude decreases with increasing x, asymptotically as $x^{-1/2}$. See Fig. 4. We list some of their properties.

$$J_0(0) = 1 \qquad \text{and} \qquad J_n(0) = 0 \quad \text{for } n = 1, 2, \ldots \tag{1.96}$$

Their power series begins

$$J_n(x) = \frac{x^n}{2^n n!} - \frac{x^{n+2}}{2^{n+2}(n+1)!} + \ldots \tag{1.97}$$

They have the *integral representations*

$$J_n(x) \;=\; \frac{1}{\pi}\int_0^\pi \cos(n\varphi - x\sin\varphi)\,d\varphi \;=\; \frac{(-i)^n}{\pi}\int_0^\pi e^{ix\cos\varphi}\cos n\varphi\,d\varphi\,. \tag{1.98}$$

Both integrals are even in φ, and periodic over 2π, so one can replace

$$\frac{1}{\pi}\int_0^\pi \;=\; \frac{1}{2\pi}\int_{-\pi}^\pi \;=\; \frac{1}{2\pi}\int_a^{a+2\pi}\,. \tag{1.99}$$

A number of *recursion formulae* relate them and their derivatives to each other:

$$\begin{aligned} J_{n-1}(x) + J_{n+1}(x) &= \frac{2n}{x}J_n(x) \\ J_{n-1}(x) - J_{n+1}(x) &= 2J_n'(x) \\ J_{n-1}(x) &= \frac{n}{x}J_n(x) + J_n'(x) \\ J_{n+1}(x) &= \frac{n}{x}J_n(x) - J_n'(x)\,. \end{aligned} \tag{1.100}$$

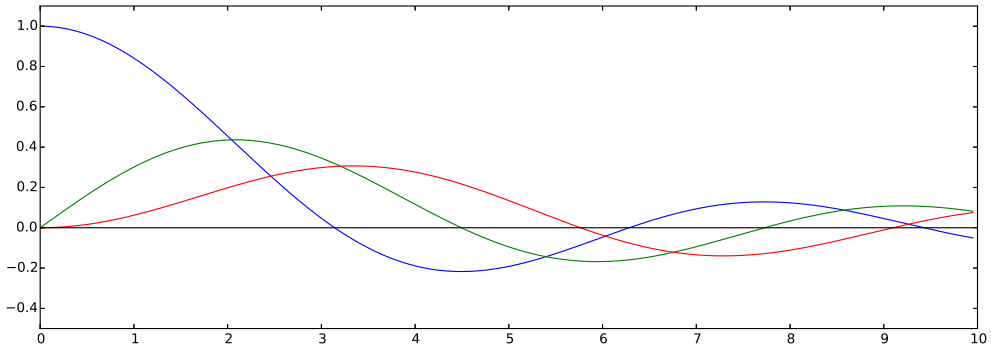


Figure 5: The first three spherical Bessel functions: j_0 (blue), j_1 (green), and j_2 (red).

Spherical Bessel functions
$j_0(x) = \dfrac{\sin x}{x}$
$j_1(x) = \dfrac{\sin x}{x^2} - \dfrac{\cos x}{x}$
$j_2(x) = \left(\dfrac{3}{x^3} - \dfrac{1}{x}\right)\sin x - \dfrac{3}{x^2}\cos x$

Table 1: Spherical Bessel functions.

As a special case of (1.100b),

$$J_0'(x) \;=\; -J_1(x)\,. \tag{1.101}$$

The Bessel function *closure relation* applies to Bessel functions with the same n but different wavelengths:

$$\int_0^\infty J_n(\alpha x)J_n(\alpha' x)x\,dx \;=\; \frac{1}{\alpha}\delta_D(\alpha-\alpha')\,. \tag{1.102}$$

A somewhat similar (note dx instead of $x\,dx$) integral with neighboring (n and $n-1$) Bessel functions gives (Gradshteyn&Ryzhik[9] 6.512.3)

$$\int_0^\infty J_n(\alpha x)J_{n-1}(\beta x)\,dx \;=\; \frac{\beta^{n-1}}{\alpha^n}\Theta(\alpha-\beta) \;=\; \begin{cases} 0 & (\alpha<\beta) \\ \frac{1}{2\alpha} & (\alpha=\beta) \\ \frac{\beta^{n-1}}{\alpha^n} & (\alpha>\beta)\,, \end{cases} \tag{1.103}$$

1.8 Spherical Bessel functions

The spherical Bessel functions $j_n(x)$ (of integer order) are related to ordinary Bessel functions of half-integer order:

$$j_n(x) = \sqrt{\frac{\pi}{2x}}J_{n+1/2}(x)\,. \tag{1.104}$$

Like J_n, they are mostly used on the positive real axis, where they are oscillating functions; their amplitude decreases faster, asymptotically as x^{-1}. See Fig. 5. Unlike J_n, they are elementary functions (for integer n), see Table 1.

We list some of their properties.

$$j_0(0) = 1 \qquad \text{and} \qquad j_n(0) = 0 \quad \text{for } n = 1, 2, \ldots \tag{1.105}$$

Their power series begins

$$j_n(x) = \frac{2^n n!\, x^n}{(2n+1)!} - \frac{2^n (n+1)!\, x^{n+2}}{(2n+3)!} + \dots \tag{1.106}$$

They have *recursion formulae* relating them and their derivatives to each other:

$$\begin{aligned} j_{n-1}(x) + j_{n+1}(x) &= \frac{2n+1}{x}\, j_n(x) \\ n j_{n-1}(x) - (n+1) j_{n+1}(x) &= (2n+1)\, j_n'(x)\,. \end{aligned} \tag{1.107}$$

As a special case of (1.107b),

$$j_0'(x) = -j_1(x)\,. \tag{1.108}$$

1.9 Power-law spectra

For certain ranges of scales, $\xi(r)$ and $P(k)$ can be approximated by a power-law form,

$$\xi(r) \propto r^{-\gamma} \qquad \text{or} \qquad P(k) \propto k^n\,. \tag{1.109}$$

Note the minus sign for ξ – we expect correlations to decrease with increasing separation, so this makes γ positive. When plotted on a log-log scale, such functions appear as straight lines with slope $-\gamma$ and n:

$$\log \xi = -\gamma \log r + const \qquad \text{and} \qquad \log P = n \log k + const\,. \tag{1.110}$$

The proportionality constant can be given in terms of a reference scale. For $\xi(r)$ we usually choose the scale r_0 where $\xi(r_0) = 1$, so that

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}\,. \tag{1.111}$$

See Fig. 6. For $P(k)$ we may write

$$P(k) = A^2 \left(\frac{k}{k_p}\right)^n \qquad \text{or} \qquad \mathcal{P}(k) = A^2 \left(\frac{k}{k_p}\right)^{n+d}\,, \tag{1.112}$$

where k_p is called a *pivot scale* (whose choice depends on the application) and $A \equiv \sqrt{P(k_p)}$ or $\sqrt{\mathcal{P}(k_p)}$ is the amplitude of the power spectrum at the pivot scale.[13]

We define the spectral index $n(k)$ as

$$n(k) \equiv \frac{d \ln P}{d \ln k}\,. \tag{1.113}$$

It gives the slope of $P(k)$ on a log-log plot. For a power-law $P(k)$, $n(k) = const = n$. We can study power-law $\xi(r)$ and $P(k)$ as a playground to get a feeling what different values of the spectral index mean, and, e.g., how γ and n are related.

The Fourier transform of a power law is a power law. For the correlation function of (1.111) we get (**exercise**)

$$\begin{aligned} \text{1D:} \quad P(k) &= \frac{2}{k}\Gamma(1-\gamma)\sin(\tfrac{1}{2}\gamma\pi)(kr_0)^\gamma && (0<\gamma<1 \quad \Rightarrow \quad -1<n<0) \\ \text{2D:} \quad P(k) &= \frac{2\pi}{k^2} 2^{1-\gamma} \frac{\Gamma(1-\frac{1}{2}\gamma)}{\Gamma(\frac{1}{2}\gamma)} (kr_0)^\gamma && (\tfrac{1}{2}<\gamma<2 \quad \Rightarrow \quad -\tfrac{3}{2}<n<0) \\ \text{3D:} \quad P(k) &= \frac{4\pi}{k^3}\Gamma(2-\gamma)\sin\frac{(2-\gamma)\pi}{2}\,(kr_0)^\gamma && (1<\gamma<3 \quad \Rightarrow \quad -2<n<0)\,, \end{aligned} \tag{1.114}$$

[13] The reason we do not follow the example of (1.111) to define a k_0 where the power spectrum would be 1 is that $P(k)$ is not dimensionless, and that $\mathcal{P}(k)$ is usually discussed at scales and times when $\mathcal{P}(k) \ll 1$, so that a k_0 such that $\mathcal{P}(k_0) = 1$ would be outside the region of interest and/or the power-law approximation.

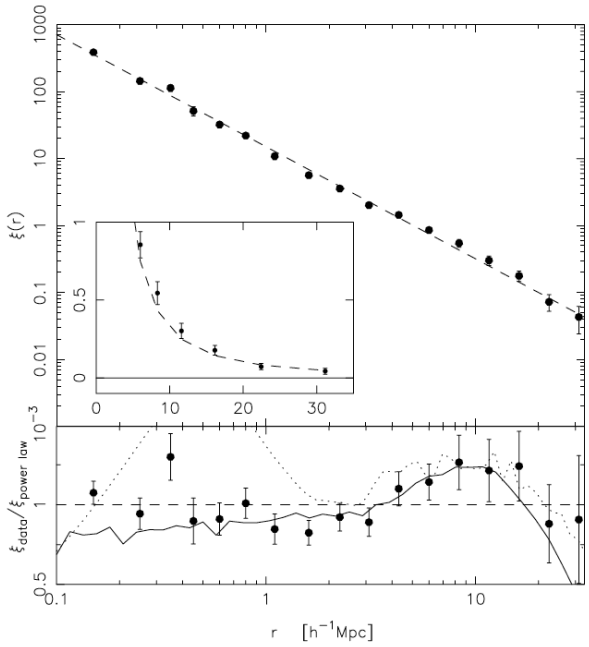


Figure 6: Top panel: The correlation function from the 2dFGRS galaxy survey in log-log scale. The dashed line is the best-fit power law ($r_0 = 5.05\,h^{-1}$Mpc, $\gamma = 1.67$). The inset shows the same in linear scale. Bottom panel: 2dFGRS data (solid circles with error bars) divided by the power-law fit. The solid line is the result from the APM survey and the dashed line from an N-body simulation. This is Fig. 11 from [10].

so that γ and n are related by

$$\boxed{n \;=\; \gamma - d}\,. \tag{1.115}$$

For $\mathcal{P}(k)$ these read

$$\begin{aligned}
\text{1D:}\quad \mathcal{P}(k) &= \frac{2}{\pi}\Gamma(1-\gamma)\sin(\tfrac{1}{2}\gamma\pi)(kr_0)^\gamma \\
\text{2D:}\quad \mathcal{P}(k) &= 2^{1-\gamma}\frac{\Gamma(1-\tfrac{1}{2}\gamma)}{\Gamma(\tfrac{1}{2}\gamma)}(kr_0)^\gamma \\
\text{3D:}\quad \mathcal{P}(k) &= \frac{2}{\pi}\Gamma(2-\gamma)\sin\frac{(2-\gamma)\pi}{2}\,(kr_0)^\gamma\,,
\end{aligned} \tag{1.116}$$

In the 3D case these expressions are undefined for $\gamma = 2$, and we have the simpler result

$$P(k) \;=\; \frac{2\pi^2}{k^3}(kr_0)^2 \qquad \text{and} \qquad \mathcal{P}(k) \;=\; (kr_0)^2\,. \tag{1.117}$$

Observationally, the 3D correlation function has $\gamma \approx 1.8$ for small separations, corresponding to $n \approx -1.2$ for high k. See Fig. 3 for the observed power spectrum from the Sloan Digital Sky Survey.

Note that a power-law correlation function is everywhere positive. This is possible when $\lim_{k\to 0} P(k) \neq 0$, which is indeed the case for the allowed spectral indices, $n < 0$, above. (In this case, there is sufficient structure at ever larger scales to maintain positive correlation at ever larger distances.) In reality, the power spectrum bends at large scales so that its spectral index becomes positive for low k, and therefore also the correlation function changes shape and will have also negative values at large enough r.

Outside these values of spectral indices the Fourier transform integrals diverge in the small- or large-scale limit (I guess); but this does not prevent $\xi(r)$ or $P(k)$ from having also such power-law forms over some limited range of scales.

The variance

$$\langle\delta^2\rangle = \xi(0) = \int_0^\infty \mathcal{P}(k)\frac{dk}{k} \propto \int_0^\infty k^{n+d-1}dk = \frac{1}{n+d}\left[k^{n+d}\right]_0^\infty \qquad \text{for} \quad n \neq -d \qquad (1.118)$$

diverges at large scales (low k) for $n \leq -d$ and at small scales (high k) for $n \geq -d$. Thus we should have $n > -d$ for $k \to 0$ and $n < -d$ for $k \to \infty$. The large scales are not an issue (since indeed $n > -d$) in cosmology. At small scales, (1.111) forces $\xi(r) \to \infty$ as $r \to 0$ for any positive γ. The solution of this issue is that arbitrarily small scales are not relevant; in practice we have a finite resolution that cuts off the smallest scales. This can be implemented with window functions, discussed in Sec. 1.10.

The case $n = -d$ ($\gamma = 0$) is a *scale-invariant* spectrum, $\mathcal{P}(k) = const$. Such a spectrum would mean that the universe would appear equally inhomogeneous at arbitrarily large scales – no asymptotic homogeneity. Note that here we discuss the spectrum of density. References to a scale-invariant or nearly scale-invariant spectrum in cosmology refer usually to the spectrum of gravitational potential (Newtonian treatment) or spacetime curvature perturbations (GR treatment). Their spectral index is lower by 4 so that such a scale-invariant spectrum will have a density spectral index $n = 1$ (in 3D).

The boundary case $n = 0$ has the same $\langle|\delta_\mathbf{k}|^2\rangle$ at all scales. From (1.115) $\gamma \to d$ as $n \to 0$. For $\gamma < d$, the integration from 0 to R in (1.14) for $\bar{\xi}(R)$ smooths over the small-scale divergence of $\xi(r)$; but for $\gamma \geq d$ the integral (1.14) diverges. Actually, also (1.114c) and (1.116c) diverge ($\Gamma(-1) = \infty$) so that a finite ξ would give infinite $P(k)$. Instead, $n = 0$ corresponds to the case of no correlations, $\xi = 0$. This is called *white noise* or a *Poisson distribution* (to be discussed in Sec. 2.3).

The larger n is, the more is the structure concentrated at small scales. Peacock comments that $n \geq 0$ spectra would seem to indicate that any large-scale structure is 'accidental', "reflecting the low-k Fourier coefficients of some small-scale process", whereas $n < 0$ means that large-scale structure is 'real' " ([1], p. 499).

Example: To do the 1D and 3D cases in (1.114) is standard FYMM I stuff, but for the 2D case I had to resort to integral tables:

$$P(k) \;=\; \int_0^\infty \xi(r)J_0(kr)\,2\pi r dr \;=\; 2\pi r_0^\gamma \int_0^\infty r^{1-\gamma}J_0(kr)dr \;=\; \frac{2\pi(kr_0)^\gamma}{k^2}\int_0^\infty x^{1-\gamma}J_0(x)dx\,. \qquad (1.119)$$

From Gradshteyn & Ryzhik ([9], formula 6.561.14) we find that

$$\int_0^\infty x^\mu J_\nu(x)dx \;=\; 2^\mu\frac{\Gamma(\frac{1}{2}+\frac{1}{2}\nu+\frac{1}{2}\mu)}{\Gamma(\frac{1}{2}+\frac{1}{2}\nu-\frac{1}{2}\mu)} \qquad \text{for } -\nu-1 < \mu < \tfrac{1}{2}\,. \qquad (1.120)$$

We have $\nu = 0$ and $\mu = 1-\gamma$, so the condition becomes $\frac{1}{2} < \gamma < 2$ and the result

$$\int_0^\infty x^{1-\gamma}J_0(x)dx \;=\; 2^{1-\gamma}\frac{\Gamma(1-\frac{1}{2}\gamma)}{\Gamma(\frac{1}{2}\gamma)}\,. \qquad (1.121)$$

1.10 Scales of interest and window functions

In (1.79) we integrated over all scales, from the infinitely large ($k = 0$ and $\ln k = -\infty$) to the infinitely small ($k = \infty$ and $\ln k = \infty$) to get the density variance. Perhaps this is not really

what we want. The average matter density today is $3 \times 10^{-27}\,\mathrm{kg/m^3}$. The density of the Earth is $5.5 \times 10^3\,\mathrm{kg/m^3}$ and that of an atomic nucleus $2 \times 10^{17}\,\mathrm{kg/m^3}$, corresponding to $\delta \approx 2 \times 10^{30}$ and $\delta \approx 10^{44}$. Probing the density of the universe at such small scales finds a huge variance in it, but this is no longer the topic of cosmology – we are not interested here in planetary science or nuclear physics.

Even the study of the structure of individual galaxies is not considered to belong to cosmology, so the smallest (comoving) scale of cosmological interest, at least when we discuss the present universe, is that of a typical separation between neighboring galaxies, of the order 1 Mpc.

To exclude scales smaller than R ($r < R$ or $k > R^{-1}$) we can *filter* the density field with a *window function* (sometimes called a *filter function*). This can be done in $\mathbf{k}$-space or $\mathbf{x}$-space.

The filtering in $\mathbf{x}$-space is done by convolution. We introduce a (usually spherically symmetric) window function $W(\mathbf{r}; R)$ such that

$$\int d^d r\, W(\mathbf{r}; R) = 1 \tag{1.122}$$

(normalization) and $W \sim 0$ for $|\mathbf{r}| \gg R$ and define the filtered density field

$$\delta(\mathbf{x}; R) \;\equiv\; (\delta * W)(\mathbf{x}) \;\equiv\; \int d^d \mathbf{x}'\, \delta(\mathbf{x}') W(\mathbf{x} - \mathbf{x}')\,. \tag{1.123}$$

Here $\delta(\mathbf{x}; R)$ and $W(\mathbf{x}; R)$ are considered as functions of $\mathbf{x}$ and R denotes the chosen resolution. To simplify notation, we write hereafter $W(\mathbf{x}; R)$ as $W(\mathbf{x})$, leaving the scale R implicit. We now also assume W is spherically symmetric, so we can write just $W(r)$.

Denote the Fourier coefficients of $\delta(\mathbf{x}; R)$ by $\delta_{\mathbf{k}}(R)$. We use the Fourier series for $\delta(\mathbf{x})$ and $\delta(\mathbf{x}; R)$, but since $W(r)$ vanishes for large r we can use the Fourier transform $W(k)$ for it. Thus we need a mixed form of the convolution theorem. Let's do it explicitly:

$$\begin{aligned} \delta_{\mathbf{k}}(R) &= \frac{1}{V}\int_V d^d x\, \delta(\mathbf{x}; R) e^{-i\mathbf{k}\cdot\mathbf{x}} \;=\; \frac{1}{V}\int_V d^d x d^d x'\, \delta(\mathbf{x}') W(\mathbf{x} - \mathbf{x}') e^{-i\mathbf{k}\cdot\mathbf{x}} \\ &= \frac{1}{V}\int_V d^d x'\, \delta(\mathbf{x}') e^{-i\mathbf{k}\cdot\mathbf{x}'} \int d^d r\, W(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} \;=\; W(\mathbf{k})\delta_{\mathbf{k}}\,, \end{aligned} \tag{1.124}$$

where

$$W(\mathbf{k}) = \int d^d r\, W(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} \tag{1.125}$$

is the Fourier transform of $W(\mathbf{r})$.[14] With our normalization, $W(\mathbf{r})$ has dimension $1/V$ and $W(\mathbf{k})$ is dimensionless with $W(\mathbf{k} = 0) = 1$. Since $W(\mathbf{r}) = W(r)$ is spherically symmetric, so is $W(\mathbf{k}) = W(k)$. Since $W(-\mathbf{r}) = W(\mathbf{r})$, $W(\mathbf{k})$ is real.

For the correlations of these filtered Fourier coefficients we get

$$\langle \delta^*_{\mathbf{k}}(R) \delta_{\mathbf{k}'}(R) \rangle \;=\; W(\mathbf{k})^* W(\mathbf{k}') \langle \delta^*_{\mathbf{k}} \delta_{\mathbf{k}'} \rangle \;=\; \frac{1}{V} \delta_{\mathbf{k}\mathbf{k}'} W(k)^2 P(k) \tag{1.126}$$

so the filtered power spectra are

$$W(k)^2 P(k) \qquad \text{and} \qquad W(k)^2 \mathcal{P}(k)\,. \tag{1.127}$$

The filtered correlation function is

$$\xi(\mathbf{r}; R) \;\equiv\; \langle \delta(\mathbf{x}; R) \delta(\mathbf{x} - \mathbf{r}; R) \rangle \;=\; \frac{1}{(2\pi)^d} \int d^d k\; e^{i\mathbf{k}\cdot\mathbf{r}} W(k)^2 P(k)\,. \tag{1.128}$$

[14] When $\mathbf{x}$ is closer than R to the edge of the volume V, the window function collects a contribution outside V. In this convolution theorem we used periodic boundary conditions. In real applications one needs to consider edge effects.

and the variance of the filtered density field is

$$\sigma^2(R) \equiv \langle \delta(\mathbf{x};R)^2 \rangle = \xi(0;R) = \int_0^\infty W(k)^2 \mathcal{P}(k) \frac{dk}{k} \,. \tag{1.129}$$

Considered as a function of R, it provides another measure of structure at different scales. Writing $W(\mathbf{k})$ and $P(\mathbf{k})$ in terms of their Fourier transforms, we get (**exercise**)

$$\sigma^2(R) = \frac{1}{(2\pi)^d} \int d^d k \, W(\mathbf{k})^2 P(\mathbf{k}) = \int d^d x d^d x' \, \xi(|\mathbf{x}' - \mathbf{x}|) W(\mathbf{x}) W(\mathbf{x}') \,. \tag{1.130}$$

Spectral moments. More generally, we define the spectral moments

$$\sigma_\ell^2(R) \equiv \int_0^\infty k^{2\ell} \mathcal{P}(k) W(k)^2 \frac{dk}{k} \,, \tag{1.131}$$

so that $\sigma^2(R) = \sigma_0^2(R)$ is the zeroth moment. The relation $\sigma^2(R) = \xi(0;R)$ can be generalized to higher moments and derivatives of $\xi(0;R)$ at $r = 0$, since, e.g.,

$$\begin{aligned} \nabla^2 \xi(\mathbf{r};R) &= \frac{1}{(2\pi)^d} \int d^d k \, (-k^2) e^{i\mathbf{k}\cdot\mathbf{r}} W(k)^2 P(k) \\ \Rightarrow \quad \nabla^2 \xi(0;R) &= \frac{1}{(2\pi)^d} \int d^d k \, (-k^2) W(k)^2 P(k) = -\sigma_1^2(R) \,. \end{aligned} \tag{1.132}$$

(The unfiltered $\xi(r)$ and its derivatives may diverge at $r = 0$, but $\xi(0;R)$ has been smoothed by the window function.) Peacock[1], p. 500 has

$$\xi^{(2\ell)}(0;R) = (-1)^\ell \frac{\sigma_\ell^2(R)}{2\ell + 1} \tag{1.133}$$

which probably holds as such only for $d = 3$, since I get

$$\xi''(0,R) = -\frac{\sigma_1^2(R)}{d} \tag{1.134}$$

(I did not try to do the higher moments).

To get (1.134) from (1.132) we need to relate $\nabla^2 \xi(\mathbf{r};R)$ to $\xi''(r,R)$ at $r = 0$. Expand

$$\xi(r,R) = \sum_{n=0}^\infty a_n r^n \,, \tag{1.135}$$

where $a_1 = 0$ so that $\xi(\mathbf{r};R)$ is smooth at $\mathbf{r} = 0$. Now

$$\partial_i r^n = n r^{n-1} \partial_i r = n r^{n-2} x_i \quad \Rightarrow \quad \partial_i \xi(r;R) = \sum_{n=2}^\infty a_n n r^{n-2} x_i \tag{1.136}$$

and

$$\partial_j \partial_i \xi(r,R) = \sum_{n=2}^\infty a_n n \left[(n-2) r^{n-4} x_i x_j + r^{n-2} \partial_j x_i \right] \tag{1.137}$$

where $\partial_j x_i = \delta_{ij}$. Thus

$$\nabla^2 \xi(r,R) \equiv \sum_i \partial_i \partial_i \xi(r) = \sum_{n=2}^\infty a_n n \left[(n-2) r^{n-2} + d \, r^{n-2} \right] \to 2d \, a_2 \tag{1.138}$$

as $r \to 0$, whereas

$$\xi''(r,R) = \sum_{n=2}^\infty a_n n(n-1) r^{n-2} \to 2a_2 \,, \tag{1.139}$$

so that

$$\nabla^2\xi(0;R) = d\cdot\xi''(0;R)\,. \tag{1.140}$$

The simplest window function is the *top-hat window function*

$$W_T(\mathbf{r}) \;\equiv\; \frac{1}{V(R)} \qquad \text{for} \quad |\mathbf{r}|\le R \tag{1.141}$$

and $W_T(\mathbf{r}) = 0$ elsewhere, i.e., $\delta(\mathbf{x})$ is filtered by replacing it with its mean value within the distance R. It's Fourier transform[15] is (**exercise**):

$$\begin{aligned}
&\text{1D:} \qquad W_T(k) &&= \frac{1}{kR}\sin kR \\
&\text{2D:} \qquad W_T(k) &&= \frac{2}{kR}J_1(kR) \\
&\text{3D:} \qquad W_T(k) &&= \frac{3}{(kR)^3}\left(\sin kR - kR\cos kR\right) \;=\; \frac{3}{kR}j_1(kR)
\end{aligned} \tag{1.142}$$

Mathematically more convenient is the *Gaussian window function*

$$W_G(r) \;\equiv\; \frac{1}{V_G(R)}e^{-\frac{1}{2}r^2/R^2}\,, \tag{1.143}$$

where

$$V_G(R) \;\equiv\; \int d^dr\, e^{-\frac{1}{2}|\mathbf{r}|^2/R^2} \tag{1.144}$$

is the *volume* of W_G. The volume of a window function is defined as what $\int d^dr\, W(\mathbf{r})$ would be if W were normalized so that $W(0) = 1$, instead of the normalization we chose in (1.122). For the top hat, this is the volume over which the filter averages; for others, a generalization of this.

The volume of W_G is (**exercise**)

$$V_G(R) = (2\pi)^{d/2}R^d\,, \tag{1.145}$$

and its Fourier transform is, for all d, (**exercise**)

$$W_G(k) \;=\; e^{-\frac{1}{2}(kR)^2}\,. \tag{1.146}$$

(The 1D case was done in FYMM Ib. From that it's easy to generalize to arbitrary d.)

We can also define the $\mathbf{k}$-*space top-hat window function*

$$W_k(\mathbf{k}) \;\equiv\; 1 \qquad \text{for} \quad k\le 1/R \tag{1.147}$$

and $W_k(\mathbf{k}) = 0$ elsewhere. In $\mathbf{x}$-space this becomes (**exercise**)

$$\text{3D:} \quad W_k(\mathbf{r}) \;=\; \frac{1}{2\pi^2R^3}\frac{\sin y - y\cos y}{y^3} \;=\; \frac{1}{6\pi^2R^3}\frac{3j_1(y)}{y}\,, \qquad \text{where} \quad y\equiv|\mathbf{r}|/R\,. \tag{1.148}$$

The integral $\int d^3r\, W_k(\mathbf{r}) = 4\pi\int r^2dr W_k(r)$ diverges: the integrand oscillates as a function of r with constant amplitude. The first zeros of the integrand are 4.493, 7.725, 10.904, 14.066, and 17.221; and integrated to these values, the integral gives 99.21, 20.645, 97.37, 21.24, and 97.10

[15] Note the emerging pattern with the Bessel functions: trigonometric functions are "Bessel functions for 1D", cos and sin corresponding to J_0 and J_1; the ordinary Bessel functions J_n are "for 2D"; and the spherical Bessel functions are "for 3D". All are oscillating functions; trigonometric functions have constant amplitude; J_n decay as $x^{-1/2}$ for large x, and $j_n(x)$ decay as x^{-1}.

times R^3. This oscillation is a disadvantage of this window function, but it has other advantages. In this case we define the volume to be

$$\text{3D:} \quad V_k(R) \;=\; 6\pi^2 R^3 \approx 59.22 R^3\,, \tag{1.149}$$

so that $W_k(0) = 1/V_k$ as for the other window functions. This value falls in the middle of the oscillation of the integral. Despite these oscillations, we expect (1.123) to converge, since for large $|\mathbf{x}-\mathbf{x}'|$ positive and negative δ tend to cancel.

For this window function the density variance is simply

$$\sigma^2(R) \;=\; \frac{1}{(2\pi)^d}\int_0^{R^{-1}} C_d k^{d-1} P(k) dk \;=\; \int_{-\infty}^{-\ln R} \mathcal{P}(k) d\ln k\,. \tag{1.150}$$

Note that the volumes of the different window functions are quite different. See Fig. 7. In 3D:

$$V_T(R) = \frac{4\pi}{3}R^3 = 4.189R^3\,, \quad V_G(R) = (2\pi)^{3/2}R^3 = 15.75R^3\,, \quad V_k(R) = 6\pi^2 R^3 = 59.22R^3\,. \tag{1.151}$$

The values of R that make the volumes equal are $R_G = 0.6431R_T$ and $R_k = 0.4136R_T$. Thus a given R corresponds to a somewhat different effective scale for the different window functions.

The different window functions also give quite different $\sigma^2(R)$. Observationally, the 3D galaxy distribution has ([1], p. 501, [2], p. 83).

$$\sigma_T^2(R) \approx 1.0 \qquad \text{for } R = 8\,h^{-1}\text{Mpc}. \tag{1.152}$$

Near these scales the slope of the correlation function is

$$\gamma \approx 1.8 \qquad \text{corresponding to } n = -1.2. \tag{1.153}$$

This slope does not hold at larger scales, and at $R = 30\,h^{-1}$Mpc, $\sigma_T^2(R)$ is already down to 10^{-2} ($\sigma \sim 0.1$ [2], p. 83). See also Fig. 3.

One may also ask, whether scales larger than the observed universe (i.e., the lower limit $k = 0$ or $\ln k = -\infty$ in the k integrals) are relevant, since we cannot observe the inhomogeneity at such scales. Due to such very-large-scale inhomogeneities, the average density in the observed universe may deviate from the average density of the entire universe. Inhomogeneities at scales somewhat larger than the observed universe could appear as an anisotropy in the observed universe. The importance of such large scales depends on how strong the inhomogeneities at these scales are, i.e., how the power spectrum behaves as $k \to 0$.

Exercise: We defined $\sigma^2(R)$ as an expectation value over the ensemble. Define $\widehat{\sigma}^2(R)$ as the volume average over a realization and show that

$$\widehat{\sigma}^2(R) \;\equiv\; \frac{1}{V}\int_V d^dx\, \widehat{\delta}(\mathbf{x},R)^2 \;=\; \frac{V}{(2\pi)^d}\int d^dk\, |\widehat{\delta}_{\mathbf{k}}|^2 W(k)^2\,. \tag{1.154}$$

Exercise: For a power-law spectrum and a Gaussian window function, show that

$$\sigma_G^2(R) \;=\; \frac{1}{2}\Gamma\left(\frac{n+d}{2}\right)\mathcal{P}(R^{-1})\,. \tag{1.155}$$

Exercise: For a power-law spectrum and **k**-space top-hat window function, show that

$$\sigma_k^2(R) \;=\; \frac{1}{n+d}\mathcal{P}(R^{-1})\,. \tag{1.156}$$

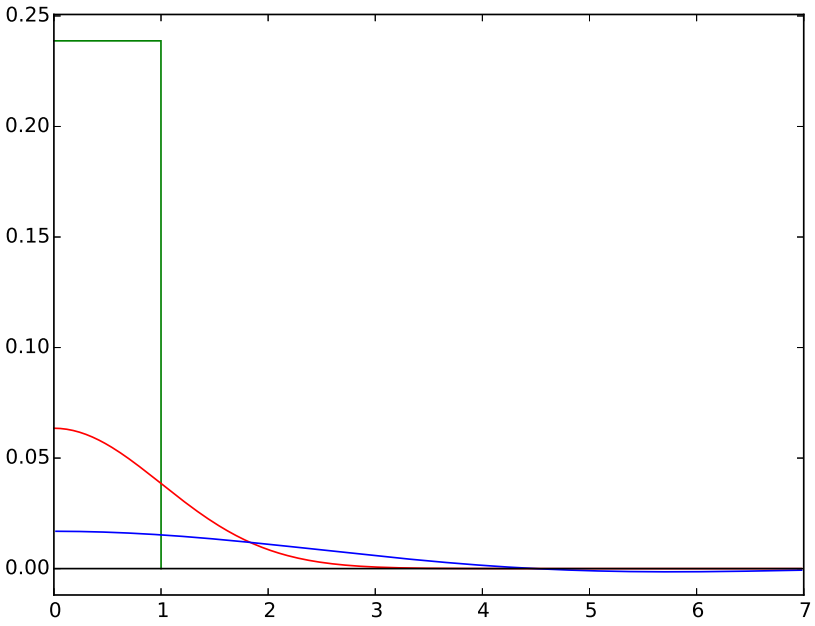


Figure 7: The 3D window functions $W(r)$, top-hat (green), Gaussian (red), and k (blue), for $R = 1$.

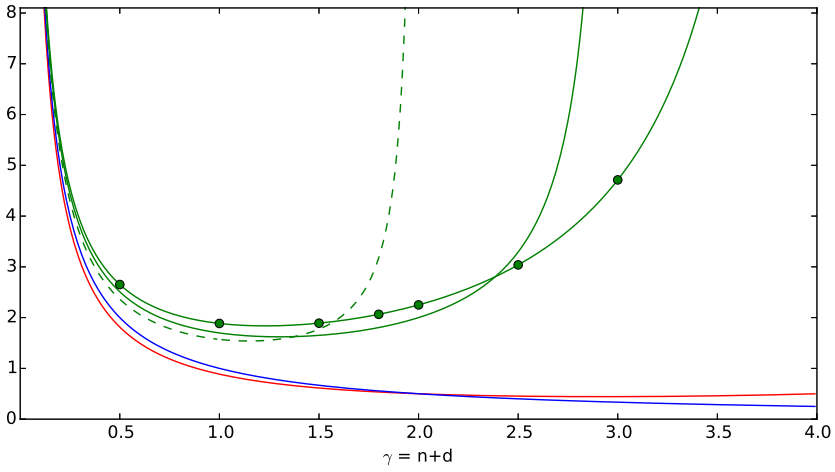


Figure 8: The ratio of $\sigma^2(R)$ to $\mathcal{P}(R^{-1})$ in the case of a power-law spectrum $\mathcal{P}(k) \propto k^{n+d}$ for the three different window functions: Gaussian (red), k (blue), 1D top-hat (green dashed), 2D top-hat (green), and 3D top-hat (green with dots). They all diverge in the limit $n \to -d$ ($\gamma \to 0$) due to the contributions of ever larger scales ($\ln k \to -\infty$). The divergence at $n \to 1$ for the top-hat window functions is a trickier thing. It has to do with their Fourier transform not dying off at high k fast enough.

Example: I wanted to do the same also for the top-hat window function, especially since (1.152). The cases seem different for different d, so try **first the 3D:**

$$\sigma_T^2(R) = \int_0^\infty \mathcal{P}(k) W_T(k)^2 \frac{dk}{k} = A^2 k_p^{-(n+3)} \int_0^\infty k^{n+3} \left(\frac{3}{kR}\right)^2 j_1^2(kR) \frac{dk}{k}$$
$$= 9A^2 (k_p R)^{-(n+3)} \int_0^\infty x^n j_1^2(x) dx = 9 I_n \mathcal{P}(R^{-1}) , \quad (1.157)$$

where

$$I_n = \int_0^\infty x^n j_1^2(x) dx . \quad (1.158)$$

I couldn't integrate I_n and didn't find it in integral tables. Wolfram Alpha said "computation time exceeded" for both I_n and I_{-n}, but for $n = -1.2$ it gave the remarkable result

$$I_{-1.2} = \frac{125\sqrt{5+\sqrt{5}}\Gamma(4/5)}{1386 \times 2^{3/10}} \approx 0.229418 . \quad (1.159)$$

This would give $\sigma_T^2(R) = 2.0648\,\mathcal{P}(R^{-1})$ for $n = -1.2$ ($\gamma = 1.8$), which appears surprisingly large, since the other two window functions give $\sigma_G^2(R) = 0.5343\,\mathcal{P}(R^{-1})$ and $\sigma_k^2(R) = 0.5556\,\mathcal{P}(R^{-1})$. Actually, I was equally surprised that σ_G^2 and σ_k^2 came so close to each other although the volumes of the two window functions are quite different. I would have been content, if σ_G^2 had been intermediate between σ_k^2 and σ_T^2. I think the explanation lies in the Fourier transform $W_T(k)$ not dying off fast enough for high k, so for large $n + d$, where there is lots of power at small scales, scales $\ll R$ keep contributing to $\sigma_T^2(R)$. When this is applied to galaxy number density, there will be another cut-off due to the finite distances between galaxies, so the effect of this high-k tail may not be fully realized...

The solution for doing the 3D case with Wolfram Alpha turned out to be to restrict the range of n to $n < 0$ and $n < -1$. This gives

$$I_n = \int_0^\infty x^n j_1^2(x) dx = 2^{-n} \frac{(n+1)\Gamma(n-1)}{n-3} \sin\frac{n\pi}{2} \qquad \text{for } -3 < n < 0 . \quad (1.160)$$

(I just assume that this result holds also for $0 < n < 1$; it diverges at $n \to 1$, see Fig. 8.) This expression is not defined for integer n. Instead we have (obtained easily with WolframAlpha)

$$I_{-2} = \frac{\pi}{15} , \qquad I_{-1} = \frac{1}{4} , \qquad I_0 = \frac{\pi}{6} . \quad (1.161)$$

Then the 1D case:

$$\sigma_T^2(R) = A^2 (k_p R)^{-(n+1)} \int_0^\infty x^{n-2} \sin^2 x \, dx = I_n \mathcal{P}(R^{-1}) , \quad (1.162)$$

where

$$I_n = \int_0^\infty x^{n-2} \sin^2 x \, dx , \quad (1.163)$$

which diverges at $x \to 0$ for $n \le -1$ and at $x \to \infty$ for $n \ge +1$. Gradshteyn&Ryzhik[9] 3.821.9 gives $I_0 = \pi/2$. Wolfram Alpha gives

$$I_n = 2^{-n} \sin\frac{n\pi}{2} \frac{\Gamma(n)}{1-n} \quad (0 < n < 1) \qquad \text{and} \qquad I_n = 2^{-n} \sin\left(\frac{-n\pi}{2}\right) \frac{\Gamma(n+1)}{n(n-1)} \quad (-1 < n < 0) \quad (1.164)$$

For $n + 1 = 1.8$ this gives $\sigma_T^2(R) = 3.1797\,\mathcal{P}(R^{-1})$, which is even more than I got for the 3D case.

Then the 2D case:

$$\sigma_T^2(R) = 4 I_n \mathcal{P}(R^{-1}) , \qquad \text{where} \quad I_n = \int_0^\infty x^{n-1} J_1(x)^2 \, dx . \quad (1.165)$$

Wolfram Alpha computation time was exceeded but Gradshteyn&Ryzhik[9] 6.574.2 gives

$$I_n = \frac{\Gamma(1-n)\Gamma(1+n/2)}{2^{1-n}\Gamma(1-n/2)^2\Gamma(2-n/2)} \qquad \text{for } n < 1. \quad (1.166)$$

These results for the different window functions are compared in Fig. 8.

2 Distribution of galaxies

Instead of a continuous density $\rho(\mathbf{x})$, we now consider a distribution of discrete objects. Their *number density* $\rho(\mathbf{x})$ is then only defined with a finite resolution.[16] The resolution depends on the application: the counts-in-cells approach of Secs. 2.3 and 2.4 has lower and the microcell approach of Sec. 2.5 has higher resolution, but anyway we stay far from infinite resolution (which for point-like objects, what galaxies in reality are not, would lead to delta functions). To make the discussion sound less abstract we call these objects galaxies, although they could also be other cosmological objects (e.g., clusters of galaxies), and another application is in numerical methods where a continuous density field is represented by a distribution of point masses.

2.1 The average number density of galaxies

It is often said that the observable universe contains about 200 billion (2×10^{11}) galaxies. If the observable universe is taken to mean everything until the last scattering surface (the origin of the cosmic microwave background) at $z = 1090$, which lies at comoving distance $r \approx 3.1 H_0^{-1} = 9300\, h^{-1}\mathrm{Mpc}$, then its comoving volume is

$$V \;=\; \frac{4\pi}{3} r^3 \;\approx\; 3400 h^{-3}\,\mathrm{Gpc}^3 \;=\; 3.4 \times 10^{12} h^{-3}\,\mathrm{Mpc}^3\,. \tag{2.1}$$

However, in this context the observable universe is taken to mean up to what distance we can actually observe galaxies. If we take this to mean $z = 8$, which lies at $r \approx 2.05 H_0^{-1} = 6100\, h^{-1}\mathrm{Mpc}$, the observable volume is

$$V_{\mathrm{obs}} \;\approx\; 970 h^{-3}\,\mathrm{Gpc}^3 \;=\; 9.7 \times 10^{11} h^{-3}\,\mathrm{Mpc}^3\,. \tag{2.2}$$

With $N_g = 2 \times 10^{11}$ galaxies this gives comoving mean galaxy number density

$$\bar\rho_g \;=\; \frac{N_g}{V_{\mathrm{obs}}} \;\approx\; \frac{0.21}{h^{-3}\,\mathrm{Mpc}^3}\,. \tag{2.3}$$

Our past light cone. In (2.2) and (2.3), I restricted the observable universe to $z \le 8$, and ignored evolution effects (higher redshifts corresponds to earlier times) to get a homogeneous galaxy mean number density to correspond to some recent $t = const$, but more appropriate would be to define the observable universe to correspond to our past light cone, all the way to $z = 1090$. The galaxies it contains are then those, whose world lines intersect this light cone. Recently I read that the number of galaxies in the observable universe (defined this way) is 10 times larger, since in the early universe galaxies were much smaller, and later they merged to form larger galaxies so the comoving galaxy number density went down. Thus (2.3) applies to the late universe, but in the early universe the density was more than 10 times larger. The page[17] links to a draft of an article by Conselice et al.[18]. Fig. 9 is from that article. The comoving galaxy number density increases towards higher z at first, because of the above evolution effect, but then begins to fall (probably near $z \sim 8$, but we do not have good data at such high redshifts) when we get to times when most galaxies had not yet formed.

The baryon density is $\bar\rho_b = \Omega_b \rho_{\mathrm{crit0}} = \omega_b h^{-2} \rho_{\mathrm{crit0}}$, where $\omega_b = 0.022$ [19], giving

$$\bar\rho_b = 6.1 \times 10^9 m_\odot / \,\mathrm{Mpc}^3 \tag{2.4}$$

[16] MBW[2], Sec. 6.1.2 takes a heavier approach here. They consider a two-step random process, where the first random process generates a continuous density field $\rho(\mathbf{x})$ and a second random process generates a point mass representation of it. The advantage of this is that there is no worry about dV being smaller than the resolution of $\rho(\mathbf{x})$. They also invoke ergodicity, but this does not seem necessary, if one refers to $\langle\rho\rangle$ instead of $\bar\rho$.

[17] `https://www.nasa.gov/feature/goddard/2016/hubble-reveals-observable-universe-contains-10-times-more-galaxies-than-previously-thought`

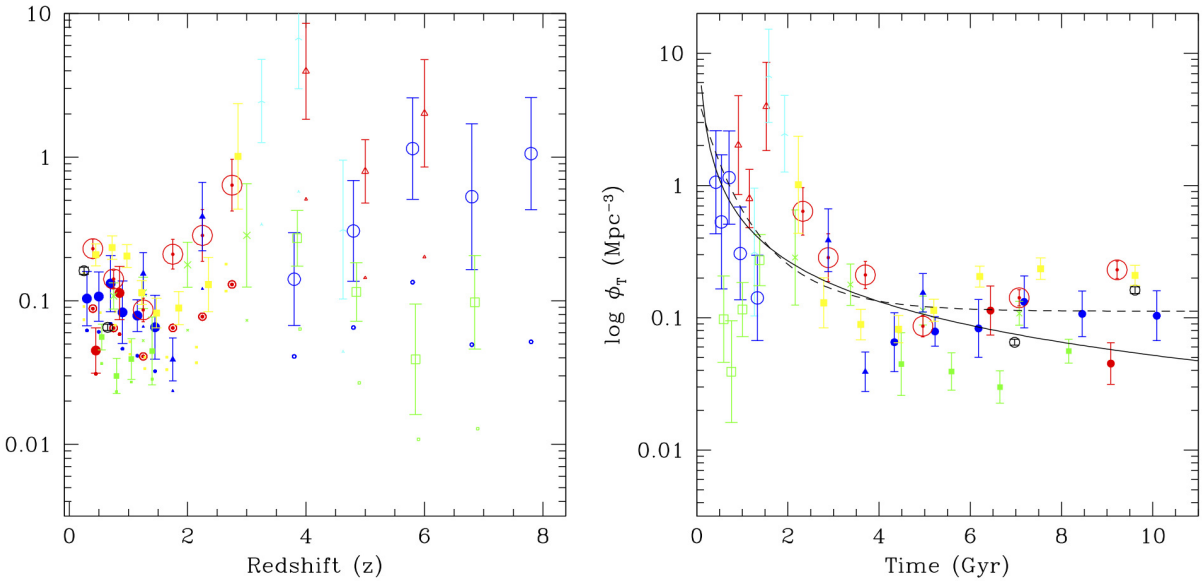


Figure 9: Evolution of the comoving total galaxy number density ϕ_T as a function of redshift and time [18]. The symbols with error bars are results from different surveys. The solid line on the right panel is a fit to the data points, and the dashed line is a fit of a galaxy merger model to them. The plot assumes $h = 0.7$, so to convert into units of $h^3\,\mathrm{Mpc}^{-3}$, multiply the vertical scale numbers (which are for ϕ_T, not $\log\phi_T$) by $h^{-3} = 2.9$.

With $h = 0.7$ this gives

$$\frac{\bar\rho_b}{\bar\rho_g} = 8.6\times10^{10} m_\odot = 1.7\times10^{41}\,\mathrm{kg} \qquad \text{baryonic matter per galaxy} \tag{2.5}$$

in the late universe. The total matter density parameter is $\omega_m = 0.14$ [19], so this gives $5.5\times10^{11} m_\odot$ total matter (baryonic + cold dark matter) per galaxy.

From [2], p.62 (Table 2.6) there are about 10 times as many dwarf galaxies in the local part of the universe as there are spiral galaxies; and the number of other types of galaxies is about half of that of spiral galaxies. The dwarf galaxies (defined as those with absolute magnitude $M_B \gtrsim -18$) thus make up most of the number of galaxies although they contain a relatively small fraction of all stars ([2], p. 57). The further out we look the larger the absolute luminosity (the smaller the absolute magnitude) of the galaxy has to be for us to be able to observe it. Thus the number density of observable galaxies is smaller than (2.3) and falls with distance.

For the 2D galaxy number density on the sky I can give a more definite number: The Euclid wide survey will cover $\Omega = 15000$ square degrees (= 36.36% of the sky) and is expected to observe 1.5 billion ($N_g = 1.5\times10^9$) galaxies (with apparent magnitude $m < 24.5$, observed sufficently well for their observed shapes to be used for weak lensing statistics)[20]. This corresponds to a 2D density of

$$\bar\rho_{g,\mathrm{Euclid}} = \frac{N_g}{\Omega} \approx \frac{30}{\mathrm{arcmin}^2}\,. \tag{2.6}$$

Most of these galaxies are at $z \lesssim 2$, corresponding to $r \lesssim 2H_0^{-1}$ or $V \approx 190h^{-3}\,\mathrm{Gpc}^3$, 1/5 of the volume to $z = 8$. Defining the "Euclid volume" as 15 000 square degrees of sky up to $z = 2$, we have

$$V_{\mathrm{Euclid}} \approx 70h^{-3}\,\mathrm{Gpc}^3\,, \tag{2.7}$$

which should contain about 14 billion galaxies, so Euclid will miss most of them.

Let's check if these numbers from the three sources[18, 2, 20] appear consistent with each other: Comparing the absolute magnitude of the brightest dwarf galaxies, $M = -18$, to the Euclid wide survey depth, $m = 24.5$, we conclude that even the brightest dwarf galaxies will be missed beyond distance modulus $m - M = 42.5 = -5 + 5 \lg d_L[\mathrm{pc}] \quad \Rightarrow \quad d_L = 10^{9.5}\mathrm{pc} \approx 3.16\,\mathrm{Gpc}$. With $h = 0.7$ this is $d_L = 2.2\,h^{-1}\mathrm{Gpc} = 0.74 H_0^{-1}$ corresponding to $z \approx 0.55$. (In a flat universe the luminosity distance and comoving distance are related by $d_L = (1+z)r$, so this corresponds to $r \approx 0.48 H_0^{-1}$.) Thus the Euclid wide survey should see some dwarf galaxies at $z < 0.55$ and none at $z > 0.55$, beyond which it will miss also some of the larger galaxies. This seems consistent with Euclid observing 1.5 billion out of a total of 14 billion galaxies in V_{Euclid}.

2.2 Galaxy 2-point correlation function

We treat individual galaxies as mathematical points, so that each galaxy has a coordinate value $\mathbf{x}$. We define the *galaxy 2-point correlation function* $\xi(\mathbf{r})$ as the *excess probability* of finding a galaxy at separation $\mathbf{r}$ from another galaxy:

$$d\mathbf{P} \equiv \langle\rho\rangle\,[1 + \xi(\mathbf{r})]\,dV \tag{2.8}$$

where $\langle\rho\rangle$ is the mean (ensemble average) galaxy number density, dV is a volume element that is a separation $\mathbf{r}$ away from a chosen reference galaxy, and $d\mathbf{P}$ is the probability that there is a galaxy within dV.

The probability of finding a galaxy in volume dV_1 at a random location $\mathbf{x}$ is

$$d\mathbf{P}_1 \;=\; \langle\rho(\mathbf{x})\rangle dV_1 \;=\; \langle\rho\rangle\langle 1 + \delta(\mathbf{x})\rangle dV_1 \;=\; \langle\rho\rangle dV_1\,. \tag{2.9}$$

The probability of finding a galaxy pair at $\mathbf{x}$ and $\mathbf{x}+\mathbf{r}$ is

$$\begin{aligned} d\mathbf{P}_{12} &= \langle\rho(\mathbf{x})\rho(\mathbf{x}+\mathbf{r})\rangle dV_1 dV_2 \;=\; \langle\rho\rangle^2\langle[1+\delta(\mathbf{x})][1+\delta(\mathbf{x}+\mathbf{r})]\rangle dV_1 dV_2 \\ &= \langle\rho\rangle^2\left[1 + \langle\delta(\mathbf{x})\rangle + \langle\delta(\mathbf{x}+\mathbf{r})\rangle + \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle\right] dV_1 dV_2 \\ &= \langle\rho\rangle^2\left[1 + \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle\right] dV_1 dV_2\,, \end{aligned} \tag{2.10}$$

since $\langle\delta(\mathbf{x})\rangle = \langle\delta(\mathbf{x}+\mathbf{r})\rangle = 0$. Dividing $d\mathbf{P}_{12}$ with $d\mathbf{P}_1$ we get the probability $d\mathbf{P}_2$ of finding the second galaxy once we have found the first one

$$d\mathbf{P}_2 \;=\; \langle\rho\rangle\left[1 + \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle\right] dV_2\,. \tag{2.11}$$

Comparing (2.11) to (2.8) we see that our new definition of ξ agrees with the old one

$$\xi(\mathbf{r}) \;=\; \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle\,. \tag{2.12}$$

Thus, for any galaxy, $\langle\rho\rangle[1+\xi(\mathbf{r})]dV$ is the expectation number of galaxies in a volume element dV at separation $\mathbf{r}$ and the mean number of neighbors within a spherical shell is

$$dN(r) = \langle\rho\rangle\,[1 + \xi(r)]\,C_d r^{d-1} dr \tag{2.13}$$

and the mean number of neighbors within distance R is

$$N(R) \;=\; \langle\rho\rangle V(R) + \langle\rho\rangle\int_0^R \xi(r) C_d r^{d-1} dr \;=\; \langle\rho\rangle V(R)\left[1 + \bar{\xi}(R)\right]\,. \tag{2.14}$$

Thus $1 + \xi(r)$ can be interpreted as the *mean (expected) density profile around each galaxy.*

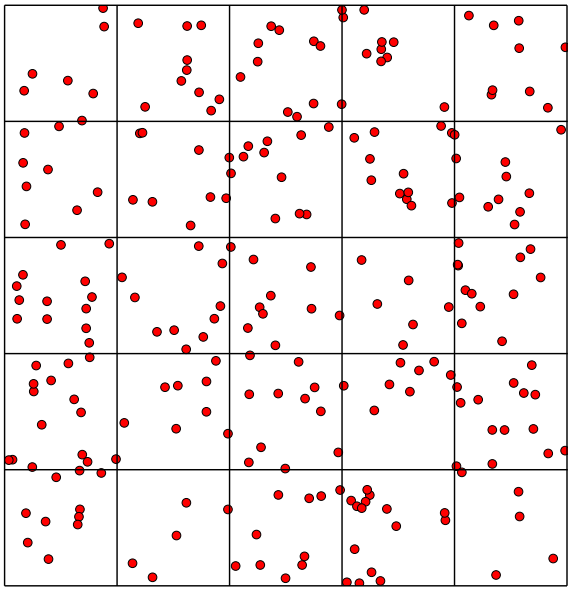


Figure 10: Poisson distribution of $N = 250$ galaxies. The 2D volume is divided into $M = 25$ cells, so that on average a cell should contain 10 galaxies. There are two cells with just 5 galaxies.

2.3 Poisson distribution

A *Poisson distribution* is an uncorrelated distribution of galaxies, which we get when we assign each galaxy i a random location $\mathbf{x}_i$ (with uniform probability density within V) independently. (This process is called a *Poisson process*.) Then

$$d\mathbf{P}_{12} = \langle\rho\rangle^2 dV_1 dV_2 \quad \Rightarrow \quad \xi(\mathbf{r}) = 0\,. \tag{2.15}$$

Divide the volume V into M subvolumes (cells) ΔV. Assign now N galaxies into V with a Poisson process. See Fig. 10. Each galaxy lands in a particular cell with probability $p = 1/M$, and somewhere else with probability $1 - p = 1 - 1/M$. The probability of the n first galaxies landing in this cell and the remaining $N-n$ elsewhere is thus $(1/M)^n(1-1/M)^{N-n}$. Since there are

$$\binom{N}{n} \equiv \frac{N!}{n!(N-n)!} \tag{2.16}$$

ways of choosing n galaxies out of N, the probability of getting exactly n galaxies in a particular cell is

$$\mathbf{P}(n) = \binom{N}{n}\left(\frac{1}{M}\right)^n\left(1-\frac{1}{M}\right)^{N-n}\,. \tag{2.17}$$

This is the nth term of the binomial expansion

$$(a+b)^N = \sum_{n=0}^{N}\binom{N}{n}a^n b^{N-n} \tag{2.18}$$

so we can easily check that the total probability is

$$\sum_{n=0}^{N}\mathbf{P}(n) = \sum_{n=0}^{N}\binom{N}{n}\left(\frac{1}{M}\right)^n\left(1-\frac{1}{M}\right)^{N-n} = \left[\frac{1}{M}+\left(1-\frac{1}{M}\right)\right]^N = 1^N = 1\,. \tag{2.19}$$

This probability distribution (2.17) of integers is called the *binomial distribution* $B(N,p)$, where $p = 1/M$. The *Poisson limit theorem* states that: if $N \to \infty$ and $p \to 0$ ($M \to \infty$) so that $Np = N/M \to \lambda$ (keep cells at fixed sized and increase the survey volume), then

$$\mathbf{P}(n) \;=\; \binom{N}{n} p^n (1-p)^{N-n} \;\to\; \frac{\lambda^n e^{-\lambda}}{n!} \;=\; \left(\frac{N}{M}\right)^n \frac{e^{-N/M}}{n!} \tag{2.20}$$

This probability distribution of integers is called the *Poisson distribution.*

For the Poisson distribution we have the expectation values (**exercise**)

$$\begin{aligned}
\langle n \rangle &= \sum_{n=0}^{\infty} n\mathbf{P}(n) \;=\; e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} \;=\; \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \;=\; \lambda \;=\; N/M \\
\langle n^2 \rangle &= \sum_{n=0}^{\infty} n^2 \mathbf{P}(n) \;=\; \ldots \;=\; \lambda + \lambda^2 \\
\langle (\Delta n)^2 \rangle &\equiv \langle (n - \langle n \rangle)^2 \rangle \;=\; \langle n^2 \rangle - \langle n \rangle^2 \;=\; \lambda \,.
\end{aligned} \tag{2.21}$$

The galaxy density in a cell is $\rho(\mathbf{x}) = n/\Delta V$ and the density perturbation is $\delta(\mathbf{x}) = (n - \langle n \rangle)/\langle n \rangle$, so that

$$\langle \delta^2 \rangle \;=\; \frac{\langle (n - \langle n \rangle)^2 \rangle}{\langle n \rangle^2} \;=\; \frac{1}{\lambda} \;=\; \frac{1}{\langle n \rangle} \,. \tag{2.22}$$

(We will later, in Sec. 8, refer to this as *Poisson variance*: the relative variance is 1/expected number of points.) Thus, although for a Poisson distribution $\xi(\mathbf{r}) = 0$ for $\mathbf{r} \neq 0$, we have

$$\xi(0) \;=\; \langle \delta^2 \rangle \;=\; \frac{M}{N} \;=\; \frac{1}{\langle n \rangle} \tag{2.23}$$

(in the limit of very large N and M). This density variance depends on the resolution, since increasing the number of cells M with fixed volume V and fixed number of galaxies N (decreasing ΔV) makes its larger.

(Note: This discussion was done with fixed N, corresponding to the subensemble $\bar{\rho} = \langle \rho \rangle$; we should also consider allowing N to vary, presumably also with a Poisson distribution $\mathbf{P}(N) = \lambda^N e^{-\lambda}/N!$, where now $\lambda = \langle N \rangle = \langle \rho \rangle V$ (TBD).)

We could continue with this approach to specify that the volume V is cubic and the division into M cells forms a rectangular grid, replacing M with M^d and doing a discrete Fourier transform to find the power spectrum of the Poisson distribution, finding that

$$P(k) \;=\; const \,, \tag{2.24}$$

but since this seems not to be the usual approach in literature, I skip this, avoiding the discussion of discrete Fourier transforms. (Update: I have now added this below.) However, we make some comments on the result: The Poisson distribution has the power spectrum of white noise, $n = 0$, the amplitude depending on the resolution. We noted in Sec. 1.9 that as $n \to 0$, $\gamma \to d$. Now we have that $\xi(\mathbf{r}) = 0$ for $\mathbf{r} \neq 0$, but $\xi(0) \propto (\Delta r)^{-d}$ (for fixed V and N), where Δr is the side of the cell, i.e., $\Delta V = (\Delta r)^d$.

Discrete Fourier transform. For simplicity, we do this here in 1D only, and I have put this in small print, since we introduce another way to do a Fourier transform for a discrete set of objects in Sec. 2.5. In spite of being in 1D, I use the vector notation for the (1D) vectors $\mathbf{x}$ and $\mathbf{k}$ before switching to the discrete notation, where they get replaced by the integers j and k. We denote the 1D volume with L (length), and divide it into M cells with size Δ, so that $L = M\Delta$. For convenience, let M be even. Let $j = 0, \ldots, M-1$ index the cells, so that the location of cell j is $\mathbf{x}_j = j\Delta$ (we may assume that this refers to the center of the cell, so that we have chosen to put the coordinate origin at the center of the first cell, so that the "survey" extends from $-\frac{1}{2}\Delta$ to $L - \frac{1}{2}\Delta$). The density perturbation in cell j is

$$\delta(\mathbf{x}_j) = \delta(j\Delta) \equiv \delta_j = \frac{n_j - \langle n \rangle}{\langle n \rangle} \,, \tag{2.25}$$

where n_j is the number of galaxies in cell j.

In place of the Fourier coefficient of a continuous function

$$\delta_{\mathbf{k}} = \frac{1}{L}\int \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} dx\,, \tag{2.26}$$

where the allowed values for the wave vector (wave number) $\mathbf{k}$ are $k(2\pi/L)$ where k is an integer, we have now the sum

$$\delta_k \;=\; \frac{1}{L}\sum_{j=0}^{M-1} \delta_j e^{-ij\Delta k(2\pi/L)}\Delta \;=\; \frac{\Delta}{L}\sum_{j=0}^{M-1} \delta_j e^{-i2\pi jk/M}\,, \tag{2.27}$$

where $\Delta/L = 1/M$.

The discretization imposes an upper limit to what Fourier modes can be resolved: the maximum value of k is such that the wavelength is equal to 2Δ:

$$k_{\max}\frac{2\pi}{L} = \frac{2\pi}{2\Delta} = \frac{\pi}{\Delta} \quad\Rightarrow\quad k_{\max} = \frac{L}{2\Delta} = \frac{M}{2}\,, \tag{2.28}$$

and the allowed values are

$$k = -\tfrac{1}{2}M,\; -\tfrac{1}{2}M+1, \ldots, \tfrac{1}{2}M\,. \tag{2.29}$$

Since $\exp(-i2\pi jk/M)$ is periodic in k, i.e., $k+M$ gives the same values as k, we can always replace k with $k+M$ or $k-M$. Also modes $k = -\frac{1}{2}M$ and $k = \frac{1}{2}M$ look the same, so we only have M different modes. We can thus use the more convenient range

$$k = 0, 1, \ldots M-1 \tag{2.30}$$

instead of (2.29).

The different Fourier modes are orthogonal (**exercise**)

$$\sum_{k=0}^{M-1} e^{i2\pi(j-j')k/M} \;=\; M\delta_{jj'}\,. \tag{2.31}$$

Using this, one can easily show (**exercise**) that

$$\delta_j = \sum_{k=0}^{M-1} \delta_k e^{i2\pi jk/M}\,. \tag{2.32}$$

Equations (2.27) and (2.32) form the *discrete Fourier transform* pair.

We can now calculate

$$\begin{aligned}
\langle \delta_k^* \delta_{k'} \rangle \;&=\; \frac{\Delta^2}{L^2}\sum_{j=0}^{M-1}\sum_{j'=0}^{M-1} \langle \delta_j \delta_{j'} \rangle e^{i2\pi(jk-j'k')/M} \\
&=\; \frac{\Delta^2}{L^2}\sum_{j=0}^{M-1}\sum_{r=0}^{M-1} \langle \delta_j \delta_{j+r} \rangle e^{-i2\pi jk/M} e^{i2\pi(j+r)k'/M} \\
&=\; \frac{\Delta^2}{L^2}\sum_{r=0}^{M-1} \xi_r e^{-i2\pi rk'/M} \sum_{j=0}^{M-1} e^{i2\pi j(k-k')/M} \\
&=\; \frac{\Delta}{L}\sum_{r=0}^{M-1} \xi_r e^{-i2\pi rk/M}\delta_{kk'} \;\equiv \frac{1}{L}\delta_{kk'}P_k\,,
\end{aligned} \tag{2.33}$$

where we have 1) written $j' = j+r$, where we can let also the integer r range from 0 to $M-1$, since everything wraps over in M; 2) defined $\xi_r \equiv \langle \delta_j \delta_{j+r} \rangle$ (invoking statistical homogeneity); 3) used (2.31); and 4) defined the *power spectrum* P_k as the discrete Fourier transform of ξ_r:

$$P_k \;\equiv\; \Delta \sum_{r=0}^{M-1} \xi_r e^{-i2\pi rk/M}\,. \tag{2.34}$$

(The difference in the forms of (2.27) and (2.34), i.e., the missing L, is just the difference in (1.57). I could have made other choices in the definitions about where to place L and Δ.)

For a Poisson distribution, $\xi_r = 0$, except

$$\xi_0 = M/L\langle\rho\rangle = \frac{1}{\Delta\langle\rho\rangle} = \frac{1}{\langle n\rangle}\,, \tag{2.35}$$

so that

$$P_k \;=\; \frac{1}{\langle\rho\rangle} \qquad \text{for all } k\,. \tag{2.36}$$

Exercise: Cox process. A Cox process refers to a combination of two Poisson processes. Consider the following examples:

1. Infinitely long lines are placed randomly (the first Poisson process) into an infinite volume. Galaxies are then assigned randomly (the second Poisson process) on these lines, so that the mean (expectation value of) linear number density on these lines is λ. What is $\xi(r)$, given in terms of λ and the resulting mean (expectation value of) 3D galaxy number density $\langle\rho\rangle$?
2. Like the previous case, but now the line segments have finite length L (all have the same length). What is $\xi(r)$, in terms of λ, L, and the number density n_s of line segments?

Answer:

1.
$$\xi(r) \;=\; \frac{\lambda}{2\pi\langle\rho\rangle} r^{-2} \tag{2.37}$$

2.
$$\xi(r) \;=\; \frac{1}{2\pi n_s L\, r^2}\left(1 - \frac{r}{L}\right) \quad \text{for } r \le L\,; \qquad \xi(r) \;=\; 0 \quad \text{for } r > L \tag{2.38}$$

Note that this does not depend on λ ($\langle\rho\rangle = n_s L\lambda$, so it cancels in the ratio $\lambda/\langle\rho\rangle$).

Note: The Cox process has been used to generate "mock" catalogs for testing computer codes to estimate $\xi(r)$. It is actually not easy to generate test catalogs which have exactly some known, but non-zero, correlation function $\xi(r)$ that the estimate could be compared to. Cox is one way to do this.

Exercise: Cox process with binning. Continuation of previous exercise: When measuring the correlation function $\xi(r)$ from data, one has to count it for bins of finite width Δr. So instead of asking what is the excess probability $\xi(r)\langle\rho\rangle dV$ for another galaxy at separation r (so that $dV = 4\pi r^2 dr$), for comparing data with theory we ask what is the excess probability $\xi(r_1, r_2)\langle\rho\rangle\Delta V$ for a separation between r_1 and r_2 (i.e., here $\Delta r = r_2 - r_1$). Find $\xi(r_1, r_2)$ for $r_1 < r_2 \le L$. ($\xi = 0$ for $r_1 \ge L$, and the answer for $r_1 < L < r_2$ is more complicated but rarely needed.)

Answer:

1.
$$\xi(r_1, r_2) \;=\; \frac{\lambda}{2\pi\langle\rho\rangle}\frac{3(r_2 - r_1)}{r_2^3 - r_1^3} \;=\; \frac{\lambda}{2\pi\langle\rho\rangle}\frac{3}{r_1^2 + r_1 r_2 + r_2^2} \tag{2.39}$$

2.
$$\xi(r_1, r_2) \;=\; \frac{1}{2\pi n_s L}\frac{3(r_2 - r_1)}{r_2^3 - r_1^3}\left(1 - \frac{r_1 + r_2}{2L}\right) \;=\; \frac{1}{2\pi n_s L}\frac{3}{r_1^2 + r_1 r_2 + r_2^2}\left(1 - \frac{r_1 + r_2}{2L}\right) \tag{2.40}$$

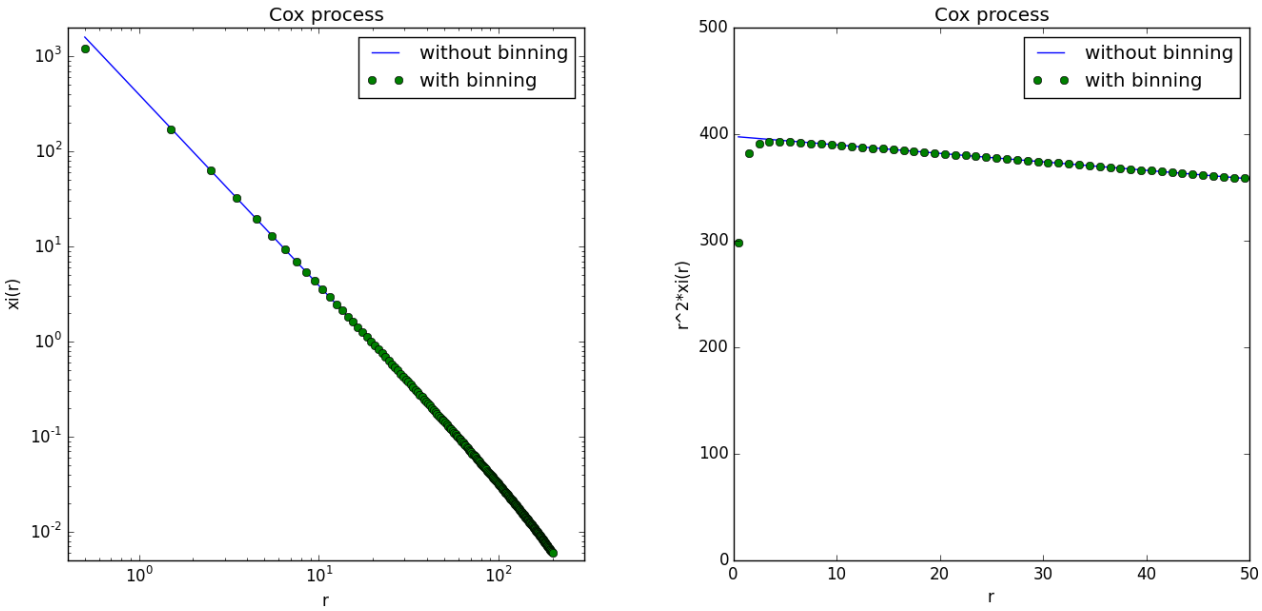


Table 2: The Cox process correlation function for $L = 500$, $n_s = 8 \times 10^{-7}$, and $\Delta r = 1$ binning. On the left: a log-log plot of $\xi(r)$. On the right: lin-lin plot of $r^2\xi(r)$.

2.4 Counts in cells

One of the first methods to measure the clustering properties of galaxies was dividing the survey volume V into cells (subvolumes) of equal size ΔV and shape, and count the number n of galaxies in each cell. Defining

$$\Delta n \equiv n - \langle n \rangle \tag{2.41}$$

the variance

$$\mu_2 \equiv \langle (\Delta n)^2 \rangle \tag{2.42}$$

and skewness

$$\mu_3 \equiv \langle (\Delta n)^3 \rangle \,, \tag{2.43}$$

we have that for a completely random (i.e., Poisson) distribution of galaxies

$$\mu_2 = \mu_3 = \langle n \rangle \,. \tag{2.44}$$

A clustered distribution will have a larger variance. We define

$$y \equiv \frac{\mu_2 - \langle n \rangle}{\langle n \rangle^2} \tag{2.45}$$

as a measure of clustering. It can be shown that

$$\langle y \rangle = \frac{\int_{\Delta V} \int_{\Delta V} \xi(\mathbf{x}_1 - \mathbf{x}_2) d^d x_1 d^d x_2}{\Delta V^2} \tag{2.46}$$

Likewise we define

$$z \equiv \frac{\mu_3 - 3\mu_2 + 2\langle n \rangle}{\langle n \rangle^3} \tag{2.47}$$

as a measure of excess skewness of the clustering. It measures nonlinear effects in structure growth and non-Gaussianity of primordial perturbations.

For an actual survey we define corresponding quantities $\widehat{\mu}_2$, $\widehat{\mu}_3$, $\widehat{y}$, and $\widehat{z}$, by replacing expectation values with survey averages. Note that these are biased estimators, $\langle\widehat{y}\rangle \neq y$, $\langle\widehat{z}\rangle \neq z$, since taking expectation values does not commute with raising to second or third power and division. One can study scale dependence of structure by using larger or smaller cells. This method is better than correlation function estimation for detecting structure at large scales, where the correlation function is small.

An important improvement of this method is to, instead of using disjoint cells, assign a much larger number of cells of the same size ΔV to random locations within the survey, allowing them to overlap. This oversampling does not change the expected variance, which is determined by (2.46), but the measured variance will be closer to this expectation value. [14, 15]

2.5 Fourier transform for a discrete set of objects

When we replaced the continuous density field with a set of discrete objects, the resolution of our description was reduced. The standard approach of a discrete Fourier transform, where one introduces a rectangular grid with finite resolution, introduces another, independent, loss of resolution, which is unnecessary. We would lose the information on the exact locations of the galaxies if we just assigned them into finite cells. The only discreteness we need to introduce is that inherent in the problem, that of the discrete point set. This is done by introducing *microcells*. We start as if we were going to do a normal discrete Fourier transform, dividing the volume into cells. But now we make the cells small compared to the scales of galaxy distribution, so that the probability of there being more than one galaxy in a cell becomes zero, and specifying in which cell a galaxy i is, specifies its "exact" location $\mathbf{x}_i$. Denote the volume of such a microcell with δV. Most microcells will be empty. The galaxy number density in microcell j is $n_j/\delta V$, where $n_j = 0$ or 1. This means that $n_j^2 = n_j$, which will be very helpful later.

The Fourier coefficients of the density field become

$$\rho_{\mathbf{k}} = \frac{1}{V}\int_V \rho(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}d^dx = \frac{1}{V}\sum_j (n_j/\delta V)e^{-i\mathbf{k}\cdot\mathbf{x}_j}\delta V = \frac{1}{V}\sum_j n_j e^{-i\mathbf{k}\cdot\mathbf{x}_j}\,, \tag{2.48}$$

where the sum is over microcells, and $\mathbf{x}_j$ is the location of the microcell. But since $n_j = 0$ for all the empty microcells ($n_j = 0$), only those terms survive, where the microcell contains a galaxy, $n_j = 1$, and the sum becomes a sum over galaxies

$$\rho_{\mathbf{k}} = \frac{1}{V}\sum_i e^{-i\mathbf{k}\cdot\mathbf{x}_i}\,, \tag{2.49}$$

where $\mathbf{x}_i$ is the location of galaxy i. Thus

$$\begin{aligned}
\delta_{\mathbf{k}} &= \frac{\rho_{\mathbf{k}}}{\langle\rho\rangle} = \frac{1}{\langle\rho\rangle V}\sum_i e^{-i\mathbf{k}\cdot\mathbf{x}_i} = \frac{1}{\langle N\rangle}\sum_{i=1}^{N} e^{-i\mathbf{k}\cdot\mathbf{x}_i} \qquad \text{for } \mathbf{k}\neq 0 \\
\delta_0 &= \bar{\delta} = \frac{\bar{\rho}}{\langle\rho\rangle} - 1 = \frac{N-\langle N\rangle}{\langle N\rangle}\,,
\end{aligned} \tag{2.50}$$

where N is the total number of galaxies in the volume V, and $\langle N\rangle = \langle\rho\rangle V$ is its expectation value.

2.5.1 Poisson distribution again

We apply now our new Fourier method to the Poisson distribution, where the locations $\mathbf{x}_i$ are independent random numbers, so that the complex numbers $e^{-i\mathbf{k}\cdot\mathbf{x}_i}$ in (2.50) are distributed

randomly on the unit circle of the complex plane. Doing the sum $\sum_{i=1}^{N} e^{-i\mathbf{k}\cdot\mathbf{x}_i}$ thus executes a random walk on the complex plane, with step length 1.

To get the power spectrum,

$$|\delta_{\mathbf{k}}|^2 = \delta_{\mathbf{k}}^*\delta_{\mathbf{k}} = \frac{1}{\langle N\rangle^2}\sum_{ij} e^{i\mathbf{k}\cdot(\mathbf{x}_i-\mathbf{x}_j)} = \frac{1}{\langle N\rangle^2}\left(\sum_{i\neq j} e^{i\mathbf{k}\cdot(\mathbf{x}_i-\mathbf{x}_j)} + \sum_i 1\right)$$
$$= \frac{1}{\langle N\rangle^2}\left(2\sum_{\text{pairs}}\cos(\mathbf{k}\cdot(\mathbf{x}_i-\mathbf{x}_j)) + N\right). \tag{2.51}$$

There is an equal probability for the $\cos(\mathbf{k}\cdot(\mathbf{x}_i-\mathbf{x}_j))$ to be positive or negative, so the expectation value of the first term vanishes, and we get

$$P(k) = V\langle|\delta_{\mathbf{k}}|^2\rangle = V\frac{\langle N\rangle}{\langle N\rangle^2} = \frac{V}{\langle N\rangle} = \frac{1}{\langle\rho\rangle}, \tag{2.52}$$

which is independent of k, i.e., the spectral index is $n = 0$.

In Sec. 9.1 we redo this for a correlated distribution, where we see that the effect of having a discrete set of galaxies instead of a continuous density adds this $V/\langle N\rangle = 1/\langle\rho\rangle$ term to the power spectrum. This added term is called *shot noise*. The higher the density of points (galaxies included in the survey) the smaller is this shot noise. For an estimate of the power spectrum of the underlying mass distribution we subtract this shot noise.

Probability distribution of $\widehat{P}(\mathbf{k})$: We will later discuss estimation of power spectrum from galaxy surveys in more detail, but let us already consider how individual realizations differ from the expectation value, i.e., how is

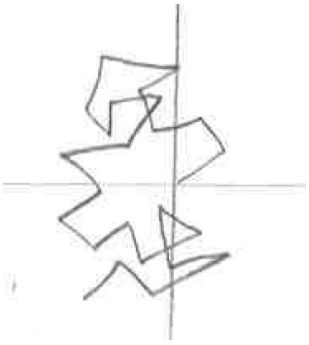
$$\widehat{P}(\mathbf{k}) \equiv V|\widehat{\delta}_{\mathbf{k}}|^2 \tag{2.53}$$

distributed around the expectation value $P(k)$. We assume a fixed N (which may be different from $\langle N\rangle$), i.e., we do not here fold in the probability distribution of N. Consider the complex number $\widehat{\delta}_{\mathbf{k}}$ as a 2D vector

$$\widehat{\delta}_{\mathbf{k}} = \frac{1}{N}\sum_{i=1}^{N} e^{-i\mathbf{k}\cdot\mathbf{x}_i} = \frac{1}{N}\left(\sum_i \cos\mathbf{k}\cdot\mathbf{x}_i\,,\; -\sum_i \sin\mathbf{k}\cdot\mathbf{x}_i\right), \tag{2.54}$$

which points to the endpoint of the random walk (or $N\widehat{\delta}_{\mathbf{k}}$ does). Consider the real part of $\widehat{\delta}_{\mathbf{k}}$:

$$\frac{1}{N}\sum_i \cos\mathbf{k}\cdot\mathbf{x}_i\,. \tag{2.55}$$



Here the terms in the sum are independent random variables with a nonuniform probability distribution ($\mathbf{k}\cdot\mathbf{x}_i$ has a uniform probability distribution[18]).

We apply the *central limit theorem*: The sum (or mean) of independent random variables approaches the normal (Gaussian) distribution,

$$\mathbf{P}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}, \tag{2.56}$$

where μ is the expectation value and σ^2 the variance, as $N\to\infty$, regardless of the probability distribution of the individual variables. If each variable has the same probability distribution, with expectation value μ and variance σ^2, then their mean will have expectation value μ and variance σ^2/N.

[18]Comment by Elina Keihänen: I do not think $\mathbf{k}\cdot\mathbf{x}$ has uniform distribution, except in 1D. It is a sum of three terms, $k_1x_1 + k_2x_2 + k_3x_3$, each of which separately is uniformly distributed, but their sum is not. Suprisingly, (2.57) still seems to hold. I checked this numerically.

The expectation value of $\cos \mathbf{k}\cdot\mathbf{x}_i$ is zero and the variance is $\frac{1}{2\pi}\int_0^{2\pi}\cos^2 x dx = \frac{1}{2}$. Thus $\mathrm{Re}\,\widehat{\delta}_{\mathbf{k}}$ has the probability distribution

$$\mathbf{P}\big(\mathrm{Re}\,\widehat{\delta}_{\mathbf{k}} = x\big) \;=\; \sqrt{\frac{N}{\pi}}e^{-Nx^2}\,. \tag{2.57}$$

The imaginary part has the same probability distribution. Now clearly the real and imaginary parts are correlated. The individual terms are fully correlated since $\sin = \sqrt{1-\cos^2}$. Some of this correlation remains for the sums, especially in the large-$|\widehat{\delta}_{\mathbf{k}}|$ tail of the probability distribution. Clearly, if the real part is close to its possible maximum value ($\cos\mathbf{k}\cdot\mathbf{x}_i$ has mostly landed near 1), then the imaginary part has to be small, and vice versa. However, far from the tail we can expect the correlation between the sums (the two components of the random walk) be negligible. Making this approximation,[19] we get the 2D probability distribution

$$\mathbf{P}\big(\widehat{\delta}_{\mathbf{k}}\big) \;=\; \frac{N}{\pi}e^{-N|\widehat{\delta}_{\mathbf{k}}|^2}\,. \tag{2.58}$$

To convert this into a probability distribution for $|\widehat{\delta}_{\mathbf{k}}|^2$, we need to integrate. Do first

$$\mathbf{P}\big(|\widehat{\delta}_{\mathbf{k}}| = r\big)dr \;=\; Ne^{-Nr^2}2rdr\,. \tag{2.59}$$

This is known as the *Rayleigh distribution*. For $s = |\widehat{\delta}_{\mathbf{k}}|^2 = r^2$, $ds = dr^2 = 2rdr$, so

$$\mathbf{P}\big(|\widehat{\delta}_{\mathbf{k}}|^2 = s\big)ds \;=\; Ne^{-Ns}ds\,, \tag{2.60}$$

and

$$\mathbf{P}\big(|\widehat{\delta}_{\mathbf{k}}|^2 > s\big) \;=\; N\int_s^\infty e^{-Ns}ds \;=\; e^{-Ns}\,. \tag{2.61}$$

The mean of this distribution is $1/N$ and the variance is $2/N^2$, i.e,

$$\langle\widehat{P}(\mathbf{k})\rangle = \frac{V}{N} = \frac{1}{\bar{\rho}} \qquad \text{and} \qquad \left\langle\Big(\widehat{P}(\mathbf{k}) - \langle\widehat{P}(\mathbf{k})\rangle\Big)^2\right\rangle = 2\left(\frac{V}{N}\right)^2 = \frac{2}{\bar{\rho}^2}\,. \tag{2.62}$$

Thus, although the expectation value of $\widehat{P}(\mathbf{k})$ agrees with $P(k)$, the variance around it is large and the most probable value is actually $\widehat{P}(\mathbf{k}) = 0$! Note, however, that this is for an individual Fourier mode $\mathbf{k}$, and to estimate $P(k)$ from a survey one would take the mean over a large number of Fourier modes for which $|\mathbf{k}|$ falls between k and $k+dk$, and the variance of this mean is then much lower.

We get the probability distribution of N by just applying (2.20) but considering the full volume V as one subvolume of an infinite universe, i.e., replacing n by N,

$$\mathbf{P}(N) = \langle N\rangle^N\frac{\exp^{-\langle N\rangle}}{N!}\,. \tag{2.63}$$

---

[19] Peacock[1] does not mention any such approximation (maybe it comes as part of the $N \gg 1$ limit), but I need to make it to get his Eq. (16.115), i.e., our (2.61).

3 Subspaces of lower dimension

Consider now how the statistics of a 1D or 2D subspace are related to the statistics in the full 3D space. A key starting point here is that because of statistical isotropy, the correlation function

$$\xi(r) = \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle \tag{3.1}$$

is the same, whether $\mathbf{x}$ and $\mathbf{x}+\mathbf{r}$ are restricted on a 1D line or 2D plane, or not.

3.1 Skewers

Consider the power spectrum $\mathcal{P}_{\rm 1D}(k)$ along a straight line ('skewer') going through 3D space. Since

$$\mathcal{P}_{\rm 1D}(k) = \frac{2k}{\pi}\int_0^\infty \xi(r)\cos krdr \qquad \text{and} \qquad \xi(r) = \int_0^\infty \mathcal{P}_{\rm 3D}(k)\frac{\sin kr}{kr}\frac{dk}{k}\,, \tag{3.2}$$

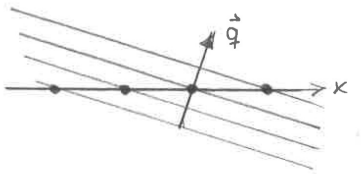
we have

$$\mathcal{P}_{\rm 1D}(k) \;=\; \frac{k}{\pi}\int_0^\infty \frac{dq}{q^2}\mathcal{P}_{\rm 3D}(q)\int_0^\infty dr\frac{2\cos kr\sin qr}{r} \;=\; k\int_k^\infty \frac{dq}{q^2}\mathcal{P}_{\rm 3D}(q)\,, \tag{3.3}$$

since (**exercise**)

$$\int_0^\infty dr\frac{2\cos kr\sin qr}{r} \;=\; \pi\Theta(q-k)\,, \tag{3.4}$$

where $\Theta(q-k)$ is the step function, 1 for $q>k$, 0 for $q<k$. Shorter wavelength $q>k$ modes in 3D contribute to the observed power in 1D at k, since when $\mathbf{q}$ is not parallel to the line, the intersection of the 3D plane wave with the line has a longer wavelength. In terms of $P_{\rm 1D}(k) = (\pi/k)\mathcal{P}_{\rm 1D}(k)$ and $P_{\rm 3D}(k) = (2\pi^2/k^3)\mathcal{P}_{\rm 3D}(k)$ this reads



$$P_{\rm 1D}(k) \;=\; \frac{1}{2\pi}\int_k^\infty \frac{dq}{q}q^2P_{\rm 3D}(q)\,. \tag{3.5}$$

This means that if $P_{\rm 3D}$ has a spectral index $n \geq -2$ ($\mathcal{P}_{\rm 3D}(q) \propto q$ or steeper), $P_{\rm 1D}(k)$ at a given scale k will be dominated by much smaller-scale (higher-k) 3D structure, and the true larger-scale 3D structure cannot be seen in a 1D survey.

From (3.5) we see that

$$\frac{dP_{\rm 1D}(k)}{dk} = -\frac{1}{2\pi}kP_{\rm 3D}(k)\,, \tag{3.6}$$

so that $P_{\rm 1D}(k)$ is necessarily monotonously decreasing. Thus we always have $n_{1D} < 0$, even if $n_{3D} > 0$.

Example: Let's redo the calculation of (3.5) in a way that is maybe easier to generalize to other cases: Start from the d-dimensional integrals

$$P_{\rm 1D}(k) \;=\; \int_{-\infty}^\infty dr\, e^{-i\mathbf{k}\cdot\mathbf{r}}\xi(\mathbf{r}) \qquad \text{and} \qquad \xi(\mathbf{r}) = \frac{1}{(2\pi)^3}\int d^3q\, e^{i\mathbf{q}\cdot\mathbf{r}}P_{\rm 3D}(q)\,, \tag{3.7}$$

so that

$$P_{\rm 1D}(k) \;=\; \frac{1}{(2\pi)^3}\int drd^3qe^{-i\mathbf{k}\cdot\mathbf{r}}e^{i\mathbf{q}\cdot\mathbf{r}}P_{\rm 3D}(q)\,, \tag{3.8}$$

where the vectors $\mathbf{r}$ and $\mathbf{k}$ lie along the 1D line (which we can take as the x axis). Thus in the exponentials $(\mathbf{q}-\mathbf{k})\cdot\mathbf{r} = (q_x-k)r$ and the r integral gives

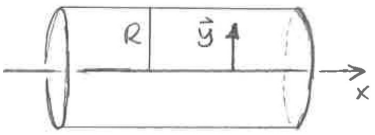
$$\frac{1}{2\pi}\int dr\, e^{i(q_x-k)r} \;=\; \delta_1^D(k-q_x)\,, \tag{3.9}$$

forcing $q_x = k$. Write now $\mathbf{q} = (k, \mathbf{h})$ where $\mathbf{h}$ is 2-dimensional. Now

$$P_{1D}(k) \;=\; \frac{1}{(2\pi)^2}\int d^2h\, P_{3D}(q) \;=\; \frac{1}{2\pi}\int hdh\, P_{3D}(q)\,. \tag{3.10}$$

Since $(q)^2 = k^2 + h^2$, $hdh = qdq$ and $q \geq k$ and we have (3.5).

For the case of discrete objects ('galaxies') we have to allow a finite thickness for the skewer (an infinitely thin line will catch zero galaxies, if they are treated as points). Consider thus a narrow cylinder[20] with radius R. The 1D density field $\delta(x, R)$ is then related to the 3D density field $\delta(x, \mathbf{y})$ by

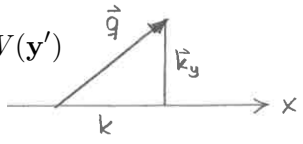


$$\delta(x, R) = \int d^2y\, \delta(x, \mathbf{y})W(\mathbf{y})\,, \tag{3.11}$$

where $W(\mathbf{y})$ is the 2D top-hat window function (we take the line as the x axis and $\mathbf{y} = (y, z)$).

Thus

$$\begin{aligned}
\xi(r) \;&=\; \langle\delta(x,R)\delta(x+r,R)\rangle \;=\; \int d^2y d^2y'\,\langle\delta(x,\mathbf{y})\delta(x+r,\mathbf{y}')\rangle W(\mathbf{y})W(\mathbf{y}') \\
&=\; \sum_{\mathbf{k}\mathbf{k}'}\int d^2y d^2y' W(\mathbf{y})W(\mathbf{y}')\langle\delta^*_{\mathbf{k}}\delta_{\mathbf{k}'}\rangle e^{-ik_x x}e^{ik'_x(x+r)}e^{-i\mathbf{k}_y\cdot\mathbf{y}}e^{i\mathbf{k}'_y\cdot\mathbf{y}'} \\
&=\; \frac{1}{V}\sum_{\mathbf{k}} e^{ik_x r}P_{3D}(k)\int d^2y\, W(\mathbf{y})e^{-i\mathbf{k}_y\cdot\mathbf{y}}\int d^2y'\, W(\mathbf{y}')e^{i\mathbf{k}_y\cdot\mathbf{y}'} \\
&=\; \frac{1}{V}\sum_{\mathbf{k}} e^{ik_x r}P_{3D}(k)|W(k_y)|^2\,.
\end{aligned} \tag{3.12}$$



We now replace the sum with an integral to get

$$\xi(r) \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} dk_x\, e^{ik_x r}\frac{1}{(2\pi)^2}\int d^2k_y\, P_{3D}(k)|W(k_y)|^2\,. \tag{3.13}$$

From here we can read the $P_{1D}(k)$ as the Fourier transform of $\xi(r)$:

$$P_{1D}(k) \;=\; \frac{1}{(2\pi)^2}\int d^2k_y\, P_{3D}(q)|W(k_y)|^2\,, \tag{3.14}$$

where $\mathbf{q} = (k, \mathbf{k}_y)$. Comparing to (3.10), we see that $P_{3D}(q)$ was replaced by $P_{3D}(q)|W(k_y)|^2$, where $k_y^2 = q^2 - k^2$. The result is thus

$$P_{1D}(k) \;=\; \frac{1}{2\pi}\int_k^{\infty}\frac{dq}{q}q^2P_{3D}(q)\left|W\left(\sqrt{q^2-k^2}\right)\right|^2\,. \tag{3.15}$$

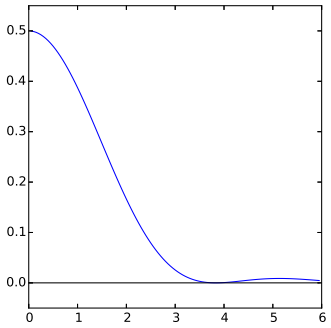


Figure 11: $2J_1^2(x)/x^2$

Inserting the Fourier transform of the 2D top-hat window function and writing in terms of $\mathcal{P}_{3D}$ not P_{3D}, this is

$$P_{1D}(k) \;=\; \pi\int_k^{\infty}\frac{dq}{q^2}\mathcal{P}_{3D}(q)\frac{2J_1^2(\sqrt{q^2-k^2}R)}{(q^2-k^2)R^2}\,. \tag{3.16}$$

[20] Concerning the observational situation a narrow cone would be a more appropriate geometry, but it ruins the homogeneity along the line (we won't get rid of x as we do now when going from the second to the third line of Eq. 3.12) and the problem gets complicated.

The Bessel function $J_1(x)$ is much like $\sin x$ except for the $x^{-1/2}$ damping. For small x, $J_1(x) \approx x/2$; it reaches its maximum at $x \approx 2$, where $J_1(x) \approx 0.2$, becomes zero again at $x \approx 3.8$ and then keeps oscillating with a damped amplitude. Thus the window function is roughly constant at first (the factor multiplying $\mathcal{P}_{3D}(q)$ in (3.16) is $\approx 1/2$) , and then falls rapidly after $q^2 - k^2 > (2/R)^2$.

Example: Peacock ([1], p. 518) mentions the $2000\,h^{-1}$Mpc long pencil-beam survey of Broadhurst et al.[16], which generated a lot of excitement[17], since it found a periodic structure of wavelength $128\,h^{-1}$Mpc, i.e., the 1D power spectrum was strongly peaked at $k = 2\pi/128\,h^{-1}\mathrm{Mpc} = 0.05\,h\mathrm{Mpc}^{-1}$. This seemed to contradict the standard view of large-scale structure, according to which, $\sigma^2(R)$ and $\mathcal{P}_{3D}(k)$ should be $\ll 1$ at these scales. However, considering the above, large 1D power at such a large scale is not unexpected. The standard view is that at smaller scales $\gamma = 1.8$, so that $n = -1.2$ and $\mathcal{P}_{3D} \propto k^\gamma$, reaching unity near $k = 0.1\,h\mathrm{Mpc}^{-1}$ (see Fig. 3). The representative transverse radius of the pencil beam was $R = 3\,h^{-1}$Mpc. Thus the observed 1D power at all scales larger than this is dominated by 3D structure at this scale. Approximating the window function with a step function that cuts off at $q^2 - k^2 = 4/R$, i.e., at $q = \sqrt{k^2 + 4/R^2}$ (3.16) becomes

$$P_{1D}(k) \;\approx\; \frac{\pi}{2}\int_k^{\sqrt{k^2+4/R^2}} \frac{dq}{q^2}\left(\frac{q}{0.1\,h\mathrm{Mpc}^{-1}}\right)^{1.8} \;=\; \left[\left(k^2 + 4/R^2\right)^{0.4} - k^{0.8}\right] 124\,h^{-1}\mathrm{Mpc}\,, \qquad (3.17)$$

where in the last step k and $1/R$ are in units of $h\mathrm{Mpc}^{-1}$. The expression in brackets is $\approx$ *const* for large scales $k \ll 2/R = 0.7\,h\mathrm{Mpc}^{-1}$ (it is 0.72 for $k = 0$ and 0.63 at $k = 0.05\,h\mathrm{Mpc}^{-1}$) and then becomes smaller as $k \to 2/R$ and beyond. At $k = 2\pi/128\,h^{-1}\mathrm{Mpc} = 0.05\,h\mathrm{Mpc}^{-1}$ this gives $\mathcal{P}_{1D}(k) = \frac{k}{\pi}P_{1D}(k) \approx (0.05/\pi) \times 0.63 \times 124 \approx 1.3$.

3.2 Slices

Left as an exercise.

Exercise: Consider a 2-dimensional slice of a 3-dimensional density field. Show that

$$\mathcal{P}_{2D}(k) \;=\; k^2 \int_k^\infty \frac{dk'}{\sqrt{(k')^2 - k^2}} \frac{1}{(k')^2}\mathcal{P}_{3D}(k')\,. \qquad (3.18)$$

Exercise: Consider instead a slice with thickness $2R$, i.e, we have a 1D top-hat window function $W_T(z;R)$ in the orthogonal direction. Show that

$$\mathcal{P}_{2D}(k) \;=\; k^2 \int_k^\infty \frac{dk'}{(k')^2\sqrt{(k')^2 - k^2}} \mathcal{P}_{3D}(k') \left|W\left(\sqrt{(k')^2 - k^2}\right)\right|^2 . \qquad (3.19)$$

The result (3.18) diverges at high k' for $\mathcal{P}_{3D}(k') \propto (k')^2$ or steeper, i.e, for $n \geq -1$. In (3.19) the finite thickness of the slice cuts off the contribution from the smallest scales, so that for $n \geq -1$ the large-scale 2D power is dominated by the structure near the scale R.

4 Angular correlation function for small angles

Suppose we have a galaxy survey with no distance (or redshift) information. For every galaxy we then have only the direction to it, which we denote by a unit vector $\hat{\mathbf{r}}$. We define the *angular correlation function* $w(\vartheta)$ as the excess probability of finding another galaxy at a separation angle ϑ from a given galaxy. In analogy with (2.10) we have

$$\langle n(\hat{\mathbf{r}}_1)n(\hat{\mathbf{r}}_2)\rangle d\Omega_1 d\Omega_2 \;\equiv\; \langle n\rangle^2\,[1+w(\vartheta)]\,d\Omega_1 d\Omega_2\,, \tag{4.1}$$

where $n(\hat{\mathbf{r}})$ is the number density of galaxies per unit solid angle ($d\Omega$) in the direction $\hat{\mathbf{r}}$ on the sky, and ϑ is the angle between directions $\hat{\mathbf{r}}_1$ and $\hat{\mathbf{r}}_2$.

4.1 Relation to the 3D correlation function

How is this related to the 3D distribution of galaxies? It is essential here that a galaxy survey does not see all galaxies that lie in a given direction. The further out a galaxy is the larger its absolute luminosity has to be for it to be observed. This is represented by the *selection function* $S(r)$ of the survey, which gives the *probability for a galaxy at distance r to be observed by the survey.*

Since the volume element is given by $dV = r^2 dr d\Omega$, $n(\hat{\mathbf{r}})$ is related to the 3D galaxy number density by

$$n(\hat{\mathbf{r}}) \;=\; \int_0^\infty \rho(\mathbf{r})S(r)r^2\,dr \tag{4.2}$$

and

$$\langle n\rangle \;=\; \langle\rho\rangle\int_0^\infty S(r)r^2\,dr \;\equiv\; \langle\rho\rangle V_S\,, \tag{4.3}$$

where

$$V_S \;\equiv\; \int_0^\infty S(r)r^2\,dr\,, \tag{4.4}$$

is the selection function volume per unit solid angle,[21] For a survey covering a solid angle Ω, the expected number of observed galaxies is thus

$$\langle N\rangle \;=\; \langle n\rangle\Omega \;=\; \langle\rho\rangle V_S\Omega\,. \tag{4.5}$$

For density perturbations, the relation between $\delta n(\hat{\mathbf{r}}) \equiv (n(\hat{\mathbf{r}}) - \langle n\rangle)/\langle n\rangle$ and the 3D $\delta(\mathbf{r}) \equiv (\rho(\mathbf{r}) - \langle\rho\rangle)/\langle\rho\rangle$ is

$$\frac{\delta n}{\langle n\rangle} \;=\; \frac{1}{V_S}\int_0^\infty \delta(\mathbf{r})S(r)r^2 dr\,. \tag{4.6}$$

Thus

$$\begin{aligned}
\langle n(\hat{\mathbf{r}}_1)n(\hat{\mathbf{r}}_2)\rangle d\Omega_1 d\Omega_2 &= \int_0^\infty dr_1\int_0^\infty dr_2\,\langle\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\rangle\,S(r_1)S(r_2)r_1^2r_2^2\,d\Omega_1 d\Omega_2\\
&= \langle\rho\rangle^2\int_0^\infty dr_1\int_0^\infty dr_2\,[1+\xi(r_{12})]\,S(r_1)S(r_2)r_1^2r_2^2\,d\Omega_1 d\Omega_2\\
&= \langle\rho\rangle^2\int_0^\infty dr_1 S(r_1)r_1^2 d\Omega_1\int_0^\infty dr_2 S(r_2)r_2^2 d\Omega_2\\
&\quad + \langle\rho\rangle^2\int_0^\infty dr_1\int_0^\infty dr_2\,\xi(r_{12})S(r_1)S(r_2)r_1^2r_2^2\,d\Omega_1 d\Omega_2\\
&\equiv \langle n\rangle^2 d\Omega_1 d\Omega_2 + \langle n\rangle^2 w(\vartheta)\,d\Omega_1 d\Omega_2\,,
\end{aligned} \tag{4.7}$$

[21] This V_S is my own definition (MBW[2] keeps writing the integral without calling it anything) and there are different normalizations of $S(r)$ (Peacock[1] normalizes it so that $V_S = 1$.)

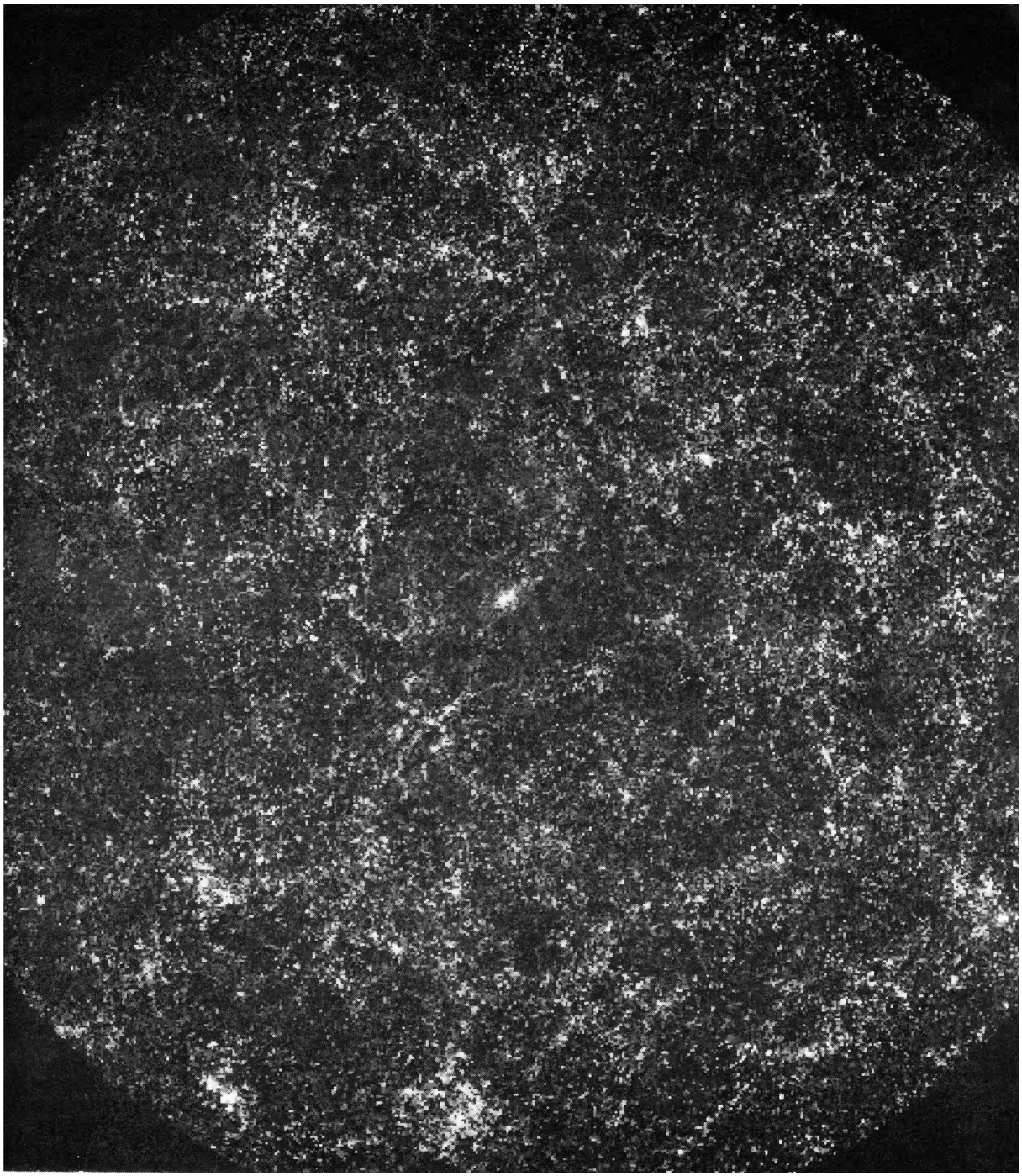


Figure 12: Shane-Wirtanen map of the galaxy distribution on the northern galactic hemisphere [21, 22]. The sky is divided into cells of $10' \times 10'$ (arcminutes) and each cell is colored in greyscale according to the number of galaxies in it, white signifying ≥ 8 galaxies. The galaxies were counted from the Lick survey [21], which has a limiting magnitude of about $m_* = 19$. The Lick survey, begun in 1950 and completed in 1967, was the largest galaxy survey of its time, containing about a million galaxies, and a major step in the study of the large-scale structure of the universe. It covered the northern galactic hemisphere and about half of the southern galactic hemisphere. The survey used $6^\circ \times 6^\circ$ photographic plates obtained with the 0.5 m refractor (lens telescope) at Lick Observatory (on a mountain top in the San Francisco Bay area). At center, near the north galactic pole: Coma Cluster. This is Fig. 3 from [23].

so that we have the relation between the angular and 3D correlation functions

$$w(\vartheta) \;=\; \frac{\langle\rho\rangle^2}{\langle n\rangle^2}\int_0^\infty dr_1\int_0^\infty dr_2\,\xi(r_{12})S(r_1)S(r_2)r_1^2r_2^2\,, \tag{4.8}$$

where

$$r_{12} \;=\; \sqrt{(\mathbf{r}_1-\mathbf{r}_2)^2} \;=\; \sqrt{r_1^2+r_2^2-2r_1r_2\cos\vartheta}\,. \tag{4.9}$$

4.1.1 Selection function

For simplicity, we now ignore all cosmological effects, like the difference between the luminosity distance and comoving distance, and the time evolution of galaxies and their clustering, and assume that the survey is *magnitude limited*, i.e., within the survey area it observes all galaxies whose (apparent) magnitude is $< m_*$, where m_* is the limiting magnitude. The absolute magnitude M and apparent magnitude m are defined as logarithmic scales

$$m \equiv -2.5\lg\frac{l}{l_0}\,, \qquad M \equiv -2.5\lg\frac{L}{L_0}\,, \tag{4.10}$$

where l is the apparent luminosity (flux density, received power per unit area), L is the absolute luminosity (radiated power), and l_0, L_0 are reference luminosities chosen so that $m = M$ for an object 10 pc away. There are actually different magnitude scales for different bands of the electromagnetic spectrum, with different reference luminosities. The *bolometric* scale refers to the entire spectrum, and for it $L_0 = 3.0\times10^{28}\,\mathrm{W}$. Thus L and M are related by

$$L = L_0\times10^{-0.4M}\,. \tag{4.11}$$

The magnitudes m and M and the distance r of a galaxy[22] are related by

$$m-M \;=\; -5+5\lg r(\mathrm{pc}) \quad\Rightarrow\quad r = 10^{(m-M)/5}\times10\ \mathrm{pc}\,. \tag{4.12}$$

Thus the galaxy can be observed if

$$M \;<\; m_*+5-5\lg r(\mathrm{pc})\,. \tag{4.13}$$

Exercise: Find the value of l_0 in $\mathrm{W/m^2}$ in the bolometric scale.

For example, if all galaxies had $M = -20$ (the Andromeda galaxy has $M = -20.3$ [24], so this is typical for a spiral galaxy), then all galaxies closer than $d_* = 10^{(m_*/5)+4}\times10\ \mathrm{pc} = 10^{m_*/5}\times0.1\,\mathrm{Mpc}$ would be observed and none beyond, so that the selection function would be a step function

$$S(r) \;=\; \Theta(d_*-r) \;=\; \Theta(1-r/d_*)\,, \tag{4.14}$$

where d_* is the survey depth. For $m_* = 20$, this gives $d_* = 1\,\mathrm{Gpc}$ (the cosmological effects are already becoming significant at this distance). The volume of this selection function is $V_S = d_*^3/3$. (For a full-sky survey, $\Omega = 4\pi$, so $V_S\Omega$ is the familiar $4\pi d_*^3/3$.)

In reality, different galaxies have different M, and the selection function is related to the galaxy luminosity function $\phi(L)$ (the probability distribution of galaxy luminosities L), as

$$S(r) \;=\; \int_{L_*(r)}^\infty \phi(L)dL \Big/ \int_0^\infty \phi(L)dL\,, \tag{4.15}$$

[22] Actually, the luminosity distance.

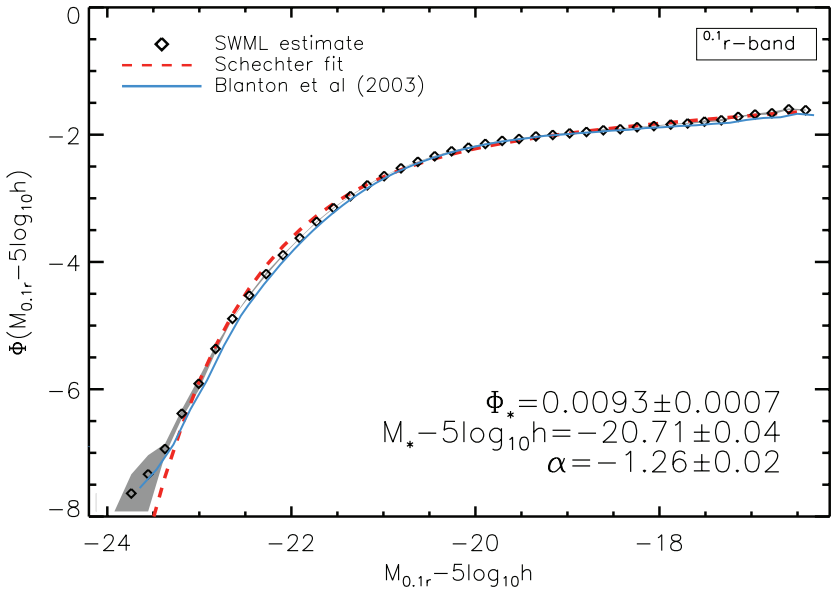


Figure 13: Luminosity function ϕ from the SDSS survey r band ("red" light, about 550–680 nm) for galaxies with redshift $z = 0.02$–0.22.. g The SDSS luminosity functions are significantly differ for different wavelength bands: typically the galaxies have fainter magnitudes (larger M value) for longer wavelengths (i, "near infrared", and z, "infrared" bands) and brighter magnitudes (more negative M value) for shorter wavelengths (g, "green", and u, "ultraviolet"). For $h = 0.7$, $-5 \lg h = 0.775$. This is Fig. 7 from [25].

where

$$L_*(r) = L_0 \times 10^{-0.4m_* - 2 + 2 \lg r(\mathrm{pc})} \tag{4.16}$$

is the smallest luminosity observable at distance r. The denominator can be omitted if $\phi(L)$ is normalized so that this integral equals 1, but apparently this is not necessarily the case.

We can still define a survey depth $d_* \propto 10^{m_*/5}$ as the maximum distance to which some reference magnitude M is observed. Within the approximation of ignoring cosmological effects (so the luminosity function does not evolve in time), the shape of the selection function is independent of m_*, except it is stretched proportional to d_*, so we can write it as

$$S(r) \equiv f(r/d_*) \tag{4.17}$$

where f is a function determined by $\phi(L)$ and independent of the survey. In general, $V_S \propto d_*^3$.

In the above we ignored the difference between different cosmological distance concepts. The distance in (4.12) is really the luminosity distance d^L, which in a flat cosmology is related to the comoving distance d^C by $d^L = (1 + z)d^C$. Using a reference magnitude $M = -22$ (most galaxies are less luminous than this, but there are enough galaxies with this magnitude to be used for statistics) we find for the Lick survey, with limiting magnitude $m_* = 19$, the depth $d_*^L = 1.58\,\mathrm{Gpc}$ and for the Euclid survey, with limiting magnitude $m_* = 24.5$, the depth $d_*^L = 19.95\,\mathrm{Gpc}$. Conversion to comoving distance requires the knowledge of the corresponding redshift, which depends on assumed cosmology and is a bit complicated. Assuming a flat cosmology with $\Omega_m = 0.3$ and $h = 0.7$ and interpolating the redshift from Table 2 given at end of Sec. 3 of Cosmology II we get Table 3.

4.1.2 Small-angle limit

Assume now that the survey depth d_* is large compared to the separations r_{12} over which there are significant correlations $\xi(r_{12})$, so that in the double integral (4.8) we get a significant

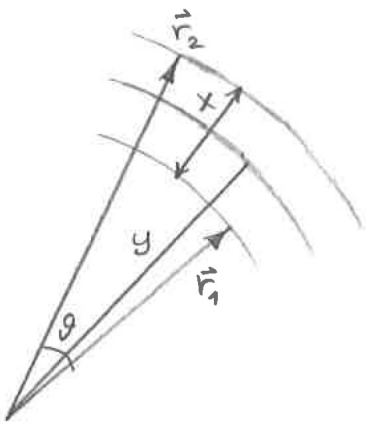
survey	m_*	d_*^L	z_*	d_*^C
Lick	19	1.58 Gpc	0.30	1.2 Gpc
Euclid	24.5	19.95 Gpc	2.45	5.8 Gpc

Table 3: Depths of the Lick and Euclid surveys for reference magnitude $M = -22$ assuming flat cosmology with $\Omega_m = 0.3$ and $h = 0.7$. For comparison, the comoving distance to the last scattering surface, from which the CMB originates, is 13.6 Gpc in this cosmology.

contribution only when $r_{12} \ll d_*$). This then means that $w(\vartheta)$ is negligible except for small angles $\vartheta \ll 1$. (Clearly "significant", "negligible", and "small" here are relative terms. The larger and better the survey, the higher accuracy we can aim at.) We can then do the integral (4.8). Note that it is a two-dimensional integral over two distances r_1, r_2. The $\xi(r_{12})$ in the integral depends also on the angle ϑ via (4.9) but we are doing the integral for a fixed ϑ. Define now

$$y \equiv \tfrac{1}{2}(r_1 + r_2) \qquad \text{and} \qquad x \equiv r_2 - r_1 \tag{4.18}$$

(the mean distance and distance difference for a pair of coordinate vectors $\mathbf{r}_1$, $\mathbf{r}_2$), and do a coordinate transformation from (r_1, r_2) to (x, y). The Jacobian determinant of this coordinate transformation is -1 (**exercise**), so that $dr_1 dr_2 = dx dy$.



Since $x \le r_{12}$, the preceding assumption means that we get a significant contribution to (4.8) only when x is small, otherwise $\xi \sim 0$. Thus the integration limits on x do not matter and we can replace

$$\int_0^\infty dr_1 \int_0^\infty dr_2 \qquad \text{by} \qquad \int_0^\infty dy \int_{-\infty}^\infty dx \tag{4.19}$$

The selection function will change slowly over distances $\ll d_*$, so we can approximate

$$S(r_1)S(r_2) \;=\; S(y + \tfrac{1}{2}x)S(y - \tfrac{1}{2}x) \;\approx\; S(y)^2 \,. \tag{4.20}$$

Making further appropriate small angle approximations for $\cos\vartheta$ where it occurs, we get (**exercise**)

$$\boxed{w(\vartheta) \;\approx\; \frac{1}{V_S^2} \int_0^\infty dy\, y^4 S^2(y) \int_{-\infty}^\infty dx\, \xi\left(\sqrt{x^2 + y^2\vartheta^2}\right) .} \tag{4.21}$$

This is known as *Limber's equation* (it is Eq. 16.100 in [1] and Eq. 6.166 in [2]; note that my notation is a compromise between these two).

What is the effect of the depth of the survey on $w(\vartheta)$? Define a scaled distance $s \equiv y/d_*$ so that $y = d_* s$. Then (4.21) becomes

$$w(\vartheta) \;\approx\; \frac{1}{V_S^2} d_*^5 \int_0^\infty ds\, s^4 f^2(s) \int_{-\infty}^\infty dx\, \xi\left(\sqrt{x^2 + d_*^2 s^2 \vartheta^2}\right) \;=\; \frac{1}{d_*} \times \mathrm{F}(d_*\vartheta)\,, \tag{4.22}$$

where the function F is independent of d_*. Thus increasing the depth of the survey shifts the correlations to smaller angles and reduces their amplitude. The first effect is because at larger distances the same physical scales (at which there are correlations) are seen in a smaller angle. The second effect is because the structures at different distances that we see on top of each other are uncorrelated and thus tend to cancel each other in $w(\vartheta)$.

Peacock ([1], p. 521) makes the comment that observing this effect of increasing the depth of the survey was important in demonstrating that the observed structure in such surveys without distance information (e.g., The Lick survey, Fig. 12) was real. Such surveys suffer from effects like anisotropy of the selection function due to, e.g., foreground extinction, that would show as fake structure if not corrected for (which might not be possible).

4.1.3 Power law

Assume now that the 3D correlation function has a power-law form

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma} . \tag{4.23}$$

We expect that $w(\vartheta)$ should have a shallower slope, since for smaller scales we have more layers of structure piled on top of each other, so the canceling effect should be bigger.

Limber's equation (4.21) becomes (**exercise**)

$$\begin{aligned} w(\vartheta) &\approx \frac{r_0^\gamma}{V_S^2} \int_0^\infty dy\, y^4 S^2(y) \int_{-\infty}^\infty dx\, \left(x^2 + y^2\vartheta^2\right)^{-\gamma/2} \\ &= \frac{2r_0^\gamma \vartheta^{1-\gamma}}{V_S^2} \int_0^\infty dx\, (1+x^2)^{-\gamma/2} \int_0^\infty dy\, y^{5-\gamma} S^2(y) \\ &= r_0^\gamma \vartheta^{1-\gamma} \frac{\sqrt{\pi}\,\Gamma(\frac{1}{2}(\gamma-1))}{\Gamma(\frac{1}{2}\gamma)} \frac{1}{V_S^2} \int_0^\infty dy\, y^{5-\gamma} S^2(y)\,. \end{aligned} \tag{4.24}$$

Thus we get a power-law form, but with a shallower slope, $\gamma - 1$, also for the angular correlation function:

$$w(\vartheta) = \left(\frac{\vartheta}{\vartheta_0}\right)^{1-\gamma} . \tag{4.25}$$

For $\gamma = 1.8$ the slope of $w(\vartheta)$ is thus 0.8. The angle ϑ_0 where $w(\vartheta) = 1$ depends on the survey depth via the selection function $S(y)$.

Exercise: Find how ϑ_0 changes as a function of survey depth d_*, i.e., by what factor does ϑ_0 change if d_* is increased by a factor D.

Exercise: For a step-function selection function and power-law correlation function with $\gamma = 9/5$, find ϑ_0 in terms of r_0 and d_*. Give the numerical value in arcmin for $r_0 = 5\,h^{-1}$Mpc and $d_* = 1000\,h^{-1}$Mpc.

4.2 Power spectrum for flat sky

If the survey area is small enough (i.e., the maximum ϑ within the area is $\ll 1$)[23] we can treat the sky in that area as a flat 2D Euclidean space. To specify coordinates on that space we can then use a 2D vector $\boldsymbol{\theta} = (\theta_x, \theta_y)$ instead of a 3D unit vector $\hat{\mathbf{r}}$. Thus $\boldsymbol{\theta} = 0$ at the center of the survey (or whatever we choose as origin), and we choose two orthogonal directions (e.g., west and north, or east and south) to specify the components θ_x and θ_y. They are given in units of degrees or radians, or rather arcminutes or milliradians. We can then define a 2D correlation function $w(\boldsymbol{\theta})$, which is the same quantity as the $w(\vartheta)$ defined earlier, with $\vartheta = |\boldsymbol{\theta}| \ll 1$.

If the survey area on this flat sky is a square, we can import the earlier results (1.83b and 1.84b) for the 2D correlation function and power spectrum from Sec. 1.6:

$$\begin{aligned} P_\vartheta(K) &= \int_0^\infty w(\vartheta) J_0(K\vartheta)\, 2\pi\vartheta d\vartheta \qquad \text{or} \qquad \mathcal{P}_\vartheta(K) = K^2 \int_0^\infty w(\vartheta) J_0(K\vartheta)\vartheta d\vartheta \\ w(\vartheta) &= \frac{1}{2\pi} \int_0^\infty P_\vartheta(K) J_0(K\vartheta) K dK \quad = \int_0^\infty \mathcal{P}_\vartheta(K) J_0(K\vartheta) \frac{dK}{K} \end{aligned} \tag{4.26}$$

[23] This can also be applied to larger survey areas, as long as we are only interested in measuring the correlation function at small scales and the power spectrum at high K. We can, e.g., divide the larger survey area into several small squares and apply the methods of this section to each of them separately and taking in the end the average of all the $P_\vartheta(K)$ from the different squares.

with some change in notation.

For a power-law correlation function,

$$w(\vartheta) = \left(\frac{\vartheta}{\vartheta_0}\right)^{-\varepsilon} \tag{4.27}$$

we thus get a power-law power spectrum, using the result (1.114),

$$\begin{aligned} P_\vartheta(K) &= \frac{2\pi}{K^2} 2^{1-\varepsilon} \frac{\Gamma(1-\frac{1}{2}\varepsilon)}{\Gamma(\frac{1}{2}\varepsilon)} (K\vartheta_0)^\varepsilon = \frac{2\pi}{K^2} 2^{2-\gamma} \frac{\Gamma(\frac{3}{2}-\frac{1}{2}\gamma)}{\Gamma(\frac{1}{2}\gamma-\frac{1}{2})} (K\vartheta_0)^{\gamma-1} \qquad (\tfrac{1}{2} < \varepsilon < 2) \\ \mathcal{P}_\vartheta(K) &= 2^{1-\varepsilon} \frac{\Gamma(1-\frac{1}{2}\varepsilon)}{\Gamma(\frac{1}{2}\varepsilon)} (K\vartheta_0)^\varepsilon = 2^{2-\gamma} \frac{\Gamma(\frac{3}{2}-\frac{1}{2}\gamma)}{\Gamma(\frac{1}{2}\gamma-\frac{1}{2})} (K\vartheta_0)^{\gamma-1}, \end{aligned} \tag{4.28}$$

where the second forms assume $\varepsilon = \gamma - 1$. For $\varepsilon = 0.8$,

$$\mathcal{P}_\vartheta(K) = 0.7712 (K\vartheta_0)^{0.8}. \tag{4.29}$$

4.2.1 Relation to the 3D power spectrum

Write the $\xi(\sqrt{x^2+y^2\vartheta^2})$ in Limber's equation (4.21) in terms of the 3D power spectrum,

$$\int_{-\infty}^{\infty} dx\, \xi\left(\sqrt{x^2+y^2\vartheta^2}\right) = \int_0^\infty \frac{dk}{k} \mathcal{P}(k) \int_{-\infty}^{\infty} dx\, \frac{\sin k\sqrt{x^2+y^2\vartheta^2}}{k\sqrt{x^2+y^2\vartheta^2}}. \tag{4.30}$$

The x integral can be written as

$$\frac{2}{k} \int_0^\infty dx \frac{\sin\sqrt{x^2+a^2}}{\sqrt{x^2+a^2}}, \tag{4.31}$$

where $a = ky\vartheta$. Wolfram Alpha couldn't do this. Gradshteyn&Ryzhik [9] 3.876.1 says

$$\int_0^\infty \frac{\sin(p\sqrt{x^2+a^2})\cos bx}{\sqrt{x^2+a^2}} dx = \frac{\pi}{2} J_0\left(a\sqrt{p^2-b^2}\right) \qquad \text{for } 0 < b < p, \tag{4.32}$$

but here we have $b = 0$ (so that $\cos bx = 1$). There's no reason to think that the result would not hold also for $b = 0$, since the behavior of the integrand or the result is not problematic as $b \to 0$. Thus we take the x integral to give $(\pi/k)J_0(ky\vartheta)$, so that Limber's equation becomes

$$w(\vartheta) \approx \frac{1}{V_S^2} \int_0^\infty dy\, y^4 S^2(y) \int_0^\infty \frac{dk}{k^2} \pi \mathcal{P}(k) J_0(ky\vartheta). \tag{4.33}$$

Writing $k = K/y$ this becomes

$$w(\vartheta) \approx \frac{1}{V_S^2} \int_0^\infty dy\, y^5 S^2(y) \int_0^\infty \frac{dK}{K^2} \pi \mathcal{P}(K/y) J_0(K\vartheta). \tag{4.34}$$

Comparing (4.33) to (4.26) we can read off

$$\boxed{\mathcal{P}_\vartheta(K) = \frac{1}{V_S^2} \frac{\pi}{K} \int_0^\infty dy\, y^5 S^2(y) \mathcal{P}(K/y),} \tag{4.35}$$

the power version of Limber's equation.

Peacock[1], p. 520, comments: "This is just a convolution in log space, and is considerably simpler to evaluate and interpret than the w–ξ version (4.21)". That is, we consider S and $\mathcal{P}$ as

functions of $\ln y$ and $\ln k$ (mathematically, these are different functions, $S \circ \exp$ and $\mathcal{P} \circ \exp$, but as is common in physics, we use the same symbol for the same quantity, even when we consider it as a function of a different variable), so that

$$\mathcal{P}_\vartheta(\ln K) \;=\; \frac{1}{V_S^2}\frac{\pi}{K}\int_{-\infty}^{\infty} e^{6\ln y} S^2(\ln y)\mathcal{P}(\ln K - \ln y) d\ln y\,. \tag{4.36}$$

Exercise: Show that if we increase the depth of the survey by a factor D, the angular power spectrum $\tilde{\mathcal{P}}_\vartheta(K)$ of the deeper survey is related to the $\mathcal{P}_\vartheta(K)$ of the shallower survey by

$$\tilde{\mathcal{P}}_\vartheta(K) \;=\; \frac{1}{D}\mathcal{P}_\vartheta(K/D)\,. \tag{4.37}$$

When we apply (4.35) to a power spectrum $\mathcal{P}(k)$ from cosmological theory, which may evolve in time, i.e., $\mathcal{P}(k,t)$, the for each y we just use $\mathcal{P}(K/y, t(y))$, where $t(y)$ is the local time at distance y.

5 Spherical sky

For correlations over larger angles on the sky we have to take into account that the geometry of the sky is a sphere. So now our 2D space is not Euclidian; it is a sphere. This makes the math more difficult but there's one benefit: when considering the full sky we do not have to worry about boundary conditions or edge effects. In practice this benefit is lost, since there are no truly full-sky surveys of galaxies; at least the galactic plane region is obscured by the Milky Way. However, we begin by considering the full sky.

To specify a point on the sphere, we can use unit vectors

$$\hat{\mathbf{r}} = (x, y, z) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \tag{5.1}$$

or spherical coordinates θ, ϕ. The spherical coordinate system is singular at $\theta = 0$ ('North Pole') and $\theta = \pi$ ('South Pole'), so that ϕ is unspecified there. Note that the 'latitude' (altitude/elevation a, declination δ, ecliptic latitude β, galactic latitude b) of traditional astronomical coordinate systems (horizontal, equatorial, ecliptic, galactic) is $\pi/2 - \theta$, so that it is zero at the equator/ecliptic etc.; θ is sometimes called 'colatitude'. (In these systems ϕ is called azimuth A[24], right ascension α, ecliptic longitude λ, galactic longitude l.)

5.1 Angular correlation function and angular power spectrum

Instead of a Fourier expansion in plane waves, we now expand the density perturbation on the sky in spherical harmonics:

$$\delta(\hat{\mathbf{r}}) \;=\; \delta(\theta,\phi) \;=\; \sum_{\ell m} a_{\ell m} Y_{\ell m}(\theta,\varphi)\,, \tag{5.2}$$

where the sum goes over $\ell = 0, 1, 2, \ldots, \infty$ and $m = 0, \pm1, \pm2, \ldots, \pm\ell$, and the $Y_{\ell m}$ are spherical harmonics, functions that form an orthonormal and complete set on the sphere. The spherical harmonic coefficients

$$a_{\ell m} \;=\; \int d\Omega\, Y^*_{\ell m}(\hat{\mathbf{r}})\delta(\hat{\mathbf{r}}) \tag{5.3}$$

take now the role the Fourier coefficients $\delta_{\mathbf{k}}$ had for Euclidean space. They are complex numbers, but from the reality of δ follows that

$$a_{\ell,-m} = (-1)^m a^*_{\ell m}\,, \tag{5.4}$$

so that for each ℓ there are $2\ell + 1$ real degrees of freedom.

From statistical homogeneity and isotropy[25] follows that the "theoretical" (i.e., ensemble average) correlation function

$$w(\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2) \;\equiv\; \langle\delta(\hat{\mathbf{r}}_1)\delta(\hat{\mathbf{r}}_2)\rangle \;=\; w(\vartheta) \tag{5.5}$$

may depend only on the angle ϑ between the directions $\hat{\mathbf{r}}_1$ and $\hat{\mathbf{r}}_2$.

Analogous to our proof (1.74), we can show (done at end of Sec. 5.5) that from statistical homogeneity and isotropy follows that

$$\langle a^*_{\ell m} a_{\ell' m'}\rangle \;=\; \delta_{\ell\ell'}\delta_{mm'} 2\pi \int_{-1}^{1} d\cos\vartheta\, L_\ell(\cos\vartheta) w(\vartheta) \;\equiv\; \delta_{\ell\ell'}\delta_{mm'} C_\ell \tag{5.6}$$

[24] Often the azimuth is defined as $A = -\phi$.

[25] In this context one usually talks only about "isotropy", since the role of locations $\mathbf{r}$ is now taken by directions $\hat{\mathbf{r}}$, but we still have two different aspects here: one is that all directions $\hat{\mathbf{r}}$ from us to the sky are equal, and the other is that all directions "along" the sky, i.e., how the angle between $\hat{\mathbf{r}}_1$ and $\hat{\mathbf{r}}_2$ is oriented, are equal.

where the

$$C_\ell \;=\; \langle |a_{\ell m}|^2 \rangle \;=\; 2\pi \int_{-1}^{1} d\cos\vartheta \, L_\ell(\cos\vartheta) w(\vartheta) \tag{5.7}$$

is the *angular power spectrum* and the L_ℓ are Legendre polynomials.[26] Note that the expectation value $\langle |a_{\ell m}|^2 \rangle$ is independent of m. The index ℓ is called the *multipole number* and it corresponds to angular scale, whereas the index m corresponds to different orientations and patterns with the same angular scale; i.e, ℓ corresponds to k and m corresponds to the direction $\hat{\mathbf{k}}$ in a Fourier expansion.

Thus

$$\begin{aligned} w(\vartheta) \;&\equiv\; \langle \delta(\hat{\mathbf{r}}_1)\delta(\hat{\mathbf{r}}_2) \rangle \;=\; \sum_{\ell m}\sum_{\ell' m'} \langle a^*_{\ell m} a_{\ell' m'} \rangle Y^*_{\ell m}(\hat{\mathbf{r}}_1) Y_{\ell' m'}(\hat{\mathbf{r}}_2) \\ &=\; \sum_\ell C_\ell \sum_m Y^*_{\ell m}(\hat{\mathbf{r}}_1) Y_{\ell m}(\hat{\mathbf{r}}_2) \;=\; \frac{1}{4\pi}\sum_\ell (2\ell+1) C_\ell L_\ell(\cos\vartheta)\,. \end{aligned} \tag{5.8}$$

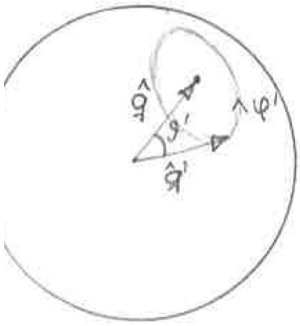
using (5.6) and (5.28). (One can show this also by using (5.7) and the completeness of Legendre polynomials (**exercise**).)

For a single realization we can define an "observed correlation function" $\widehat{w}(\vartheta)$ as the sky average[27]

$$\widehat{w}(\vartheta) \equiv \langle \delta(\hat{\mathbf{r}})\delta(\hat{\mathbf{r}}') \rangle_{\rm sky} \tag{5.9}$$

where the $\langle\cdot\rangle_{\rm sky}$ refers to average over all pairs $\hat{\mathbf{r}}$,$\hat{\mathbf{r}}'$ separated by angle ϑ. Pick first a reference direction $\hat{\mathbf{r}}$ and integrate over the circle of directions $\hat{\mathbf{r}}'$ that are ϑ away from $\hat{\mathbf{r}}$:

$$\widehat{w}(\hat{\mathbf{r}},\vartheta) \;\equiv\; \frac{1}{2\pi}\int d\varphi'\, \delta(\hat{\mathbf{r}})\delta(\hat{\mathbf{r}}')\,. \tag{5.10}$$



Then integrate this over the whole sky (the $d\Omega$ integration is over $\hat{\mathbf{r}}$):

$$\widehat{w}(\vartheta) \;\equiv\; \frac{1}{4\pi}\int d\Omega\, \widehat{w}(\hat{\mathbf{r}},\vartheta) \;=\; \frac{1}{8\pi^2}\int d\Omega \int d\varphi' \delta(\hat{\mathbf{r}})\delta(\hat{\mathbf{r}}')\,. \tag{5.11}$$

While the expectation value of $|a_{\ell m}|^2$ is the same for all m, the actual realized values are different. Define the "observed" angular power spectrum $\widehat{C}_\ell$ as their average:

$$\widehat{C}_\ell \;\equiv\; \frac{1}{2\ell+1}\sum_m |a_{\ell m}|^2\,. \tag{5.12}$$

We want to relate $\widehat{w}$ and $\widehat{C}_\ell$ the same way we did for the theoretical w and C_ℓ in (5.7) and (5.8). Starting from (5.12),

$$\begin{aligned} \widehat{C}_\ell \;&\equiv\; \frac{1}{2\ell+1}\sum_m |a_{\ell m}|^2 \;=\; \int d\Omega d\Omega'\, \delta(\mathbf{r})\delta(\mathbf{r}') \frac{1}{2\ell+1}\sum_m Y^*_{\ell m}(\hat{\mathbf{r}}) Y_{\ell m}(\hat{\mathbf{r}}') \\ &=\; \frac{1}{4\pi}\int d\Omega \int d\Omega'\, \delta(\mathbf{r})\delta(\mathbf{r}') L_\ell(\cos\vartheta') \;=\; \int d\cos\vartheta'\, L_\ell(\cos\vartheta') \frac{1}{4\pi}\int d\Omega d\varphi'\, \delta(\hat{\mathbf{r}})\delta(\hat{\mathbf{r}}') \\ &=\; 2\pi \int \widehat{w}(\vartheta) L_\ell(\cos\vartheta) d\cos\vartheta\,, \end{aligned} \tag{5.13}$$

[26] The usual notation for Legendre polynomials is P_ℓ, but we will later need this notation for multipoles of the power spectrum.

[27] In principle we should write here $\widehat{\delta}$ instead of δ, i.e., perturbation with respect to mean density on the sky, not the expectation value, but this looked too messy.

and therefore also

$$\widehat{w}(\vartheta) \;=\; \frac{1}{4\pi}\sum_\ell (2\ell+1)\widehat{C}_\ell L_\ell(\cos\vartheta)\,. \tag{5.14}$$

For $\vartheta = 0$ we get

$$w(0) \equiv \langle\delta^2\rangle = \sum_\ell \frac{2\ell+1}{4\pi} C_\ell \qquad \text{and} \qquad \widehat{w}(0) \equiv \langle\delta^2\rangle_{\rm sky} = \sum_\ell \frac{2\ell+1}{4\pi}\widehat{C}_\ell\,. \tag{5.15}$$

5.2 Legendre polynomials

The Legendre polynomials $L_\ell(x)$ ([8], Chapter 12) form a complete and orthogonal set of functions over the interval $[-1,1]$. $L_\ell(x)$ is a polynomial of order ℓ. The orthogonality relation is

$$\int_{-1}^{1} L_\ell(x) L_{\ell'}(x) dx = \frac{2}{2\ell+1}\delta_{\ell\ell'} \tag{5.16}$$

and the completeness relation[28] is

$$\sum_{\ell=0}^{\infty} \frac{2\ell+1}{2} L_\ell(x) L_\ell(y) \;=\; \delta_D(x-y)\,. \tag{5.17}$$

The orthonormal set is thus

$$\left\{ \sqrt{\frac{2\ell+1}{2}} L_\ell(x) \right\}. \tag{5.18}$$

The Legendre polynomials usually occur as $L_\ell(\cos\vartheta)$, the range $[-1,1]$ mapping into $[0,\pi]$ for ϑ. At the two boundaries they have the values

$$L_\ell(-1) = (-1)^\ell \qquad \text{and} \qquad L_\ell(1) = 1\,. \tag{5.19}$$

Their parity is determined by ℓ:

$$L_\ell(-x) = (-1)^\ell L_\ell(x)\,. \tag{5.20}$$

Legendre polynomials
$L_0(x) = 1$
$L_1(x) = x$
$L_2(x) = \frac{1}{2}(3x^2-1)$
$L_3(x) = \frac{1}{2}(5x^3-3x)$
$L_4(x) = \frac{1}{8}(35x^4-30x^2+3)$

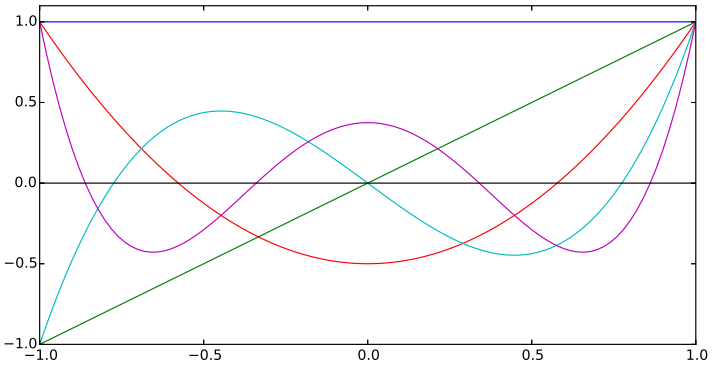


Table 4: Legendre polynomials: L_0 (blue), L_1 (green), L_2 (red), L_3 (cyan), L_4 (magenta).

Associated Legendre functions $P_\ell^m(z) = P_\ell^m(\cos\theta)$
$P_1^1(z) = \sqrt{1-z^2} = \sin\theta$
$P_2^1(z) = 3z\sqrt{1-z^2} = 3\cos\theta\sin\theta$
$P_2^2(z) = 3(1-z^2) = 3\sin^2\theta$

Spherical harmonics
$Y_0^0(\theta,\phi) = \frac{1}{\sqrt{4\pi}}$
$Y_1^1(\theta,\phi) = -\sqrt{\frac{3}{8\pi}}\sin\theta e^{i\phi} = -\sqrt{\frac{3}{4\pi}}\cdot\frac{1}{\sqrt{2}}(x+iy)$
$Y_1^0(\theta,\phi) = \sqrt{\frac{3}{4\pi}}\cos\theta = \sqrt{\frac{3}{4\pi}}\cdot z$
$Y_2^2(\theta,\phi) = \sqrt{\frac{5}{96\pi}}\,3\sin^2\theta e^{i2\phi} = \frac{1}{4}\sqrt{\frac{15}{2\pi}}\cdot(x+iy)^2$
$Y_2^1(\theta,\phi) = -\sqrt{\frac{5}{24\pi}}3\sin\theta\cos\theta e^{i\phi} = -\frac{1}{2}\sqrt{\frac{15}{2\pi}}\cdot z(x+iy)$
$Y_2^0(\theta,\phi) = \sqrt{\frac{5}{4\pi}}\left(\frac{3}{2}\cos^2\theta-\frac{1}{2}\right) = \frac{1}{4}\sqrt{\frac{5}{\pi}}\cdot(3z^2-1)$

Table 5: Spherical harmonics.

5.3 Spherical harmonics

The spherical harmonics ([8], Chapter 12) form a complete and orthonormal set of functions over the sphere. The orthonormality relation is

$$\int d\Omega\, Y^*_{\ell m}(\hat{\mathbf{r}}) Y_{\ell' m'}(\hat{\mathbf{r}}) \;=\; \delta_{\ell\ell'}\delta_{mm'} \tag{5.21}$$

and the completeness relation is

$$\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} Y^*_{\ell m}(\theta,\phi) Y_{\ell m}(\theta',\phi') \;=\; \delta_D(\phi-\phi')\delta_D(\cos\theta-\cos\theta')\,. \tag{5.22}$$

They are elementary functions and have the explicit form[29]

$$Y_{\ell m}(\theta,\phi) \;=\; (-1)^m\sqrt{\frac{2\ell+1}{4\pi}}\sqrt{\frac{(\ell-m)!}{(\ell+m)!}}P_\ell^m(\cos\theta)e^{im\phi}\,. \tag{5.23}$$

where

$$\begin{aligned} P_\ell^m(z) &= (1-z^2)^{m/2}\frac{d^m}{dz^m}L_\ell(z) \qquad (\text{for } m\geq 0) \\ P_\ell^{-m}(z) &= (-1)^m\frac{(\ell-m)!}{(\ell+m)!}P_\ell^m(z) \end{aligned} \tag{5.24}$$

are *associated Legendre functions* so that

$$P_\ell^0(z) = L_\ell(z)\,. \tag{5.25}$$

[28] Again, it is difficult to find this in literature. It is in Wikipedia.

[29] There are different phase conventions for the $Y_{\ell m}$. We follow the *Condon-Shortley phase convention* of [8]. [26] uses the same phase convention for $Y_{\ell m}$, but they put the $(-1)^m$ already in the definition of P_ℓ^m, so it is missing in their version of Eq. (5.23).

Note that for $z = \cos\theta$, $(1-z^2)^{1/2} = \sin\theta$. Thus the θ-dependence is in $P_\ell^m(\cos\theta)$ and the ϕ-dependence is in $e^{im\phi}$. The functions P_ℓ^m are real and

$$Y_{\ell,-m} = (-1)^m Y^*_{\ell m}\,, \tag{5.26}$$

so that

$$Y_{\ell 0}(\theta,\phi) = \sqrt{\frac{2\ell+1}{4\pi}} L_\ell(\cos\theta) \qquad \text{is real.} \tag{5.27}$$

Summing over the m corresponding to the same multipole number ℓ gives the *addition theorem*

$$\sum_m Y^*_{\ell m}(\theta',\phi') Y_{\ell m}(\theta,\phi) = \frac{2\ell+1}{4\pi} L_\ell(\cos\vartheta)\,, \tag{5.28}$$

where ϑ is the angle between $\hat{\mathbf{n}} = (\theta,\phi)$ and $\hat{\mathbf{n}}' = (\theta',\phi')$, i.e., $\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}' = \cos\vartheta$. For $\hat{\mathbf{n}} = \hat{\mathbf{n}}'$ this becomes

$$\sum_m |Y_{\ell m}(\theta,\phi)|^2 = \frac{2\ell+1}{4\pi} \tag{5.29}$$

(since $L_\ell(1) = 1$ always).

We shall also need the expansion of a plane wave in terms of spherical harmonics,

$$e^{i\mathbf{k}\cdot\mathbf{x}} = 4\pi \sum_{\ell m} i^\ell j_\ell(kx) Y_{\ell m}(\hat{\mathbf{x}}) Y^*_{\ell m}(\hat{\mathbf{k}})\,. \tag{5.30}$$

Here $\hat{\mathbf{x}}$ and $\hat{\mathbf{k}}$ are the unit vectors in the directions of $\mathbf{x}$ and $\mathbf{k}$, and j_ℓ is the spherical Bessel function.

5.4 Euler angles

[Note: Euler angles and Wigner functions are needed here only for the proof of Eq. (5.6). If you like, you can skip to Sec. 5.6. Whether they are needed later in the course, I don't yet know. The Wigner functions would appear in the treatment of weak lensing in spherical geometry, but we may not get that deep.]

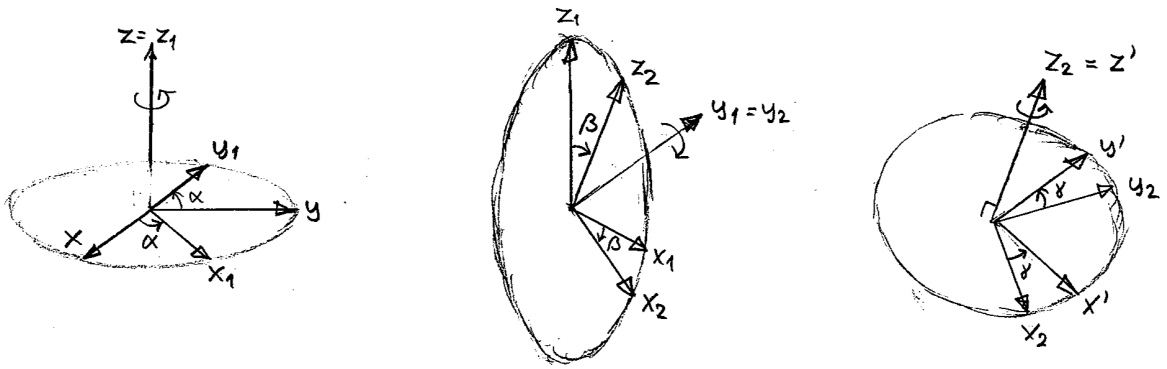
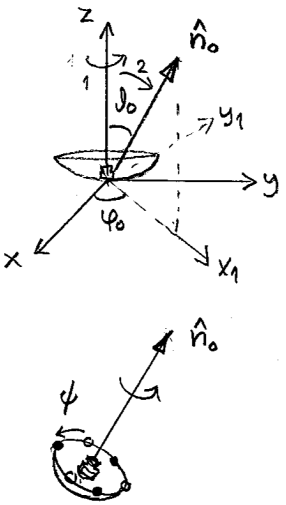


Figure 14: Constructing a rotation of the coordinate system from the Euler angles: 1) Rotate the coordinate system around the original z axis by an angle α ($0 \le \alpha < \pi$). Call the new coordinates x_1, y_1, z_1. 2) Rotate around the new y_1 axis by an angle β ($0 \le \beta \le \pi$). Call the new coordinates x_2, y_2, z_2. 3) Rotate around the z_2 axis by an angle γ ($0 \le \gamma < \pi$). Call the new coordinates x', y', z'.



To specify the orientation of a coordinate system, or in general, the orientation of a rigid object, in 3D, requires three angles: two to specify the direction where the z axis points to, and a third angle to fix the degree of freedom of rotating the x and y axes around it. These are called *Euler angles*, commonly denoted α, β, γ. (Depending on application, they may also be denoted ϕ, θ, ψ/χ.) There are two ways to specify them, i.e., how to construct a rotation from an original coordinate system to the new one in terms of them ([26], Sec. 1.4); both ways lead to the same final orientation. I present just the first one, see Fig. 14. (The side figure relates to where I first needed Euler angles, the orientation of the Planck satellite beams. One can think of rotating a radio telescope, which was initially pointing upward: the angles (β, α) are the spherical coordinates (θ, ϕ) of the $\hat{\mathbf{z}}'$ unit vector (the final pointing direction $\hat{\mathbf{n}}_0$) in the original coordinate system, and γ (ψ) rotates around it.)

5.5 Wigner D-functions

Wigner functions ([26], Chapter 4) are related to rotation of spherical coordinates. For a given sky (realization) its spherical harmonic coefficients $a_{\ell m}$ depend on the chosen spherical coordinate system (e.g., equatorial, ecliptic, galactic). Write

$$\delta(\hat{\mathbf{r}}) \;=\; \sum_{\ell m} a_{\ell m} Y_{\ell m}(\theta,\phi) \;=\; \sum_{\ell m} \tilde{a}_{\ell m} Y_{\ell m}(\tilde{\theta},\tilde{\phi})\,, \tag{5.31}$$

where (θ, ϕ) are the coordinates of $\hat{\mathbf{r}}$ in the first coordinate system and $(\tilde{\theta}, \tilde{\phi})$ in the second coordinate system. The relation between the $a_{\ell m}$ of different coordinate systems is given by

$$a_{\ell m} \;=\; \sum_{m'} D^{\ell}_{mm'}(\alpha,\beta,\gamma)\tilde{a}_{\ell m'} \tag{5.32}$$

where the $D^{\ell}_{mm'}$ are *Wigner functions* and the α, β, γ are the *Euler angles* of the rotation from the first to the second coordinate system. This means that the spherical harmonics are related

$$Y_{\ell m'}(\tilde{\theta},\tilde{\phi}) \;=\; \sum_{m} Y_{\ell m}(\theta,\phi) D^{\ell}_{mm'}(\alpha,\beta,\gamma)\,. \tag{5.33}$$

The Euler angles have the ranges[30]

$$0 \le \alpha < 2\pi\,, \qquad 0 \le \beta \le \pi\,, \qquad 0 \le \gamma < 2\pi\,. \tag{5.34}$$

In calculations one may want to use angles outside these ranges. Any value of α or γ is equivalent to the same value mod 2π, and the Wigner functions are periodic in them (just like sin and cos). Negative values of β cover the same rotations again, since rotation $(\alpha+\pi, \beta, \gamma-\pi)$ is the same rotation as $(\alpha, -\beta, \gamma)$.

We can think of the set of $a_{\ell m}$ for a fixed ℓ as a $2\ell+1$ component vector, and the $(2\ell+1)\times(2\ell+1)$ matrix $D^{\ell}(\alpha,\beta,\gamma)$ as the matrix for rotating these components to a new basis. The set of matrices $D^{\ell}(\alpha,\beta,\gamma)$ (i.e., for all values of the Euler angles) form a representation of the rotation group SO(3). These matrices are unitary

$$D^{\ell}(\alpha,\beta,\gamma)^{\dagger} \;=\; D^{\ell}(\alpha,\beta,\gamma)^{-1}\,, \tag{5.35}$$

[30] We consider only Wigner functions with integer values of ℓ, m, and m'. In quantum mechanics one introduces also Wigner functions with half-integer ℓ, which have periods of 4π instead of 2π. Some of the equations given here do not apply to them as such.

so that the inverse relations to (5.32) and (5.33) are

$$\begin{aligned} \tilde{a}_{\ell m} &= \sum_{m'} a_{\ell m'} D^{\ell}_{m'm}(\alpha,\beta,\gamma)^* \\ Y_{\ell m}(\theta,\phi) &= \sum_{m'} D^{\ell}_{mm'}(\alpha,\beta,\gamma)^* Y_{\ell m'}(\tilde{\theta},\tilde{\phi})\,. \end{aligned} \tag{5.36}$$

The Wigner functions are orthogonal and complete[27] in this 3D space of rotations:

$$\int_0^{2\pi} d\alpha \int_0^{\pi} d\cos\beta \int_0^{2\pi} d\gamma\, D^{\ell}_{mn}(\alpha,\beta,\gamma)^* D^{\ell'}_{m'n'}(\alpha,\beta,\gamma) = \frac{8\pi^2}{2\ell+1}\delta_{\ell\ell'}\delta_{mm'}\delta_{nn'} \tag{5.37}$$

$$\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\sum_{m'=-\ell}^{\ell} \frac{2\ell+1}{8\pi^2} D^{\ell}_{mm'}(\alpha,\beta,\gamma)^* D^{\ell}_{mm'}(\alpha',\beta',\gamma') = \delta_D(\alpha-\alpha')\delta_D(\cos\beta-\cos\beta')\delta_D(\gamma-\gamma')\,.$$

The orthonormal set is thus

$$\left\{ \sqrt{\frac{2\ell+1}{8\pi^2}} D^{\ell}_{mm'}(\alpha,\beta,\gamma) \right\}. \tag{5.38}$$

The Wigner functions are elementary functions and they can be given in terms of the real functions $d^{\ell}_{mn}(\beta)$ as

$$D^{\ell}_{mn}(\alpha,\beta,\gamma) = e^{-im\alpha} d^{\ell}_{mn}(\beta) e^{-in\gamma}\,, \tag{5.39}$$

where [27]

$$d^{\ell}_{mn}(\beta) = \sum_{t=\max(0,m-n)}^{\min(\ell+m,\ell-n)} \frac{(-1)^{\ell}\sqrt{(\ell+m)!(\ell-m)!(\ell+n)!(\ell-n)!}}{(\ell+m-t)!(\ell-n-t)!t!(t+n-m)!} \left(\frac{\cos\beta}{2}\right)^{2\ell+m-n-2t} \left(\frac{\sin\beta}{2}\right)^{2t+n-m}. \tag{5.40}$$

(the summation limits are equivalent to avoiding factorials of negative numbers).

These d-matrices have a number of symmetries that relate different components:

$$\begin{aligned} d^{\ell}_{mm'}(\beta) &= (-1)^{m-m'} d^{\ell}_{-m,-m'}(\beta) = (-1)^{m-m'} d^{\ell}_{m'm}(\beta) = d^{\ell}_{-m',-m}(\beta) \\ d^{\ell}_{mm'}(-\beta) &= d^{\ell}_{m'm}(\beta)\,. \end{aligned} \tag{5.41}$$

The spherical harmonics can be expressed (are special cases of) the Wigner functions:

$$Y_{\ell m}(\theta,\phi) = (-1)^m \sqrt{\frac{2\ell+1}{4\pi}} D^{\ell}_{-m,0}(\phi,\theta,\chi) = \sqrt{\frac{2\ell+1}{4\pi}} D^{\ell}_{m,0}(\phi,\theta,\chi)^* \tag{5.42}$$

(which is independent of the third angle χ as $m'=0$).

Proof of Eq. (5.6): The proof parallels (1.74), which we repeat here:

$$\begin{aligned} \langle \delta^*_{\mathbf{k}} \delta_{\mathbf{k}'} \rangle &= \frac{1}{V^2} \int d^d x e^{i\mathbf{k}\cdot\mathbf{x}} \int d^d x' e^{-i\mathbf{k}'\cdot\mathbf{x}'} \langle \delta(\mathbf{x})\delta(\mathbf{x}') \rangle \\ &= \frac{1}{V^2} \int d^d x e^{i\mathbf{k}\cdot\mathbf{x}} \int d^d r e^{-i\mathbf{k}'\cdot(\mathbf{x}+\mathbf{r})} \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r}) \rangle \\ &= \frac{1}{V^2} \int d^d r e^{-i\mathbf{k}'\cdot\mathbf{r}} \xi(\mathbf{r}) \int d^d x e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} \\ &= \frac{1}{V} \delta_{\mathbf{k}\mathbf{k}'} \int d^d r e^{-i\mathbf{k}\cdot\mathbf{r}} \xi(\mathbf{r}) \equiv \frac{1}{V} \delta_{\mathbf{k}\mathbf{k}'} P(\mathbf{k})\,, \end{aligned} \tag{5.43}$$

but is more complicated, since instead of the simple translation of the coordinate system $\mathbf{x}' = \mathbf{x}+\mathbf{r}$ which replaces the plane wave $e^{-i\mathbf{k}'\cdot\mathbf{x}'}$ by $e^{-i\mathbf{k}'\cdot(\mathbf{x}+\mathbf{r})}$ we now have to introduce a rotation of the coordinate system by an angle ϑ.

So let's start along the same lines:

$$\langle a^*_{\ell m} a_{\ell' m'} \rangle \;\equiv\; \int d\Omega\, Y_{\ell m}(\theta,\phi) \int d\Omega'\, Y^*_{\ell' m'}(\theta',\phi') \langle \delta(\hat{\mathbf{r}}) \delta(\hat{\mathbf{r}}') \rangle \,. \tag{5.44}$$

Now $\langle \delta(\hat{\mathbf{r}}) \delta(\hat{\mathbf{r}}') \rangle = w(\vartheta)$, but to continue, we would need to rewrite $Y^*_{\ell' m'}(\theta',\phi')$ in terms of θ, ϕ, and ϑ. In the preceding, (θ,ϕ) and (θ',ϕ') are the coordinates of two different points in the same coordinate system. Now we introduce another coordinate system $(\tilde\theta,\tilde\phi)$, which is a spherical coordinate system relative to the first point $\hat{\mathbf{r}}$, i.e., $\hat{\mathbf{r}}$ is the "North Pole" of this coordinate system, with $\tilde\theta = 0$, so that now $\tilde\theta' = \vartheta$. The rotation to the new coordinate system is given by Euler angles $\alpha = \phi$, $\beta = \theta$, γ can be anything.

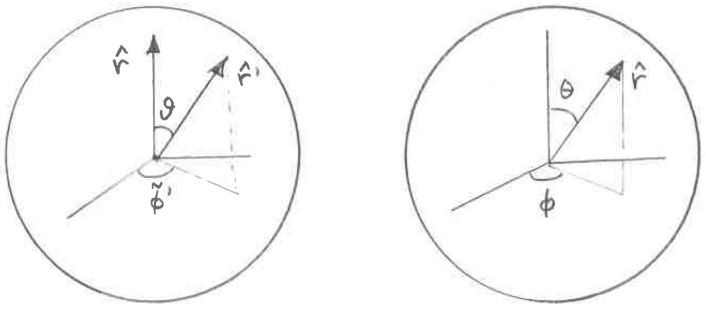


Figure 15: On the left: $\hat{\mathbf{r}}$ and $\hat{\mathbf{r}}'$ in the new coordinate system. On the right: $\hat{\mathbf{r}}$ in the old one.

The new coordinates of $\hat{\mathbf{r}}'$ are $(\vartheta,\tilde\phi')$ and

$$Y^*_{\ell' m'}(\theta',\phi') \;=\; \sum_{m''} D^{\ell'}_{m' m''}(\phi,\theta,\gamma) Y^*_{\ell' m''}(\vartheta,\tilde\phi') \,. \tag{5.45}$$

Now we can continue:

$$\begin{aligned} \langle a^*_{\ell m} a_{\ell' m'} \rangle &= \int d\Omega\, Y_{\ell m}(\theta,\phi) \int d\tilde\Omega' \sum_{m''} D^{\ell'}_{m' m''}(\phi,\theta,\gamma) Y^*_{\ell' m''}(\vartheta,\tilde\phi') w(\vartheta) \qquad (5.46) \\ &= \sqrt{\frac{2\ell+1}{4\pi}} \sum_{m''} \int d\tilde\Omega'\, Y^*_{\ell' m''}(\vartheta,\tilde\phi') w(\vartheta) \int d\Omega\, D^{\ell}_{m0}(\phi,\theta,\chi)^* D^{\ell'}_{m' m''}(\phi,\theta,\gamma) \end{aligned}$$

We then use the fact that $D^\ell_{m0}(\phi,\theta,\chi)$ does not depend on χ to set $\chi = \gamma$, and since γ was allowed to be anything, integrate over it, so that

$$\begin{aligned} \langle a^*_{\ell m} a_{\ell' m'} \rangle &= \sqrt{\frac{2\ell+1}{4\pi}} \sum_{m''} \int d\tilde\Omega'\, Y^*_{\ell' m''}(\vartheta,\tilde\phi') w(\vartheta) \times \frac{1}{2\pi} \int d\gamma \int d\Omega\, D^{\ell}_{m0}(\phi,\theta,\gamma)^* D^{\ell'}_{m' m''}(\phi,\theta,\gamma) \\ &= \sqrt{\frac{2\ell+1}{4\pi}} \sum_{m''} \int d\tilde\Omega'\, Y^*_{\ell' m''}(\vartheta,\tilde\phi') w(\vartheta) \times \frac{4\pi}{2\ell+1} \delta_{\ell\ell'} \delta_{mm'} \delta_{0m''} \\ &= \delta_{\ell\ell'} \delta_{mm'} \sqrt{\frac{4\pi}{2\ell+1}} \int d\tilde\Omega'\, Y^*_{\ell' 0}(\vartheta,\tilde\phi') w(\vartheta) \;=\; \delta_{\ell\ell'} \delta_{mm'} \int d\tilde\Omega'\, L_\ell(\cos\vartheta) w(\vartheta) \\ &= \delta_{\ell\ell'} \delta_{mm'} 2\pi \int d\cos\vartheta\, L_\ell(\cos\vartheta) w(\vartheta) \;=\; \delta_{\ell\ell'} \delta_{mm'} C_\ell \,. \qquad (5.47) \end{aligned}$$

5.6 Relation to the 3D power spectrum

Relate now the angular power spectrum C_ℓ to the 3D power spectrum $\mathcal{P}(k)$. From (4.2) follows

$$\delta(\hat{\mathbf{r}}) \equiv \frac{n(\hat{\mathbf{r}}) - \langle n \rangle}{\langle n \rangle} = \frac{1}{V_S} \int_0^\infty \delta(\mathbf{r}) S(r) r^2 \, dr \,. \tag{5.48}$$

Expand

$$\delta(\mathbf{r}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi \sum_{\mathbf{k}\ell m} \delta_{\mathbf{k}} i^\ell j_\ell(kr) Y_{\ell m}(\hat{\mathbf{r}}) Y^*_{\ell m}(\hat{\mathbf{k}}) \,, \tag{5.49}$$

and insert these into

$$a_{\ell m} = \int d\Omega \, Y^*_{\ell m}(\hat{\mathbf{r}}) \delta(\hat{\mathbf{r}}) \tag{5.50}$$

to obtain (**exercise**)

$$\begin{aligned}
a_{\ell m} &= \sum_{\mathbf{k}} \delta_{\mathbf{k}} \frac{4\pi i^\ell}{V_s} \int_0^\infty dr \, S(r) r^2 j_\ell(kr) Y^*_{\ell m}(\hat{\mathbf{k}}) \,, \\
\langle |a_{\ell m}|^2 \rangle &= \frac{16\pi^2}{V} \sum_{\mathbf{k}} P(k) \left| Y_{\ell m}(\hat{\mathbf{k}}) \right|^2 \left[\frac{1}{V_S} \int_0^\infty dr \, S(r) r^2 j_\ell(kr) \right]^2 , \qquad \text{and} \\
C_\ell &= \frac{1}{2\ell+1} \sum_m \langle |a_{\ell m}|^2 \rangle = \frac{4\pi}{V} \sum_{\mathbf{k}} P(k) \left[\frac{1}{V_S} \int_0^\infty dr \, S(r) r^2 j_\ell(kr) \right]^2 .
\end{aligned} \tag{5.51}$$

We replace the sum $(1/V)\sum_{\mathbf{k}}$ with the integral $1/(2\pi)^3 \int d^3k$ and $k^2/(2\pi^2)P(k)$ with $\mathcal{P}(k)/k$ to get the final exact result[31]

$$\begin{aligned}
C_\ell &= \frac{2}{\pi} \int dk \, k^2 P(k) \left[\frac{1}{V_S} \int_0^\infty dr \, S(r) r^2 j_\ell(kr) \right]^2 \\
&= 4\pi \int \frac{dk}{k} \mathcal{P}(k) \left[\frac{1}{V_S} \int_0^\infty dr \, S(r) r^2 j_\ell(kr) \right]^2 .
\end{aligned} \tag{5.52}$$

[31] MBW[2] (6.177) has an extra factor $1/(2\ell+1)$ here, but this seems an error. Peacock[1] (16.98) agrees with (5.52).

6 Dynamics

6.1 Linear perturbation theory

In Cosmology II we derived the linear perturbation equations for subhorizon scalar perturbations in the flat universe. For pressureless matter they are

$$\begin{aligned} \dot{\delta}_{\mathbf{k}} + \frac{ikv_{\mathbf{k}}}{a} &= 0 && (6.1)\\ \frac{d}{dt}(av_{\mathbf{k}}) + ik\Phi_{\mathbf{k}} &= 0 && (6.2)\\ -k^2\Phi_{\mathbf{k}} &= 4\pi G a^2 \bar{\rho}_m \delta_{\mathbf{k}}\,, && (6.3) \end{aligned}$$

in Fourier space, and

$$\begin{aligned} \dot{\delta} + \frac{1}{a}\nabla\cdot\mathbf{v} &= 0 && (6.4)\\ \frac{\partial}{\partial t}(a\mathbf{v}) &= -\nabla\Phi && (6.5)\\ \nabla^2\Phi &= 4\pi G a^2 \bar{\rho}_m \delta\,, && (6.6) \end{aligned}$$

in coordinate space (we use the normalization $a_0 = 1$ for the scale factor). Here $\delta = \delta_m$ is the matter density perturbation, $\mathbf{v}$ is the matter flow velocity (with respect to the homogeneously expanding background universe), and Φ is the gravitational potential due to the perturbations. Note that since we assumed scalar perturbations,

$$\nabla\times\mathbf{v} = 0 \quad\Rightarrow\quad \mathbf{v}_{\mathbf{k}} || \mathbf{k} \quad\Rightarrow\quad \mathbf{v}_{\mathbf{k}} = v_{\mathbf{k}}\hat{\mathbf{k}} \quad\Rightarrow\quad \mathbf{k}\cdot\mathbf{v}_{\mathbf{k}} = kv_{\mathbf{k}} \quad\Rightarrow\quad (\nabla\cdot\mathbf{v})_{\mathbf{k}} = ikv_{\mathbf{k}}\,. \quad (6.7)$$

As discussed in Cosmology II, vector perturbations decay, so we can ignore them.

In addition to the matter-dominated flat universe (Einstein–de Sitter) and the flat universe with matter and a cosmological constant (ΛCDM), these equations can be applied to the closed and open matter-dominated universe at scales much smaller than the curvature radius, and for other dark energy models than cosmological constant (vacuum energy) in an approximation where the effect of dark energy perturbations are ignored.

Inserting (6.1) and (6.3) into (6.2) we get the pressureless Jeans equation

$$\begin{aligned} \ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} - 4\pi G\bar{\rho}_m\delta_{\mathbf{k}} &= 0\\ \ddot{\delta}(\mathbf{x}) + 2H\dot{\delta}(\mathbf{x}) - 4\pi G\bar{\rho}_m\delta(\mathbf{x}) &= 0\,. && (6.8) \end{aligned}$$

Since the Fourier-space equation has no k-dependence, the equation holds also in coordinate space. The coefficients in these differential equations depend only on time, i.e., on the background solution. The perturbations $\delta(\mathbf{x})$ in different locations evolve independently; differing only due to their different initial conditions δ and $\dot{\delta}$.

The equation has two independent solutions, the growing mode δ_+ and the decaying mode δ_-. The solutions depend on cosmology via the Hubble parameter $H(t)$. For the flat matter-dominated universe they are

$$\delta_+ \propto a\,, \qquad \delta_- \propto a^{-3/2} \qquad (\Omega_m = \Omega = 1)\,. \quad (6.9)$$

We derived them also for ΛCDM in Cosmology II, Sec. 9.3.6:

$$\begin{aligned} \delta_+ &\propto H\int^a \frac{da}{H^3a^3} \propto \sqrt{\Omega_m a^{-3} + \Omega_\Lambda}\int^a \frac{a^{3/2}da}{[1 + (\Omega_\Lambda/\Omega_m)a^3]^{3/2}}\\ \delta_- &\propto H \propto \sqrt{\Omega_m a^{-3} + \Omega_\Lambda} \qquad (\Omega_m + \Omega_\Lambda = \Omega = 1)\,. && (6.10) \end{aligned}$$

(The freedom of the lower limit of the integral for δ_+, i.e., the integration constant, corresponds to the decaying mode.)

The decaying mode dies out in a while, eliminating one of the two degrees of freedom in the initial conditions. We can then ignore the decaying mode and keep just the growing mode. At all locations and for all Fourier components the time evolution is then the same, $\delta(a) \propto D(a)$ where $D(a)$ is the *growth function* (we leave the normalization of D unspecified here)[32]; i.e.,

$$\delta(\mathbf{x}, a) = D(a)\delta(\mathbf{x}, a_{\rm ref}) \qquad \text{and} \qquad \delta_{\mathbf{k}}(a) = D(a)\delta_{\mathbf{k}}(a_{\rm ref})\,, \tag{6.11}$$

where $a_{\rm ref}$ is the scale factor at some chosen reference time, the choice of which fixes the normalization of D so that $D(a_{\rm ref}) = 1$. The growth function D is a growing-mode solution of (6.8). After the decaying mode has died out, *the density perturbation field does not change in any other way except that it is scaled by this time-dependent factor.* The potential Φ is related to δ by (6.3), where $a^2\bar\rho_m \propto a^{-1}$. Therefore the potential has the time evolution

$$\Phi \propto \frac{D(a)}{a} \quad \Rightarrow \quad \Phi(\mathbf{x}, a) = \frac{D(a)}{a} a_{\rm ref}\Phi(\mathbf{x}, a_{\rm ref}) \quad \text{and} \quad \Phi_{\mathbf{k}}(a) = \frac{D(a)}{a} a_{\rm ref}\Phi_k(a_{\rm ref})\,. \tag{6.12}$$

Since $a = 1/(1+z)$,

$$\frac{d\ln(1+z)}{dt} \;=\; -\frac{d\ln a}{dt} \;=\; -H \tag{6.13}$$

and

$$H^{-1}\frac{\dot\delta}{\delta} \;=\; H^{-1}\frac{\dot D}{D} \;=\; H^{-1}\frac{d\ln D}{dt} \;=\; -\frac{d\ln D}{d\ln(1+z)} \;\equiv f\ , \tag{6.14}$$

where we defined the *growth rate* (of the growing mode)

$$f \;\equiv\; -\frac{d\ln D}{d\ln(1+z)} \;=\; \frac{d\ln\delta}{d\ln a}\,. \tag{6.15}$$

In the flat matter-dominated universe $D \propto a$ and $f = 1$. In cosmologies where the only relevant independent background cosmology parameters are H_0 and Ω_m, like the matter-dominated universe, or ΛCDM, $f(z)$ depends only on $\Omega_m(z)$,[33] so it is often denoted $f(\Omega_m)$. We define the *growth index*

$$\gamma \;\equiv\; \frac{d\ln f}{d\ln\Omega_m(z)}\,. \tag{6.16}$$

The growth index turns out to stay almost constant over a wide range of a or z (at least for ΛCDM and the open Friedmann model), so that

$$f(z) \;\approx\; \Omega_m(z)^\gamma\,. \tag{6.17}$$

[32] MBW[2] normalizes it to 1 at an arbitrary initial time t_i, so that $\delta(\mathbf{x}, a) = D(a)\delta_i(\mathbf{x})$ for the growing mode. Then in the matter-dominated Friedmann model $D(a) = a/a_i$.

[33] For the open Friedmann model, Eq. (6.22) shows this. For ΛCDM the math is more difficult because of the integral in (6.10a), so I was not able to write f in such form. My intuition says that f should depend only on the (time-dependent) density parameters. The only other background quantity affecting the growth is the Hubble parameter, which is related to the time scale, but I expect that when the derivative is taken with respect to $\ln a$ the result should become independent of it.

For the ΛCDM model (from Cosmology II, or (6.10))

$$\begin{aligned} D(a) &= \sqrt{\Omega_m a^{-3}+\Omega_\Lambda}\int^a \frac{x^{3/2}dx}{[1+(\Omega_\Lambda/\Omega_m)x^3]^{3/2}} \\ f(a) &= \frac{1}{1+\frac{\Omega_\Lambda}{\Omega_m}a^3}\left[\frac{a}{\left(a^{-3}+\frac{\Omega_\Lambda}{\Omega_m}\right)^{1/2}\int_0^a \frac{x^{3/2}dx}{\left(1+\frac{\Omega_\Lambda}{\Omega_m}x^3\right)^{3/2}}}-\frac{3}{2}\right] \\ \Omega_m(a) &= \frac{1}{1+(\Omega_\Lambda/\Omega_m)a^3} = \frac{(1+z)^3}{(1+z)^3+\Omega_\Lambda/\Omega_m} \end{aligned} \tag{6.18}$$

and $\gamma \approx 0.55$ (see Fig. 16) so that[34]

$$f(z) \approx \Omega_m(z)^{0.55}\,. \tag{6.19}$$

For the ($\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$) ΛCDM model $f(z=0) \approx 0.52$, and is more for higher redshift (when Ω_m was larger).

Example: For the open Friedmann model

$$\Omega_m(a) = \frac{1}{1+x}\,, \qquad \text{where} \quad x \equiv \frac{1-\Omega_m}{\Omega_m}a = \frac{1}{\Omega_m(a)}-1\,, \tag{6.20}$$

and Dodelson[11] (Exercise 7.8) gives the growth function as

$$D(a) = \frac{5\Omega_m}{2(1-\Omega_m)}\left[3\frac{\sqrt{1+x}}{x^{3/2}}\ln\left(\sqrt{1+x}-\sqrt{x}\right)+1+\frac{3}{x}\right]. \tag{6.21}$$

Derivating this, I find

$$f(a) = -\frac{\frac{9}{2x}\left[\frac{1+(2x/3)}{\sqrt{x}\sqrt{1+x}}\ln\left(\sqrt{1+x}-\sqrt{x}\right)+1\right]}{3\frac{\sqrt{1+x}}{x^{3/2}}\ln\left(\sqrt{1+x}-\sqrt{x}\right)+1+\frac{3}{x}}\,. \tag{6.22}$$

This turns out to give $\gamma \approx 0.59$ (in the literature this is given usually as 0.6, but 0.59 seems to be closer, from trying both for Fig. 17). See Fig. 17. From (6.22) it is clear that f depends on $\Omega_m(a)$ only, since x can be written in terms of $\Omega_m(a)$.

In General Relativity, the growth rate of linear perturbations of pressureless matter is determined by the expansion rate. An alternative (to dark energy) explanation to the acceleration of the expansion of the universe is to modify the theory of gravity. Such modified gravity should lead to a different relation between the expansion rate and growth rate. Thus we can differentiate between these two explanations by measuring both the expansion rate and growth rate accurately.

It is one of the goals of the Euclid mission to test gravity by determining the *growth index* γ to an accuracy better than ± 0.02.

In Sec. 7 we shall discuss how the growth rate can be determined from the redshift space distortion (RSD) of the galaxy distribution. For that we need a relation between velocity and density perturbations that we discuss below.

From (6.4) and the definition of the growth rate f (6.14), we have (once the decaying mode has died out) a relation between the density perturbation and the velocity divergence:

$$\delta(\mathbf{x}) = \frac{-\nabla\cdot\mathbf{v}}{aHf}\,. \tag{6.23}$$

[34] In literature, e.g. [33], I have seen this written as $\gamma = 6/11$ (which is ≈ 0.5455). I do not know whether this is supposed to be somehow an exact number.

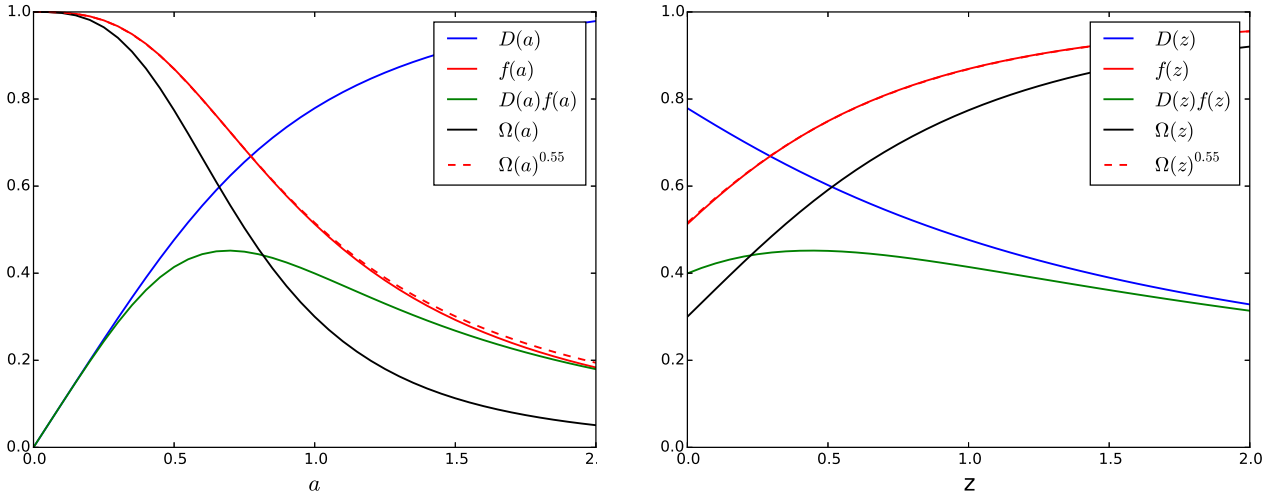


Figure 16: The growth function $D(a)$ and growth rate $f(a)$ for the ΛCDM model (with $\Omega_m = 0.3$).

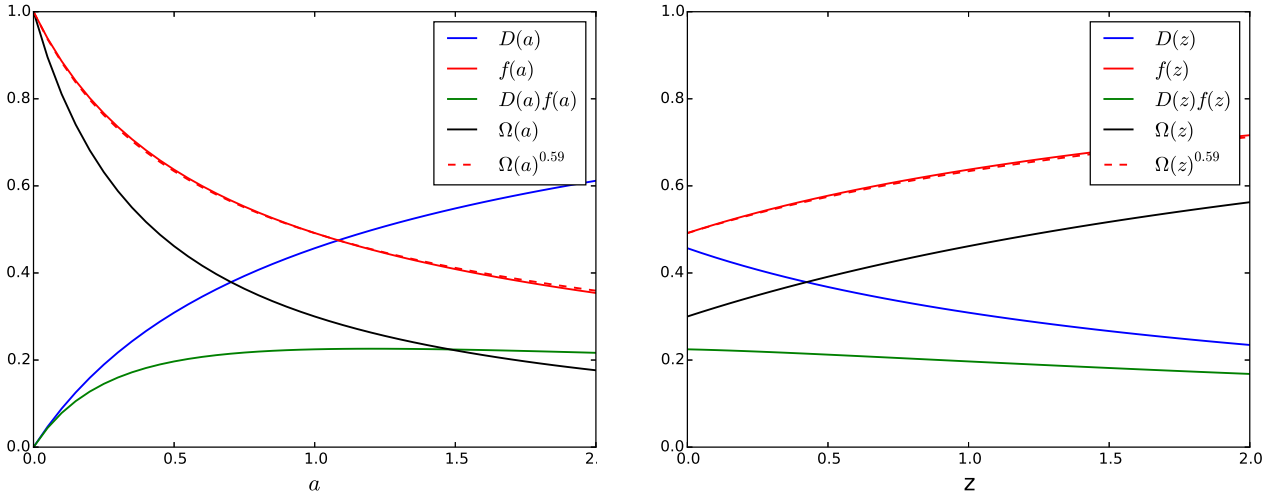


Figure 17: The growth function $D(a)$ and growth rate $f(a)$ for the open Friedmann model (with $\Omega_m = 0.3$).

Likewise, we get a relation between the velocity and the gradient of the gravitational potential (**exercise**)

$$\mathbf{v} \;=\; -\frac{1}{4\pi G\bar{\rho}_m a} H f \nabla\Phi \;=\; -\frac{1}{4\pi G\bar{\rho}_m a}\frac{\dot{D}}{D}\nabla\Phi\,, \tag{6.24}$$

so that at any given time the velocity is proportional to the potential gradient. Both of these properties hold while the perturbations are linear ($\delta \ll 1$).

In the above $\delta = \delta_m$ refers to the matter density field (including dark matter). Galaxy formation is thought to be more efficient in overdense regions, so that the perturbation in galaxy number density ρ_g is enhanced compared to matter density,

$$\delta_g \;=\; b\delta_m\,, \qquad \text{where } b > 1 \tag{6.25}$$

is called *bias*.[35] We treat it as a constant. (This is clearly an approximation, one would expect it to be somewhat redshift dependent and the relation between δ_g and δ_m is not exactly linear.) It is different for different classes of objects: for ordinary galaxies it is close to 1, for large old galaxies it is larger, and for clusters of galaxies even larger. The velocity is caused by the gravity of the total matter perturbation, which affects galaxies likewise. Thus

$$\delta_g \;=\; b\delta_m \;=\; \frac{-b\nabla\cdot\mathbf{v}}{aHf} \;\equiv\; \frac{-\nabla\cdot\mathbf{v}}{aH\beta}\,, \qquad \text{where } \beta \equiv f/b. \tag{6.27}$$

Going back to Fourier space, this reads

$$-ikv_{\mathbf{k}} \;=\; aH\beta\delta_{g\mathbf{k}}\,. \tag{6.28}$$

6.2 Zeldovich approximation

For this section, let us normalize the growth function D so that $D = 1$ at some early matter-dominated time, when $a \ll 1$, but we can already ignore the decaying solution. We refer this time as the "initial time" and denote it with subscript i. Thus

$$\delta(\mathbf{x}, a) \;=\; D(a)\delta_i(\mathbf{x}) \qquad \text{and} \qquad \Phi(\mathbf{x}, a) \;=\; \frac{D(a)}{a} a_i \Phi_i(\mathbf{x}) \tag{6.29}$$

(where the scale factor a is used in place of the time coordinate).

Consider now the position $\mathbf{x}$ of a fluid element as a function of time. Since $\mathbf{x}$ is a comoving coordinate,

$$\dot{\mathbf{x}} \;=\; \frac{\mathbf{v}}{a} \;=\; -\frac{1}{4\pi G\bar{\rho}_m a^2}\frac{\dot{D}}{D}\nabla\Phi \;=\; -\frac{a_i\dot{D}}{4\pi G\bar{\rho}_m a^3}\nabla\Phi_i\,. \tag{6.30}$$

Make the approximation $\nabla\Phi_i(\mathbf{x}) \approx \nabla\Phi_i(\mathbf{x}_i)$, where $\mathbf{x}_i$ is the position of the fluid element at the initial time, and integrate (note that $\bar{\rho}_m a^3$ is constant)

$$\mathbf{x}(t) \;=\; \mathbf{x}_i - \frac{a_i}{4\pi G\bar{\rho}_m a^3}\nabla\Phi_i(\mathbf{x}_i)\int \dot{D}dt\,, \tag{6.31}$$

[35] Maybe there should be a separate subsection on bias at some point. Since we can directly observe the galaxy distribution and the mass distribution is more difficult to infer, we know the amplitude of the galaxy power spectrum and correlation function much more accurately than the corresponding quantities for the mass distribution. Assuming a constant bias, these quantities are related by the factor b^2. Likewise the variances are related by $\sigma_m^2 = b^{-2}\sigma_g^2$. Observationally for optically detected galaxies $\sigma_{g,T}^2(8\,h^{-1}\mathrm{Mpc}) = 1$ (top-hat window function, smoothing scale $8\,h^{-1}$Mpc). Therefore one commonly used measure of the amplitude of the mass distribution inhomogeneity is

$$\sigma_8 \;\equiv\; \sigma_{m,T}(8\,h^{-1}\mathrm{Mpc}) = 1/b_g\,. \tag{6.26}$$

Since we do not know the value of b_g, we do not know the value of σ_8. We can try to infer it from cosmological measurements of quantities that respond to the mass distribution, like weak gravitational lensing or peculiar velocities; and it is predicted by cosmological models when the values of their parameters are fixed.

where we can approximate

$$\int \dot{D} dt = D(a) - D(a_i) \approx D(a) \tag{6.32}$$

at late times, when $D(a) \gg D(a_i)$. Thus we have

$$\mathbf{x}(t) \;\approx\; \mathbf{x}_i - \frac{a_i D(a)}{4\pi G \bar{\rho}_m a^3} \nabla \Phi_i(\mathbf{x}_i)\,, \tag{6.33}$$

the *Zeldovich approximation.*

This result was obtained using first-order perturbation theory, but it turns out that this approach of following the motion of fluid elements does in practice not fail as soon as the normal approach of integrating the density perturbation at a fixed location, when perturbations become large. In the Zeldovich approximation we get the evolution of the density perturbation by following the paths of the neighboring fluid elements as they get closer to each other or move away from each other. You can consider an initial cube of fluid and follow how its corners move to change the volume and shape of the cube.

We get the evolution of the density field from the Jacobian determinant of the transformation $\mathbf{x}_i \to \mathbf{x}(t)$, which gives the change of comoving volume of the fluid element, so that the comoving density changes by its inverse:

$$1 + \delta \;=\; \left| \frac{\partial \mathbf{x}}{\partial \mathbf{x}_i} \right|^{-1} \;=\; \frac{1}{(1 - D\lambda_1)\,(1 - D\lambda_2)\,(1 - D\lambda_3)}\,, \tag{6.34}$$

where

$$\left| \frac{\partial \mathbf{x}}{\partial \mathbf{x}_i} \right| \;=\; (1 - D\lambda_1)\,(1 - D\lambda_2)\,(1 - D\lambda_3)\;, \tag{6.35}$$

where the λ_i are the eigenvalues of

$$\frac{a_i}{4\pi G \bar{\rho}_m a^3} \partial_j \partial_k \Phi_i\,. \tag{6.36}$$

While $\lambda_i D \ll 1$, we have $1 + \delta \approx 1 + D(\lambda_1 + \lambda_2 + \lambda_3)$, so that

$$\delta \approx D(a)(\lambda_1 + \lambda_2 + \lambda_3)\,. \tag{6.37}$$

Rotating the coordinate system so that (6.36) is diagonal, we see that

$$\lambda_1 + \lambda_2 + \lambda_3 \;=\; \frac{a_i}{4\pi G \bar{\rho}_m a^3} \nabla^2 \Phi_i \;=\; \delta_i\,, \tag{6.38}$$

so that (6.37) is consistent with (6.29).

Order the eigenvalues so that $\lambda_1 > \lambda_2 > \lambda_3$. The eigenvalues λ_i may be positive or negative, but for an overdensity $\lambda_i > 0$, their sum is positive, so at least $\lambda_1 > 0$. Following (6.34) as D grows, we find that $\delta \to \infty$ as $D \to 1/\lambda_1$. Clearly the Zeldovich approximation must fail before that. But before it fails, we see that λ_1 corresponds to the direction along which the overdensity compresses the fastest, so that the overdensity flattens towards a planar structure orthogonal to that direction; the resulting flattened overdensity is called a "Zeldovich pancake". We see that overdensities that are not perfectly spherical (the eigenvalues λ_i are not equal) collapse in such a way that the deviation from isotropy grows.

As numerical simulations became larger and better the Zeldovich approximation was relegated to a pedagogical tool. However, recently the Zeldovich approximation has again found use in front-line research as a tool for analyzing observational large-scale structure data, called *reconstruction.* In reconstruction, observed galaxies are moved backwards along their inferred direction of motion to produce a density distribution closer to its primordial state. This helps in locating the baryon acoustic oscillation (BAO) peak (Sec. 10) more accurately.

6.3 Nonlinear growth

When δ grows the evolution becomes nonlinear, requiring a more complicated discussion. One can get further with higher-order perturbation theory, or the Zeldovich approximation; but eventually one has to resort to numerical simulations. We shall not discuss these in this course. The spherically symmetric special case can be done analytically; we offer a simple example below.

The closed Friedmann model. In Cosmology I we derived[36] the expansion law for the closed ($\Omega > 1$) matter-dominated FRW universe. It cannot be given in closed form as $a(t)$, but can be given in terms of an auxiliary variable, the *development angle* ψ, as

$$\begin{aligned} a(\psi) &= a_i \frac{\Omega_i}{2(\Omega_i - 1)}(1 - \cos\psi) = a(\psi)\frac{\Omega(\psi)}{2[\Omega(\psi) - 1]}(1 - \cos\psi) \\ t(\psi) &= H_i^{-1}\frac{\Omega_i}{2(\Omega_i - 1)^{3/2}}(\psi - \sin\psi) = H(\psi)^{-1}\frac{\Omega(\psi)}{2[\Omega(\psi) - 1]^{3/2}}(\psi - \sin\psi)\,, \end{aligned} \tag{6.39}$$

where a_i, Ω_i, and H_i are the scale factor, density parameter, and Hubble parameter at some reference time t_i (usually chosen as the present time t_0, but below we will instead choose t_i to be some early time, when Ω is still very close to 1). In the second forms we took advantage of the fact that we can choose t_i to be any time during the development and replaced it with the "current" time. See Fig. 18 for the shape of $a(t)$. This curve is called a *cycloid*. (It is the path made by a point at the rim of a wheel.) From (6.39) we solve

$$\Omega(\psi) = \frac{2}{1 + \cos\psi}\,. \tag{6.40}$$

Calculating $da/dt = da/d\psi \times d\psi/dt$ we find (**exercise**)

$$H(\psi) = 2H_i\frac{(\Omega_i - 1)^{3/2}}{\Omega_i}\frac{\sin\psi}{(1 - \cos\psi)^2}\,. \tag{6.41}$$

The matter density is given by

$$\rho(\psi) = \rho_i\left(\frac{a_i}{a(\psi)}\right)^3 = 8\rho_i\frac{(\Omega_i - 1)^3}{\Omega_i^3(1 - \cos\psi)^3}\,. \tag{6.42}$$

The scale factor reaches a maximum $a_{\max}$ (and the density a minimum) at $t_{\max}$, when $\psi = \pi$, so that

$$a_{\max} = a_i\frac{\Omega_i}{\Omega_i - 1}\,, \qquad t_{\max} = \frac{\pi}{2}H_i^{-1}\frac{\Omega_i}{(\Omega_i - 1)^{3/2}}\,, \qquad \text{and} \qquad \rho(t_{\max}) = \rho_i\frac{(\Omega_i - 1)^3}{\Omega_i^3}\,. \tag{6.43}$$

At this point $H = 0$ and then the universe begins to shrink. Since

$$\rho_i = \frac{3\Omega_i H_i^2}{8\pi G} \qquad \text{we have} \qquad \rho(t_{\max}) = \frac{3\pi}{32Gt_{\max}^2}\,. \tag{6.44}$$

The universe ends at $t_{\rm end} = 2t_{\max}$, when $\psi = 2\pi$ and $a = 0$ again.

Spherical collapse model. The expansion law (6.39) will hold also for a spherically symmetric overdense region within a flat ($\Omega = 1$) matter-dominated FRW universe. Denote the quantities for this flat background universe by $\bar{a}, \bar{H}, \bar{\rho}$. (Time t is the same for both solutions and $\bar{\Omega} = 1$, so we don't need notations for them.) The background universe has

$$\bar{H}^2 = \frac{8\pi G}{3}\bar{\rho} = \left(\frac{2}{3t}\right)^2 \quad \Rightarrow \quad \bar{\rho} = \frac{1}{6\pi G t^2} \tag{6.45}$$

Thus we see that at $t = t_{\max} \equiv t_{\rm ta}$ (we call it now the turnaround time), the density of the overdense region is

$$\rho(t_{\rm ta}) = \frac{9\pi^2}{16}\bar{\rho}(t_{\rm ta}) \approx 5.5517\bar{\rho}(t_{\rm ta})\,, \tag{6.46}$$

[36] Or should have done so.

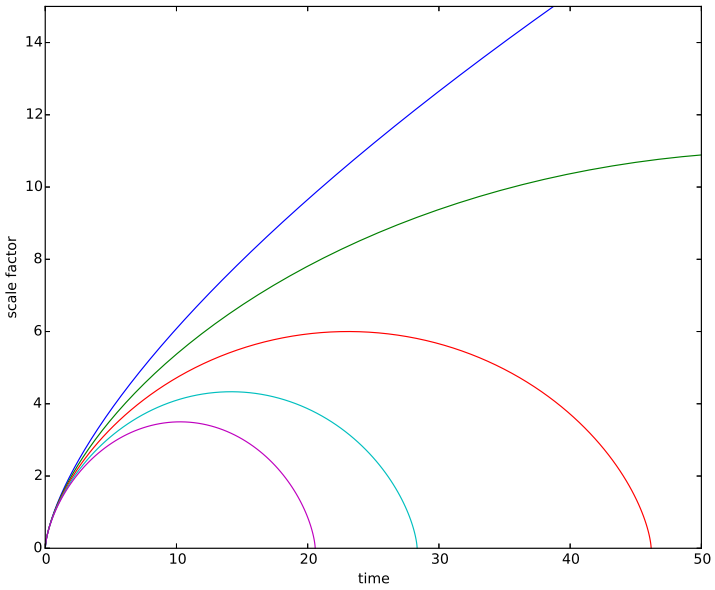


Figure 18: The expansion law for the flat matter-dominated universe (blue, the top curve) and for closed matter-dominated universes with different initial values $\Omega_i > 1$ for the density parameter. Both axes are linear, the units are arbitrary.

i.e., at the turnaround time the density contrast has the value

$$\delta_{\rm ta} = \frac{9\pi^2}{16} - 1 \approx 4.4417\,. \tag{6.47}$$

Until then the overdense region has been expanding, although slower than the surrounding background universe. At turnaround the overdense region begins to collapse.

The preceding applies both for an overdense region with homogeneous density and for one with a spherically symmetric density profile. In the latter case, we have to apply it separately for each spherical shell, and the density ρ refers, not to the density of the shell, but to the mean density within the shell, as it is the total mass within the shell that is responsible for the gravity affecting the expansion or contraction of the shell. To avoid shell crossing the density profile has to decrease outward, so that outer shells do not collapse before inner shells.[37]

In linear perturbation theory, which applies when $\delta \ll 1$, density perturbations in the flat matter-dominated universe grow as

$$\delta^{\rm lin} \propto a \propto t^{2/3}\,. \tag{6.48}$$

When the density contrast δ becomes large it begins to grow faster. Compare now the linear growth law to the above result for δ at turnaround.

The initial density contrast δ_i is given by $\rho_i = (1+\delta_i)\bar\rho_i$. On the other hand

$$\bar H_i^2 = \frac{8\pi G}{3}\bar\rho_i \qquad \text{and} \qquad \Omega_i H_i^2 = \frac{8\pi G}{3}\rho_i \tag{6.49}$$

so that

$$1+\delta_i = \Omega_i \frac{H_i^2}{\bar H_i^2} \qquad \text{or at any time} \qquad 1+\delta = \Omega\frac{H^2}{\bar H^2}\,. \tag{6.50}$$

Thus the density contrast is not simply given by $\Omega - \bar\Omega = \Omega - 1$, since also the Hubble parameters are different for the two solutions. We can sort out the separate contributions from $\Omega_i - 1$ and $(H_i/\bar H_i)^2$ at

[37] We should also include in our model an underdense region around our overdense region so that their combined mean density equals that of the background universe, so as not to affect the evolution of the surroundings.

an early time when $\Omega - 1 \ll 1$ and $\psi \ll 1$, by expanding Ω, H and $\bar{H}$ from (6.40), (6.41) and (6.45&6.39) in terms of ψ (**exercise**) to get

$$\Omega_i \approx 1 + \tfrac{1}{4}\psi_i^2 \qquad \text{and} \qquad \frac{H_i^2}{\bar{H}_i^2} \approx 1 - \tfrac{1}{10}\psi^2 \quad \Rightarrow \quad 1 + \delta_i \approx 1 + \tfrac{3}{20}\psi^2 \quad \Rightarrow \quad \delta_i \approx \tfrac{3}{5}(\Omega_i - 1)\,. \tag{6.51}$$

We can now give the linear prediction for the density contrast at turnaround time:[38]

$$\delta_{\rm ta}^{\rm lin} \;=\; \frac{\bar{a}_{\rm ta}}{\bar{a}_i}\delta_i \;=\; \left(\frac{t_{\rm ta}}{t_i}\right)^{2/3}\delta_i \;\approx \left(\frac{3\pi}{4}\right)^{2/3}\frac{\delta_i}{\Omega_i - 1} \;\approx\; \frac{3}{5}\left(\frac{3\pi}{4}\right)^{2/3} \;\approx\; 1.0624\,, \tag{6.52}$$

where we approximated

$$t_{\rm ta} \;\approx\; \frac{\pi}{2}\bar{H}_i^{-1}\frac{1}{(\Omega_i - 1)^{3/2}} \qquad \text{and} \qquad t_i = \tfrac{2}{3}\bar{H}_1^{-1}\,. \tag{6.53}$$

Thus we conclude that density perturbations begin to collapse when the linear prediction is $\delta \sim 1$, at which time the true density perturbation is already over 4 times stronger.

The collapse is completed at $t_{\rm coll} = 2t_{\rm ta}$ (corresponding to $t_{\rm end}$ of the closed universe model), when the linear prediction gives

$$\delta_{\rm coll}^{\rm lin} \;=\; 2^{2/3}\delta_{\rm ta}^{\rm lin} \;\approx\; 1.6865\,. \tag{6.54}$$

The summary from the above small print is that, for a matter-dominated flat background universe, spherically symmetric overdense regions have collapsed by the time $t_{\rm coll}$ (each overdense region has its own collapse time $t_{\rm coll}$), when the linear perturbation theory prediction would give the magnitude of the density perturbation to be

$$\delta_{\rm coll}^{\rm lin} \;\approx\; 1.6865\,. \tag{6.55}$$

The above special case can be extended to the situation where the background universe is a closed or open Friedmann model (i.e., a matter-dominated FRW universe) [29], and to the ΛCDM model [30]. See also [2], Sec. 5.1, which gives

$$\begin{aligned} \delta_{\rm coll}^{\rm lin} &\approx 1.686\,[\Omega(t_{\rm coll})]^{0.0185} \qquad \text{(Open or closed universe)} \\ \delta_{\rm coll}^{\rm lin} &\approx 1.686\,[\Omega_m(t_{\rm coll})]^{0.0055} \qquad (\Lambda\text{CDM})\,, \end{aligned} \tag{6.56}$$

so that in both cases the dependence on Ω_m is weak, and $\delta_{\rm coll}^{\rm lin} \approx 1.68$ can be used in general.

I suppose these idealized cases would lead to a black hole at the center. In reality overdensities are not exactly spherically symmetric. The deviation from spherical symmetry increases as the collapse progresses. For an ellipsoidal overdensity the flattest direction collapses first leading first to a "Zeldovich pancake". In the situation where the density refers to a number density of galaxies instead of a smooth continuous density, the galaxies will pass the center point at various distances (instead of colliding at the center as in the perfectly spherically symmetric case), after which they will move away from the center and will be decelerated, eventually falling back in and ending up orbiting the center, forming a cluster of galaxies.

6.4 Virialization

In the idealized perfectly spherically symmetric case we ended up with all mass collapsed into a point. In reality we expect to end up at $t_{\rm coll}$ with a system where the masses orbit the center point within half the radius at turnaround (this is called *virialization*). The density is then 8 times larger than the turnaround density, $\frac{9}{2}\pi^2\bar{\rho}(t_{\rm ta}) \approx 44.4\bar{\rho}(t_{\rm ta})$. In the matter-dominated case the universe has meanwhile expanded by a factor $(t_{\rm coll}/t_{\rm ta})^{2/3}$ or the volume by a factor of 4,

[38] Note that Kolb&Turner[28], p. 328, misses the factor 3/5 .

so that the density contrast has grown to $1 + \delta = 18\pi^2 \approx 178$. Quoting from Lyth&Liddle [31] Sec. 9.4: "This number is well verified by numerical simulations as dividing the virialized regions from those where matter is still infalling, but needs some adjustment in the presence of cosmological constant."

Often the rounder number 200 is used; defining the boundary of a halo to be at the radius where the mean density inside the radius is 200 times the mean density of the universe at the time the halo formed. *Halo* means a virialized overdensity of cold dark matter, which may contain one or more galaxies. The common notations r_{200} and M_{200} refer to the radius and mass of a halo defined this way. According to the *virial theorem*, the kinetic energy of the system is minus half the gravitational potential energy. This way one can estimate the typical velocities of the CDM particles from the halo mass. This assumes that the particles have not lost energy during virialization, which should hold for CDM. Baryonic matter will radiate energy as a result of particle interactions, which causes the baryonic component to fall deeper into the system forming the smaller luminous galaxy (or galaxies) inside the halo. At least in the inner parts of galaxies the density of baryonic matter is expected to exceed that of dark matter.

After virialization the density contrast will keep growing since, while the virialized system stays at constant density, the background universe keeps expanding. For example, an overdensity that turned around at $z \approx 3$ or 7, would have virialized at $z \approx 1$ or 3, and by now the matter density contrast would have grown by another factor of $\approx 2^3 = 8$ or $4^3 = 64$. The halo has a density profile, so that the density is higher deeper in.

Let us compare these densities to what we know about the density in the solar neighborhood of our galaxy. Wikipedia (List of nearest stars and brown dwarfs 13.1.2023) lists 110 stars within 20 light years ($r = 6.13$ pc) from the Solar System, with a total mass of $45.96 M_\odot$. In addition the list contains 20 brown dwarfs and two "sub-brown dwarfs or rogue planets", with a total mass of less than $1 M_\odot$. The volume within 20 light years is $V = 965.8\,\mathrm{pc}^3$. For stars, I suppose we can assume the list is complete, giving the local density of baryonic matter in stars

$$\rho_{\mathrm{stars}} = 0.048 M_\odot/\mathrm{pc}^3\,. \tag{6.57}$$

Peacock [1] writes in Sec. 12.3 about our neighborhood: "Including contributions from gas and dust clouds yields a total density in *luminous material* of $\rho_{\mathrm{lum}} = 0.1 M_\odot/\mathrm{pc}^3$, see e.g. Freedman (1987)" and that dynamical determinations of the density "tend to give more mass that is seen directly, but the discrepancy is a factor two at most, and may not exist", which would be consistent with roughly an equal amount of cold dark matter, or less of it. (I would like to find a more recent estimate.) Assuming $h = 0.7$ the present critical density is $\rho_{\mathrm{cr0}} = 1.36 \times 10^{-7} M_\odot/\mathrm{pc}^3$, and further assuming $\Omega_m = 0.3$ this gives $\bar\rho_m = 4.08 \times 10^{-8} M_\odot/\mathrm{pc}^3$, so that a local density of 0.1–$0.2 M_\odot/\mathrm{pc}^3$ would correspond to a density contrast of 2–5×10^6.

7 Redshift space

In galaxy surveys the information about the distance to the galaxy is usually its redshift z, which we use as proxy for its distance r. The distance-redshift relation depends on cosmology; in ΛCDM it is

$$r(z) \;=\; H_0^{-1} \int_0^z \frac{dz'}{\sqrt{\Omega_m(1+z')^3 + (1-\Omega_m)}} \,. \tag{7.1}$$

(where r is the comoving distance). For simplicity, we assume here the linear Hubble relation

$$r \;=\; H_0^{-1} z \qquad \text{or} \qquad z \;=\; H_0 r \,, \tag{7.2}$$

which is an approximation valid for $z \ll 1$. This ignores the peculiar velocities of galaxies. For an individual galaxy,

$$z \;=\; H_0 r + v_r \,, \qquad \text{where } v_r \equiv \hat{\mathbf{r}} \cdot \mathbf{v} \tag{7.3}$$

is the radial component of the peculiar velocity $\mathbf{v}$.

(The following formalism and results can be applied to higher redshifts also, as long as they are applied to a redshift range $\Delta z \ll 1$ (a galaxy survey is divided into redshift bins $[z, z+\Delta z]$), so that the Hubble parameter can be approximated by a constant for that range, and is used in place of H_0.)

We follow [2], Sec. 6.3, and give distances in this section in velocity units, i.e., we set $H_0 = 1$. We also use $c = 1$, so velocity, and now also distance, is dimensionless. Thus

$$1 \;=\; H_0^{-1} \;=\; 2997.92458\, h^{-1}\text{Mpc} \;=\; c \;=\; 299792.458\,\text{km/s} \,. \tag{7.4}$$

(We may call this system of units *Hubble units*.) The redshift and distance of a galaxy are thus related

$$z \;=\; r + v_r \,. \tag{7.5}$$

We define a redshift vector $\mathbf{z}$ for each galaxy,

$$\mathbf{z} \;\equiv\; z\hat{\mathbf{r}} \;=\; \mathbf{r} + v_r\hat{\mathbf{r}} \;=\; \left(1 + \frac{v_r}{r}\right)\mathbf{r} \,. \tag{7.6}$$

It gives the position of the galaxy in *redshift space*.

Because of the contribution $v_r\hat{\mathbf{r}}$ the distribution of galaxies looks different in redshift space. For example, consider a shell of galaxies some distance away from us in the spherical collapse model: In real space they form a sphere. The more distant part of the sphere has a peculiar velocity towards us, the near part away from us. Thus the shell appears flattened in redshift space. At turnaround time, the v_r part exactly cancels the difference in r, so in redshift space the shell is flattened to a flat disk transverse to the line of sight. After turnaround the part that is further away in real space is closer in redshift space and wise versa; the flat disk swells first into a flattened sphere, but as v grows we get eventually a redshift-space structure elongated along the line of sight. A cluster of galaxies also appears similar in redshift space, although the relation between their locations in real and redshift space is more complicated and mixed; galaxies near the center can be moving in our out. Their velocities with respect to the center of the cluster are larger than the radius of the cluster (remember, we are comparing dimensionless quantities here); thus the cluster appears as an elongated structure pointing towards the observer, "a finger of God".

This *redshift space distortion* (RSD) may at first appear as just a nuisance, preventing us from measuring accurately the true galaxy number density field. However, since we have a theoretical understanding how RSD comes about, we can use it as a cosmological tool. While it has erased some information about the galaxy distribution, it has replaced it with another kind of information. RSD tells us about the rate of growth of structure and is thus a test of General Relativity or modified gravity.

7.1 Redshift space as a distortion of real space

In the following we shall only consider objects which are not part of a larger collapsed system (like individual galaxies in clusters), and structures that are large enough so that the overdensities have not yet reached turnaround, so that we avoid the inversion of distance order in redshift space. Thus the peculiar velocities are usually small, of the order 10^{-3} or less. They are caused by the gravity of the density perturbation field (Sec. 6) and form a continuous velocity field $\mathbf{v}(\mathbf{r})$. From (7.5),

$$\frac{dz}{dr} \;=\; 1 + \frac{\partial v_r}{\partial r}\,. \tag{7.7}$$

The redshift space $\{\mathbf{z}\}$ is a distorted version of the real space $\{\mathbf{r}\}$. We denote the number density and its relative perturbation in redshift space as $\rho^{(z)}$ and $\delta^{(z)}$. The mapping $\mathbf{r} \to \mathbf{z}(\mathbf{r})$ maps the volume elements $d^3r \to d^3z$. We require the mapping to be one-to-one; the condition for this is that dz/dr is positive everywhere (i.e., no inversion). These volume elements contain the same galaxies,

$$\rho^{(z)}(\mathbf{z})d^3z \;=\; \rho(\mathbf{r})d^3r \quad\Rightarrow\quad [1+\delta^{(z)}(\mathbf{z})]d^3z \;=\; [1+\delta(\mathbf{r})]d^3r \tag{7.8}$$

(the mean density is the same in both spaces – in our model where $z = r$ holds for the background – since the effect of peculiar velocities on the volume cancel for sufficiently large volumes). Since $d^3r = r^2 dr d\Omega$ and $d^3z = z^2 dz d\Omega$,

$$\frac{d^3r}{d^3z} \;=\; \left(\frac{r}{z}\right)^2 \frac{dr}{dz} \;=\; \frac{1}{(1+v_r/r)^2}\left(1+\frac{\partial v_r}{\partial r}\right)^{-1} \tag{7.9}$$

so that

$$1+\delta^{(z)}(\mathbf{z}) \;=\; \frac{r^2}{(r+v_r)^2}\left(1+\frac{\partial v_r}{\partial r}\right)^{-1} [1+\delta(\mathbf{r})]\,. \tag{7.10}$$

7.2 Linear perturbations and the power spectrum

We make now the linear perturbation theory approximation, i.e., that the perturbations and their gradients are small,

$$|\delta| \ll 1\,, \quad |\mathbf{v}| \ll 1\,, \quad |\nabla\delta| \ll 1\,, \quad |\nabla\mathbf{v}| \ll 1\,, \tag{7.11}$$

so that products of these quantities are second-order small and we can drop them. We also make a separate approximation, assuming that $|v_r| \ll r$, i.e., that we consider sufficiently large distances that the peculiar velocities are small compared to the Hubble recession velocity. We can then approximate (7.10) by

$$\delta^{(z)}(\mathbf{z}) \;\approx\; \delta(\mathbf{r}) - 2\frac{v_r}{r} - \frac{\partial v_r}{\partial r}\,. \tag{7.12}$$

Now also the difference between z and r is small, so that the difference between $\delta^{(z)}(\mathbf{z})$ and $\delta^{(z)}(\mathbf{r})$ is second-order small, and we can ignore it. Thus

$$\delta^{(z)}(\mathbf{r}) \;\approx\; \delta^{(z)}(\mathbf{z})\,. \tag{7.13}$$

We then make the further assumption that

$$2\left|\frac{v_r}{r}\right| \;\ll\; \left|\frac{\partial v_r}{\partial r}\right|\,, \tag{7.14}$$

so that we can drop the $2v_r/r$. This says that the structures (e.g., their characteristic wavelength) that we are looking at are smaller than their distance from us. We have now

$$\delta^{(z)}(\mathbf{r}) \;\approx\; \delta(\mathbf{r}) - \frac{\partial v_r}{\partial r}\,. \tag{7.15}$$

Next we want to go to Fourier space so that

$$\nabla \;\to\; i\mathbf{k}\,. \tag{7.16}$$

How to deal with the $\partial v_r/\partial r$? Here

$$\frac{\partial}{\partial r} = \hat{\mathbf{r}}\cdot\nabla \qquad \text{and} \qquad v_r = \hat{\mathbf{r}}\cdot\mathbf{v}\,. \tag{7.17}$$

The direction vector $\hat{\mathbf{r}}$ appears problematic because replacing ∇ with $i\mathbf{k}$ leaves this coordinate-space quantity so that we are not properly in Fourier space. The way out is the *plane-parallel approximation*,[39] where we assume that we are studying structures that are small compared to their distance from us, so that we can make the Fourier expansion in a reference box V that is small and distant so that it appears in a small angle from us, i.e., the direction vectors $\hat{\mathbf{r}}$ to it are almost parallel. Thus we approximate

$$\hat{\mathbf{r}} \;\approx\; const \qquad \text{for the purpose of taking } \hat{\mathbf{r}}\cdot\mathbf{k}\,. \tag{7.18}$$

We have thus "externalized" the line-of-sight direction $\hat{\mathbf{r}}$ so that it is no longer a coordinate-space variable. We assume scalar perturbations so that

$$\mathbf{v}_\mathbf{k} \;=\; v_\mathbf{k}\hat{\mathbf{k}}\,. \tag{7.19}$$

Now

$$\frac{\partial v_r}{\partial r} \;\to\; i(\hat{\mathbf{r}}\cdot\mathbf{k})(\hat{\mathbf{r}}\cdot\hat{\mathbf{k}})v_\mathbf{k} \tag{7.20}$$

and (7.15) becomes

$$\delta^{(z)}_\mathbf{k} \;\approx\; \delta_\mathbf{k} - i(\hat{\mathbf{r}}\cdot\hat{\mathbf{k}})^2 k v_\mathbf{k}\,. \tag{7.21}$$

The quantity $\hat{\mathbf{r}}\cdot\hat{\mathbf{k}} = \cos\vartheta_\mathbf{k}$ is the cosine of the angle ϑ the wave vector $\mathbf{k}$ makes with the direction from us to the surveyed region. MBW [2] denotes it by $\mu_\mathbf{k}$ and Peacock [1] just by μ. From here on I denote it by $\cos\vartheta_\mathbf{k}$. The point in the plane-parallel approximation is that we consider it as depending just on the direction of $\mathbf{k}$.

Now we assume that the decaying mode has died out so we can use the result (6.28) for the linear growing mode,

$$-ikv_\mathbf{k} \;=\; aH\beta\delta_\mathbf{k} \;\approx\; \beta\delta_\mathbf{k}\,, \tag{7.22}$$

where we approximated $aH \approx a_0H_0 = H_0 = 1$, which is in line with our initial $z \ll 1$ approximation of using a linear Hubble relation. Thus (7.21) becomes

$$\delta^{(z)}_\mathbf{k} \;\approx\; [1+\beta(\cos\vartheta_\mathbf{k})^2]\delta_\mathbf{k}\,. \tag{7.23}$$

Here $\beta = f/b$, where f is the growth rate, which we want to determine to constrain cosmology and theory of gravity; and b is the galaxy bias factor.

We define now the power spectrum $P^{(z)}(\mathbf{k})$ in redshift space and find that, although the power spectrum in real space is isotropic,

$$P^{(z)}(\mathbf{k}) \;\equiv\; V\langle|\delta^{(z)}_\mathbf{k}|^2\rangle \;=\; (1+2\beta\cos^2\vartheta_\mathbf{k} + \beta^2\cos^4\vartheta_\mathbf{k})P(k) \tag{7.24}$$

[39] I suppose the name comes from approximating the transverse directions as flat planes and the radial directions as parallel.

is anisotropic, i.e., depends on the direction of $\mathbf{k}$ via $\vartheta_{\mathbf{k}}$.

We can expand the direction dependence of $P^{(z)}(\mathbf{k}) = P^{(z)}(k, \cos\vartheta_{\mathbf{k}})$ in Legendre polynomials L_ℓ,

$$P^{(z)}(k, \cos\vartheta_{\mathbf{k}}) \;=\; \sum_\ell P_\ell^{(z)}(k) L_\ell(\cos\vartheta_{\mathbf{k}})\,, \tag{7.25}$$

where the

$$P_\ell^{(z)}(k) \;\equiv\; \frac{2\ell+1}{2}\int_{-1}^{1} d\cos\vartheta_{\mathbf{k}} L_\ell(\cos\vartheta_{\mathbf{k}}) P^{(z)}(k, \cos\vartheta_{\mathbf{k}}) \tag{7.26}$$

are the *multipoles* of $P^{(z)}(\mathbf{k})$. We find (**exercise**) that only the monopole, quadrupole, and hexadecapole[40] are nonzero:

$$\begin{aligned} P_0^{(z)}(k) &= (1 + \tfrac{2}{3}\beta + \tfrac{1}{5}\beta^2)P(k) \\ P_2^{(z)}(k) &= (\tfrac{4}{3}\beta + \tfrac{4}{7}\beta^2)P(k) \\ P_4^{(z)}(k) &= \tfrac{8}{35}\beta^2 P(k)\,. \end{aligned} \tag{7.27}$$

The monopole is just $P^{(z)}(\mathbf{k})$ averaged over the directions of $\mathbf{k}$,

$$P_0^{(z)}(k) = \frac{1}{4\pi}\int d\Omega_{\mathbf{k}} P^{(z)}(\mathbf{k})\,. \tag{7.28}$$

We see that redshift-space effects makes the structure look stronger, $P_0^{(z)}(k) > P(k)$. This is because infall towards high density regions make them appear flattened, i.e., smaller in redshift space. Flattening along the line of sight corresponds to a positive quadrupole, i.e., the structure appears stronger along the line of sight.

From a galaxy survey we can determine $P_0^{(z)}(k)$, $P_2^{(z)}(k)$, and $P_4^{(z)}(k)$, and then solve from them, using (7.27), the real-space $P(k)$ and $\beta = f/b$.

The plane-parallel approximation is similar in spirit to the flat-sky approximation of Sec. 4.2. To go beyond it, one would have to take into account the spherical geometry and use spherical harmonics and spherical Bessel functions.

7.3 Correlation function

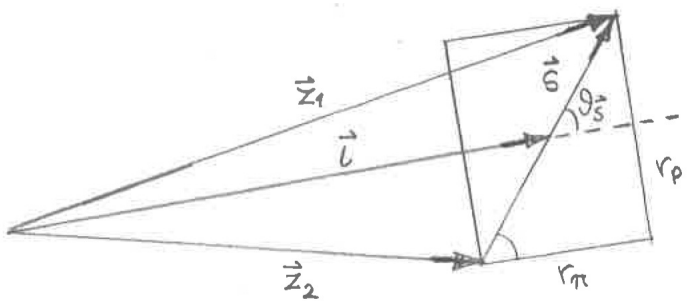


Figure 19: Parallel and perpendicular components in redshift space (here $s_\parallel, s_\perp$ are called r_π, r_p).

Because of the distortion, the redshift-space correlation function

$$\xi^{(z)}(\mathbf{z}_1, \mathbf{z}_2) \;\equiv\; \langle \delta^{(z)}(\mathbf{z}_1)\delta^{(z)}(\mathbf{z}_2)\rangle \tag{7.29}$$

[40] The naming scheme for multipoles goes according to 2^ℓ; they are called "2^ℓ-poles" (mono $= 2^0 = 1$, quadru $= 2^2 = 4$, hexadeca $= 2^4 = 16$).

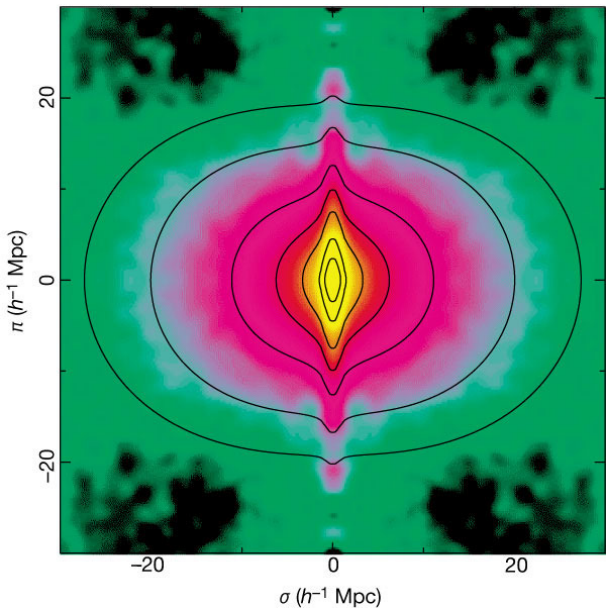


Figure 20: Measured redshift-space correlation function $\xi^{(z)}(s_\parallel, s_\perp)$ from the 2dFGRS survey (here $s_\parallel, s_\perp$ are called π, σ). Note that $s_\perp$ is positive by definition, since it's the radial coordinate of a cylindrical coordinate system. Likewise $s_\parallel$ is positive, since the separation $\mathbf{s}$ of a pair is only defined up to a sign. Here the plot is just reflected with respect to the middle vertical and horizontal to produce a more pleasing image and a 3D interpretation. One can imagine a rotation around this vertical to represent the angular coordinate of which $\xi^{(z)}$ is assumed to be independent. The resulting 3D image gives the expectation value of the overdensity around a galaxy in redshift space. The observer is looking at this from the direction of the bottom of the page. From Peacock et al.[32]

is not isotropic. We define

$$\begin{aligned}
\mathbf{l} &\equiv \tfrac{1}{2}(\mathbf{z}_1 + \mathbf{z}_2) && \text{line-of-sight vector} \\
\mathbf{s} &\equiv \mathbf{z}_1 - \mathbf{z}_2 && \text{separation vector in redshift space} \\
s_\parallel &\equiv \hat{\mathbf{l}} \cdot \mathbf{s} && \text{parallel component of separation} \\
s_\perp &\equiv \sqrt{s^2 - s_\parallel^2} && \text{perpendicular component of separation} \\
\mu &\equiv \frac{s_\parallel}{s} = \cos\vartheta_{\mathbf{s}} && \text{direction cosine}
\end{aligned} \tag{7.30}$$

($s_\parallel, s_\perp$ may also be called π, σ or r_π, r_p). In the plane-parallel approximation ($s_\parallel, s_\perp \ll l$) $\xi^{(z)}$ depends only on $s_\parallel$ and $s_\perp$, or equivalently on s and μ:

$$\xi^{(z)}(\mathbf{z}_1, \mathbf{z}_2) \approx \xi^{(z)}(\mathbf{s}) = \xi^{(z)}(s_\parallel, s_\perp) = \xi^{(z)}(s, \mu)\,. \tag{7.31}$$

(The point here is that we approximate the redshift distortion happening along the $\mathbf{l}$, i.e., $s_\parallel$, direction instead of the $\mathbf{z}_1$ and $\mathbf{z}_2$ directions; so that the distortion is isolated into $s_\parallel$.)

7.3.1 Linear growing mode

Linear perturbation theory can be applied to ξ when correlations are relatively weak, i.e., for large s (but for the plane-parallel approximation to be valid, it still has to be small compared to the depth of the survey).

For the linear growing mode we can do the expansion of $\xi^{(z)}(s, \mu)$ in Legendre polynomials analogous to (7.25):

$$\xi^{(z)}(s, \mu) = \sum_\ell \xi_\ell^{(z)}(s) L_\ell(\mu)\,. \tag{7.32}$$

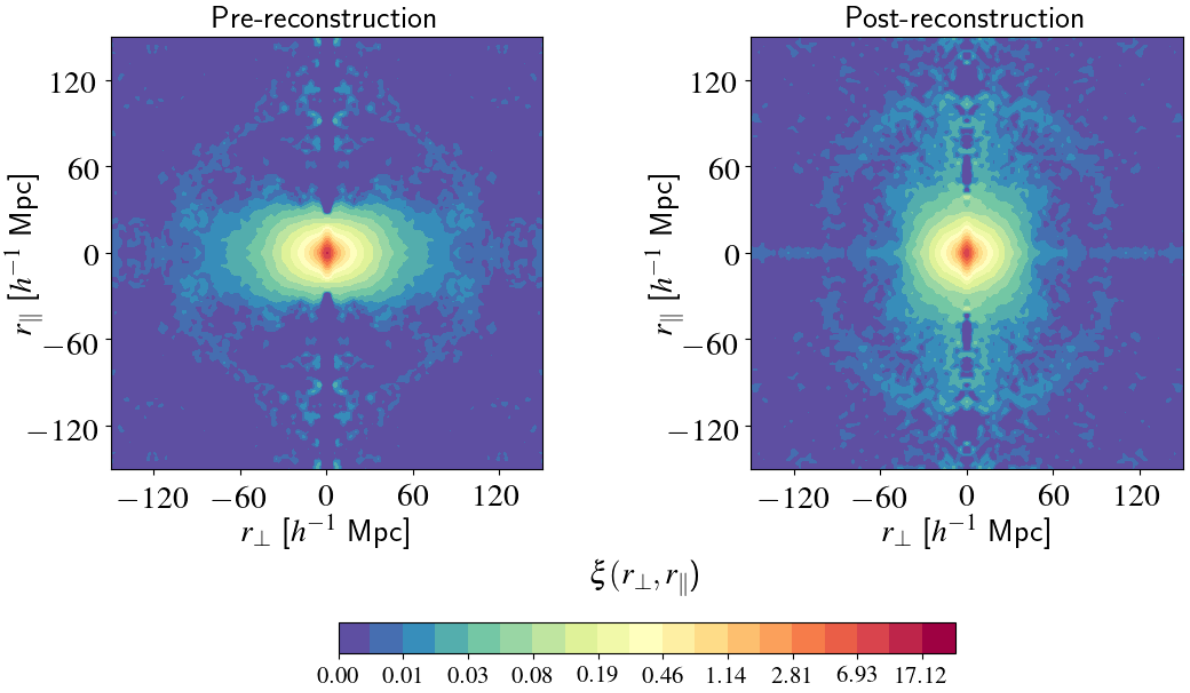


Figure 21: Left panel: Same as Fig. 20, put from the newer and larger eBOSS survey. Based on 377 458 luminous red galaxies (LRG) with redshifts $0.6 < z < 1$ from BOSS and eBOSS. Right-panel: after an attempt to correct the galaxy positions for the redshift-space distortion. Note the ring at $r \approx 100\,h^{-1}$Mpc: this is the BAO peak discussed in Sec. 10 and visible also in Fig. 2. From Bautista et al. [33].

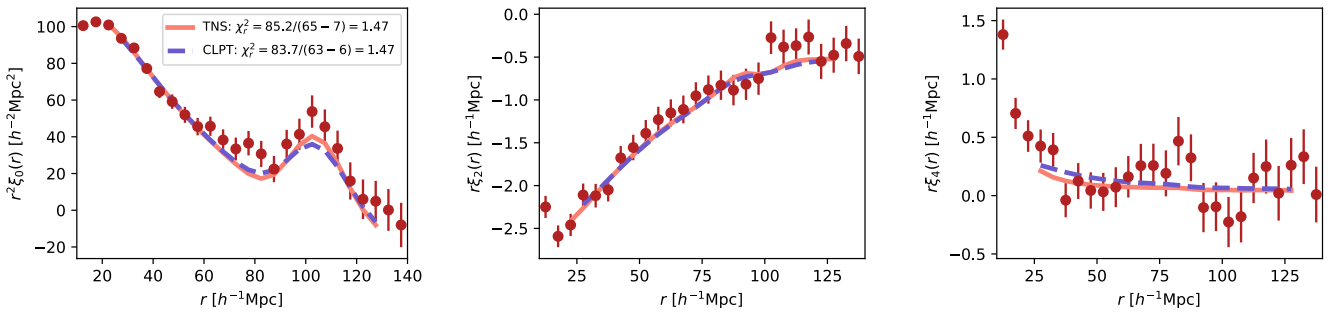


Figure 22: The multipoles $\xi_\ell^{(z)}(s)$ from the eBOSS survey, i.e., of the correlation function plotted in the left panel of Fig. 21. Note that the plotted quantities are $s^2\xi_0^{(z)}(s)$ (left), $s\xi_2^{(z)}(s)$ (middle), and $s\xi_4^{(z)}(s)$ (right). The two solid lines in each plot are fits to the data of two theoretical models, which include also non-linear effects. From Bautista et al. [33].

This is not as easy as for the power spectrum since the relation between $\delta^{(z)}$ and δ, which was simple in Fourier space, Eq. (7.23),

$$\delta_{\mathbf{k}}^{(z)} \;\approx\; [1+\beta(\cos\vartheta_{\mathbf{k}})^2]\delta_{\mathbf{k}} \;=\; \left[1+\beta\left(\frac{k_z}{k}\right)^2\right]\delta_{\mathbf{k}}\,, \tag{7.33}$$

becomes

$$\delta^{(z)}(\mathbf{z}) \;\approx\; \left[1+\beta\left(\frac{\partial}{\partial z}\right)^2\left(\nabla^2\right)^{-1}\right]\delta(\mathbf{x}) \tag{7.34}$$

in coordinate space, so that the relation

$$P^{(z)}(\mathbf{k}) \;=\; \left[1+\beta\left(\frac{k_z}{k}\right)^2\right]^2 P(k) \tag{7.35}$$

in Fourier space becomes

$$\begin{aligned}\xi^{(z)}(\mathbf{s}) \;&\approx\; \left[1+\beta\left(\frac{\partial}{\partial z}\right)^2\left(\nabla^2\right)^{-1}\right]^2\xi(s)\\ &=\; \left[1+2\beta\left(\frac{\partial}{\partial z}\right)^2\left(\nabla^2\right)^{-1}+\beta^2\left(\frac{\partial}{\partial z}\right)^4\left(\nabla^2\right)^{-2}\right]\xi(s)\\ &=\; \xi(s)+2\beta\left(\frac{\partial}{\partial z}\right)^2 F(\mathbf{s})+\beta^2\left(\frac{\partial}{\partial z}\right)^4 G(\mathbf{s})\end{aligned} \tag{7.36}$$

in coordinate space. Here where $F(\mathbf{s})$ and $G(\mathbf{s})$ are the solutions of the Poisson equations

$$\nabla^2 F=\xi \qquad \text{and} \qquad \nabla^2 G=F\,. \tag{7.37}$$

The solution of the Poisson equation is given by[41]

$$F(\mathbf{s}) \;=\; \frac{-1}{4\pi}\int d^3r\,\frac{\xi(r)}{|\mathbf{s}-\mathbf{r}|}\,, \tag{7.41}$$

where we can expand ([8], Eq. 12.4)

$$\frac{1}{|\mathbf{s}-\mathbf{r}|} \;=\; \sum_{\ell=0}^{\infty}\frac{r_<^\ell}{r_>^{\ell+1}}L_\ell(\cos\gamma)\,, \tag{7.42}$$

[41] This is common stuff in electrodynamics. It's been a long time since I studied electrodynamics, but let's consider Newtonian gravity, with which I am more familiar and which is closer to the topic of cosmology. The gravitational potential Φ is given by the Poisson equation

$$\nabla^2\Phi=4\pi G\rho \quad\Rightarrow\quad \Phi(\mathbf{s})=-G\int d^3r\,\frac{\rho(\mathbf{r})}{|\mathbf{s}-\mathbf{r}|}\,, \tag{7.38}$$

where $\rho(\mathbf{r})$ is the mass density. The result (7.43) below says that for an isotropic (spherically symmetric) mass distribution

$$\Phi(s) \;=\; -\frac{G}{s}\int_0^s 4\pi r^2\rho(r)dr - G\int_s^\infty 4\pi r\rho(r)dr \;=\; -\frac{GM(s)}{s}-G\int_s^\infty 4\pi r\rho(r)dr\,, \tag{7.39}$$

where $M(s)$ is the mass inside radius s. The gravitational field is $-\nabla\Phi(s)=-\Phi'(s)\hat{\mathbf{s}}$, so we get for it

$$-|\nabla\Phi(s)| \;=\; -\Phi'(s) \;=\; -\frac{GM(s)}{s^2}+\frac{G}{s}M'(s)-4\pi Gs\rho(s) \;=\; -\frac{GM(s)}{s^2}\,, \tag{7.40}$$

the familiar distance-squared law with the familiar property of gravity that the gravitational field of a spherically symmetric mass distribution depends just on the mass inside, $M(s)$. The second integral from the outside mass distribution for the potential is needed so that the contributions from the integration limits cancel when taking the gradient, i.e., we need it to cancel the contribution from $M'(s)$.

where $r_<$ and $r_>$ are the smaller and larger of s and r and γ is the angle between $\mathbf{s}$ and $\mathbf{r}$. This was maybe overkill here, since because of the isotropy of $\xi(r)$ only the $\ell = 0$ term survives in (7.41):

$$\begin{aligned} F(\mathbf{s}) &= \frac{-1}{4\pi}\sum_{\ell=0}^{\infty}\int r^2 dr \frac{r_<^\ell}{r_>^{\ell+1}}\xi(r)\int d\phi\int_{-1}^{1} d\cos\gamma L_\ell(\cos\gamma) \\ &= -\int_0^\infty r^2 dr\frac{1}{r_>}\xi(r) \;=\; -\frac{1}{s}\int_0^s r^2\xi(r)dr - \int_s^\infty r\xi(r)dr \\ &= -\frac{1}{s}J_3(s) - K_2(s)\,, \end{aligned} \tag{7.43}$$

where we introduced the J_ℓ and K_ℓ integrals (no relation to Bessel functions although the notation looks the same)

$$J_\ell(x) \;\equiv\; \int_0^x \xi(y)y^{\ell-1}dy\,, \qquad K_\ell(x) \;\equiv\; \int_x^\infty \xi(y)y^{\ell-1}dy\,, \tag{7.44}$$

whose derivatives are

$$J'_\ell(x) \;=\; x^{\ell-1}\xi(x)\,, \qquad K'_\ell(x) \;=\; -x^{\ell-1}\xi(x)\,. \tag{7.45}$$

We see that $F(\mathbf{s}) = F(s)$ is isotropic. Likewise,

$$\begin{aligned} G(s) &= -\frac{1}{s}\int_0^s r^2F(r)dr - \int_s^\infty rF(r)dr \\ &= \frac{1}{s}\int_0^s rdr\int_0^r y^2\xi(y)dy + \frac{1}{s}\int_0^s r^2dr\int_0^\infty \xi(y)ydy + \int_s^\infty [J_3(r)+rK_2(r)]dr \\ &= \frac{1}{s}\int_0^s dy\,y^2\xi(y)\int_y^s rdr + \frac{1}{s}\int_0^\infty dy\,y\xi(y)\int_0^{\min(s,y)} r^2dr + \int_s^\infty [J_3(r)+rK_2(r)]dr \\ &= \frac{s}{2}J_3(s) - \frac{1}{6s}J_5(s) + \frac{s^2}{3}K_2(s) + \int_s^\infty [J_3(r)+rK_2(r)]dr\,. \end{aligned} \tag{7.46}$$

Next we have to do the $\partial/\partial z$'s. Here z is the line-of-sight direction. Since $s^2 = x^2+y^2+z^2$, $2sds = 2xdx + 2ydy + 2zdz$ and

$$\frac{\partial s}{\partial z} \;=\; \frac{z}{s} \;=\; \cos\vartheta_{\mathbf{s}} \;=\; \mu\,. \tag{7.47}$$

We get (**exercise**)

$$\begin{aligned} \partial_z^2 F(\mathbf{s}) &= \ldots = \;\; \mu^2\xi(s) + (1-3\mu^2)\frac{J_3(s)}{s^3} \\ \partial_z^4 G(\mathbf{s}) &= \ldots = \;\; \mu^4\xi(s) + (\tfrac{3}{2} - 9\mu^2 + \tfrac{15}{2}\mu^4)\frac{J_3(s)}{s^3} + (-\tfrac{3}{2} + 15\mu^2 - \tfrac{35}{2}\mu^4)\frac{J_5(s)}{s^5} \end{aligned} \tag{7.48}$$

and finally

$$\begin{aligned} \xi^{(z)}(\mathbf{s}) &\approx \xi(s) + 2\beta\partial_z^2F(\mathbf{s}) + \beta^2\partial_z^4G(\mathbf{s}) \\ &= \left[1+2\beta\mu^2+\beta^2\mu^4\right]\xi(s) + \left[\beta\left(2-6\mu^2\right) + \beta^2\left(\tfrac{3}{2} - 9\mu^2 + \tfrac{15}{2}\mu^4\right)\right]\frac{J_3(s)}{s^3} \\ &\quad +\beta^2\left[-\tfrac{3}{2} + 15\mu^2 - \tfrac{35}{2}\mu^4\right]\frac{J_5(s)}{s^5} \\ &= \left[1+\beta\left(\tfrac{4}{3}L_2 + \tfrac{2}{3}\right) + \beta^2\left(\tfrac{8}{35}L_4 + \tfrac{4}{7}L_2 + \tfrac{1}{5}\right)\right]\xi(s) + \left[\beta\left(-4L_2\right) + \beta^2\tfrac{12}{7}(L_4 - L_2)\right]\frac{J_3(s)}{s^3} \\ &\quad +\beta^2\left(-4L_4\right)\frac{J_5(s)}{s^5}\,, \end{aligned} \tag{7.49}$$

so that the multipoles of the correlation function are

$$\begin{aligned}
\xi_0^{(z)}(s) &= \left(1+\tfrac{2}{3}\beta+\tfrac{1}{5}\beta^2\right)\xi(s) \\
\xi_2^{(z)}(s) &= \left(\tfrac{4}{3}\beta+\tfrac{4}{7}\beta^2\right)\xi(s) + \left(-4\beta-\tfrac{12}{7}\beta^2\right)\frac{J_3(s)}{s^3} \;=\; \left(\tfrac{4}{3}\beta+\tfrac{4}{7}\beta^2\right)\left(\xi-\bar{\xi}\right) \\
\xi_4^{(z)}(s) &= \tfrac{8}{35}\beta^2\xi(s) + \tfrac{12}{7}\beta^2\frac{J_3(s)}{s^3} - 4\beta^2\frac{J_5(s)}{s^5}\,,
\end{aligned} \tag{7.50}$$

where $\bar{\xi}$ is the volume average (Eq. 1.15) of ξ.

Since $\xi(r)$ typically decreases with increasing r, $\xi(r) < \bar{\xi}(r)$ and the quadrupole $\xi_2^{(z)}(s)$ is negative. This corresponds to a flattened shape. We see this in Fig. 20 for large scales ($r_p > 10\,h^{-1}$Mpc). (At small scales the linear theory does not apply, and we get a positive quadrupole, i.e., elongation along the line of sight, due to the Finger-of-God effect.)

From the measured $\xi^{(z)}(r_\pi, r_p) = \xi^{(z)}(s,\mu)$ one can determine the multipoles

$$\xi_\ell^{(z)}(s) \;=\; \frac{2\ell+1}{2}\int_{-1}^{1} d\mu\, L_\ell(\mu)\xi^{(z)}(s,\mu)\,, \tag{7.51}$$

the volume average of the redshift-space monopole,

$$\bar{\xi}_0^{(z)}(s) \;\equiv\; \frac{3}{s^3}\int_0^s \xi_0^{(z)}(r)r^2dr \;=\; \left(1+\tfrac{2}{3}\beta+\tfrac{1}{5}\beta^2\right)\bar{\xi}(s)\,, \tag{7.52}$$

and the quadrupole-to-monopole ratio

$$q(s) \;\equiv\; \frac{\xi_2^{(z)}(s)}{\bar{\xi}_0^{(z)}(s) - \xi_0^{(z)}(s)} \;=\; \frac{-\frac{4}{3}\beta-\frac{4}{7}\beta^2}{1+\frac{2}{3}\beta+\frac{1}{5}\beta^2}\,, \tag{7.53}$$

where the second equalities are the linear results, which should apply at large s. The last equality can be used to determine β from the measured $q(s)$ at large s.

In [33], eBOSS used the range $20\,h^{-1}\text{Mpc} < s < 130\,h^{-1}\text{Mpc}$ to determine[42] $\beta = 0.473\pm0.044$ for the effective redshift of $z_{\rm eff} = 0.698$ using the redshift bin $0.6 < z < 1.0$. The linear approximation does not hold at smaller separations and at larger separations the correlations are so small, and thus inaccurate, that they do not add anything useful. The result is consistent with the ΛCDM model with standard general relativity.

Exercise: Relate the multipoles of the 3D correlation function and power spectrum to each other. Assuming the correlation function and power spectrum have a directional dependence only on the angle ϑ with respect to the line-of-sight direction, they can be expanded

$$\xi(r,\vartheta) = \sum_\ell \xi_\ell(r)L_\ell(\cos\vartheta) \qquad \text{and} \qquad P(r,\vartheta) = \sum_\ell P_\ell(r)L_\ell(\cos\vartheta)\,. \tag{7.54}$$

Using the fact that ξ and P form a 3D Fourier transform pair, show that

$$\xi_\ell(r) \;=\; \frac{i^\ell}{2\pi^2}\int k^2dk j_\ell(kr)P_\ell(k)\,. \tag{7.55}$$

Hint: I generalized the expansion to a spherical harmonics expansion and used (5.30),

$$e^{i\mathbf{k}\cdot\mathbf{x}} = 4\pi\sum_{\ell m} i^\ell j_\ell(kx)Y_{\ell m}(\hat{\mathbf{x}})Y^*_{\ell m}(\hat{\mathbf{k}})\,. \tag{7.56}$$

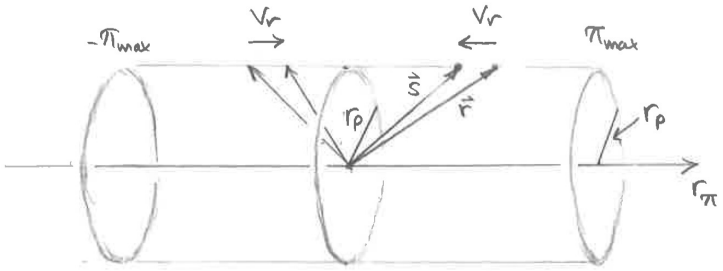


Figure 23: Fixed r_p corresponds to a cylinder centered on the reference galaxy. Redshift distortions move the relative positions of the other galaxies in the r_π direction.

7.3.2 Projected correlation function

In the plane-parallel approximation the redshift effects move the galaxies along the $s_\parallel$ (r_π in Fig. 23) direction, so that $s_\perp = r_\perp$ (r_p in Fig. 23). Thus when we integrate over $s_\parallel$,

$$w_p(r_\perp) \equiv \int_{-\pi_{\max}}^{\pi_{\max}} \xi^{(z)}(s_\parallel, s_\perp) ds_\parallel = \int_{-\pi_{\max}}^{\pi_{\max}} \xi(r(r_\parallel, r_\perp)) dr_\parallel \,, \tag{7.57}$$

where $\xi(r(r_\parallel, r_\perp)) \equiv \xi(\sqrt{r_\parallel^2 + r_\perp^2})$, these effects cancel and we get the same integral from $\xi^{(z)}$ and ξ (we catch the same galaxies along the line corresponding to fixed $r_\perp$). The function $w_p(r_\perp)$ is called the *projected correlation function* and $\pi_{\max}$ should be large enough that correlations are negligible beyond, so that we can replace $\pi_{\max}$ with ∞. Since

$$r = \sqrt{r_\parallel^2 + r_\perp^2} \,, \qquad r_\parallel = \sqrt{r^2 - r_\perp^2} \qquad \text{and} \qquad dr_\parallel = \frac{r dr}{\sqrt{r^2 - r_\perp^2}} \tag{7.58}$$

we have

$$w_p(r_\perp) = 2 \int_{r_\perp}^{\infty} \xi(r) \frac{r dr}{\sqrt{r^2 - r_\perp^2}} \,. \tag{7.59}$$

The factor 2 comes from including both negative and positive r_π.

We can determine w_p from the measured $\xi^{(z)}$ according to (7.57) and solve the real-space correlation function $\xi(r)$ from (7.59) (this is based on the assumption that ξ is isotropic). Eq. (7.59) is a version of Abel's integral equation ([8], Exercise 16.2.7). Write it as

$$w_p(y) = \int_y^{\infty} \xi(r) \frac{2 r dr}{(r^2 - y^2)^\alpha} \,. \tag{7.60}$$

(Here $\alpha = \frac{1}{2}$, but I felt the solution became more clear from this generalization.) We solve it (**exercise**) by multiplying with $(y^2 - r_0^2)^{\alpha - 1}$ and integrating

$$\int_{r_0}^{\infty} y dy \, (y^2 - r_0^2)^{\alpha-1} w_p(y) = \int_{r_0}^{\infty} y dy \, (y^2 - r_0^2)^{\alpha-1} \int_y^{\infty} \xi(r) \frac{2 r dr}{(r^2 - y^2)^\alpha} \,. \tag{7.61}$$

This two-dimensional integral covers a wedge in the (y, r) plane, where $r > y > r_0$, so it rearranges into

$$\int_{r_0}^{\infty} r dr \, \xi(r) \int_{r_0}^{r} 2 y dy \, \frac{(y^2 - r_0^2)^{\alpha-1}}{(r^2 - y^2)^\alpha} \,. \tag{7.62}$$

[42] This value is actually quoted for $f\sigma_8$, which is almost the same as $\beta = f/b$, since for galaxies $b \approx 1/\sigma_8$ as explained in a footnote in Sec. 6.1.

The second integral gives the beta function $B(\alpha, 1-\alpha)$ with the substitution $x = (y^2 - r_0^2)/(r^2 - r_0^2)$. Since $B(\frac{1}{2}, \frac{1}{2}) = \pi$, we have

$$\int_{r_0}^{\infty} r dr\, \xi(r) \;=\; \frac{1}{\pi} \int_{r_0}^{\infty} y dy \frac{w_p(y)}{\sqrt{y^2 - r_0^2}}\,. \tag{7.63}$$

Partial integration and derivating with respect to r_0 gives the final result, the *deprojected correlation function* (**exercise**)

$$\xi(r) \;=\; -\frac{1}{\pi} \int_{r}^{\infty} \frac{w_p'(y) dy}{\sqrt{y^2 - r^2}}\,. \tag{7.64}$$

These results apply both at small and large scales, they do not require ξ to be small.

7.4 Small scales

At small scales where structures have collapsed and $\xi(r)$ is large, galaxies are orbiting the centers of these structures. The radial velocities are more or less random and characterized by their dispersion σ_v. Their effect on the redshift-space correlation function becomes a convolution with the radial velocity distribution. Should add a discussion of this here some time.

7.5 Redshift space in Friedmann–Robertson–Walker universe

We should take into account the metric of the background universe. We review and extend here some results from Cosmology I.

The metric of a homogeneous and isotropic universe is the Robertson–Walker metric

$$\begin{aligned} ds^2 &= -dt^2 + a^2(t) \left[\frac{dr^2}{1 - Kr^2} + r^2 d\vartheta^2 + r^2 \sin^2\vartheta\, d\varphi^2 \right] \\ &= -dt^2 + a^2(t) \Big[d\chi^2 + f_K^2(\chi) \big(d\vartheta^2 + \sin^2\vartheta\, d\varphi^2 \big) \Big]\,, \end{aligned} \tag{7.65}$$

where

$$f_K(\chi) \;\equiv\; \begin{cases} K^{-1/2} \sin(K^{1/2}\chi)\,, & (K > 0) \\ \chi\,, & (K = 0) \\ |K|^{-1/2}\, \sinh(|K|^{1/2}\chi)\,, & (K < 0) \end{cases} \tag{7.66}$$

K is a constant determining the spatial curvature of the universe, $a(t)$ is the scale factor describing the expansion of the universe, and r and χ are two possible choices of radial coordinate. The spatial curvature radius of the universe at time t is $a(t)/\sqrt{|K|}$, or a comoving curvature radius of $r_{\rm curv} = 1/\sqrt{|K|}$. We use the normalization $a_0 = 1$ so that $a = 1/(1+z)$.

The relation of the relevant distance concepts to the redshift are: *comoving distance*

$$D_C(z) = \int_0^z \frac{dz'}{H(z')}\,, \tag{7.67}$$

angular diameter distance

$$D_A(z) = \frac{1}{1+z} f_K \left(\int_0^z \frac{dz'}{H(z')} \right)\,, \tag{7.68}$$

and *comoving angular diameter distance*

$$D_M(z) = f_K \left(\int_0^z \frac{dz'}{H(z')} \right)\,. \tag{7.69}$$

These are related to the coordinates used in the metric as

$$D_C(z) = \chi \qquad \text{and} \qquad D_M(z) = r\,. \tag{7.70}$$

We observe the separation $\mathbf{s} = (s_\parallel, s_\perp)$ of two galaxies as an angular separation $\Delta\vartheta$ on the sky and a redshift difference Δz, so that

$$s_\parallel \approx \frac{\Delta z}{H(z)} \qquad \text{and} \qquad s_\perp \approx D_M(z)\Delta\vartheta\,. \tag{7.71}$$

Both results are small-separation approximations, for $s_\parallel$ we approximated $H \approx const$ within Δz and for $s_\perp$ we used the small-angle approximation (for $K = 0$, $\sin(\frac{1}{2}\Delta\vartheta) \approx \frac{1}{2}\Delta\vartheta$; for $K \neq 0$ something different).

The Hubble parameter $H(t) \equiv \dot{a}/a$ is given by the Friedmann equation

$$H^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2}\,. \tag{7.72}$$

For a universe containing radiation, matter, and dark energy with equation of state[43]

$$w \equiv w_0 + w_a(1-a)\,, \tag{7.73}$$

the Hubble parameter evolves (**exercise**) as

$$H(z) = H_0\sqrt{\Omega_r(1+z)^4 + \Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_w(1+z)^{3(1+w_0+w_a)}e^{-3w_a z/(1+z)}}\,. \tag{7.74}$$

where Ω_r, Ω_m, and Ω_w are density parameters giving the present density of radiation, matter, and dark energy as a fraction of the critical energy density, and

$$\Omega_k \equiv -\frac{K}{H_0^2}\,. \tag{7.75}$$

For a flat universe,

$$D_M(z) = D_C(z) = r = \chi \qquad (K = 0)\,, \tag{7.76}$$

the spatial geometry is Euclidean, and we can convert the observed coordinates z, θ, ϕ of a galaxy into Cartesian redshift space coordinates simply by

$$x = r\sin\theta\cos\phi\,, \quad y = r\sin\theta\sin\phi\,, \quad z = r\cos\theta\,, \tag{7.77}$$

where

$$r = D_C(z)\,. \tag{7.78}$$

We thus work with comoving coordinates, so that structures which expand with the universe have fixed size.

However, we need to know the cosmology to know $D_C(z)$. What should we do, if we are trying to determine the cosmology from the galaxy distribution? The usual approach is to assume a *fiducial cosmology* for converting the observed galaxy positions and redshifts into Cartesian redshift coordinates. To compare the data to predictions from other cosmologies, one then calculates theoretically how the observed correlation function and power spectrum should

[43] This is not a physically motivated equation of state for dark energy (there aren't any), but just a phenomenological two-parameter model that we will try to fit to the data for the redshift range $z = 0$ to a few. For larger redshifts dark energy should have a negligible effect on the geometry of spacetime (if its equation-of-state parameter w stays sufficiently less than 0).

appear if the conversion to Cartesian coordinates is done using this (wrong) fiducial cosmology. This works best if the fiducial cosmology is not too far from the true cosmology.

For an open ($K < 0$) or closed ($K > 0$) universe the situation is trickier. Fortunately Planck and BAO observations constrain the universe to be close to flat, $|\Omega_k| < 0.0026$ (68% C.L.) [34], so that the curvature radius is much larger than the present Hubble distance,

$$r_{\rm curv} \;=\; H_0^{-1}/\sqrt{|\Omega_k|} \;>\; 19.6 H_0^{-1} \;=\; 58\,800\,h^{-1}{\rm Mpc} \tag{7.79}$$

(this assumes extended ΛCDM cosmology, where this curvature is the only extension). Thus the fiducial cosmology can be chosen to be flat. We still need to consider how the curvature affects the theoretical predictions. Because of evolution effects, for deep surveys the correlation functions and powers spectra are estimated separately for different redshift ranges (redshift bins). To assess how significantly the geometry within the survey volume of such a bin $[z_{\rm min}, z_{\rm max}]$ deviates from flat one has to compare the radial comoving extent $D_C(z_{\rm max}) - D_C(z_{\rm min})$, which is usually less than $1000\,h^{-1}$Mpc, to $r_{\rm curv}$. For the correlation function we are usually only interested in galaxy separations less than $r_{\rm max} \sim 200\,h^{-1}{\rm Mpc} \ll r_{\rm curv}$, so that the discussion of $\xi(r)$in Sec. 1 based on Euclidean geometry should hold well. Curvature affects the expansion rate and therefore also the growth of density perturbations, and therefore Ω_K can be constrained indirectly by measurements at small scales where the direct geometrical effect of curvature is negligible.

The other concern regarding the analysis presented in Secs. 7.1–7.3 is that the Hubble parameter is a function of redshift. This analysis had to do with how redshift differences are converted to radial (line-of-sight) distance differences. Instead of the $s_\parallel = H_0^{-1}\Delta z$ of (7.2) we now have

$$s_\parallel = \int_{z_1}^{z_2} \frac{dz'}{H(z')} \approx H(z)^{-1}\Delta z \tag{7.80}$$

so the analysis and results for correlation function and power spectrum multipoles hold to the extent H can be approximated as a constant within the redshift bin. Of course also $\xi(r)$, $P(k)$ and $\beta = f/b$ evolve in time and thus depend on redshift, so that we measure some kind of average value over the redshift bin. These evolution effects can be minimized by using narrow redshift bins, but this way we also lose statistics especially in the line-of-sight direction, and the thickness of the redshift bin should be at least as large as $r_{\rm max}$.

The results on $\beta = f/b$ from galaxy surveys are usually presented as $f(z)\sigma_8(z)$, where σ_8 is the linear theory prediction for the rms mass density fluctuation (square root of variance) in $R = 8\,h^{-1}$Mpc spheres, i.e., the top-hat $\sigma_T(8\,h^{-1}{\rm Mpc})$. Galaxy surveys measure $\sigma_g = b\sigma$, so that multiplying the two measurements, β and σ_g, gives $f\sigma_8$, which is independent of bias, and can be directly compared to the prediction of a cosmological model.[44] The redshift dependence of σ_8 is $\propto D(z)$ (the growth function), so that

$$\sigma_8(z) \;=\; \frac{D(z)}{D(z=0)}\sigma_8\,, \tag{7.81}$$

and the present value is $\sigma_8(z=0) \sim 0.8$ (assuming ΛCDM, Planck finds[45] $\sigma_8 = 0.811 \pm 0.006$ [34], with $\sigma_8 = 0.8120$ for the best-fit ΛCDM model). In Fig. 24 we plot the $f(z)\sigma_8(z)$ of the Planck 2018 best-fit ΛCDM model, and compare it to modified gravity with a larger or smaller growth index γ.

[44] There is a missing piece in this discussion, since the measured σ_g includes also a contribution from non-linear effects. The relation between the nonlinear and linear σ is obtained from numerical simulations.

[45] Note that this is not a *measurement* of σ_8. Planck looks at the early universe, whereas σ_8 relates to the present universe. One fits the parameters of a chosen cosmological model to Planck data, and for this fitted model one calculates theoretically what the present-day σ_8 ought to be.

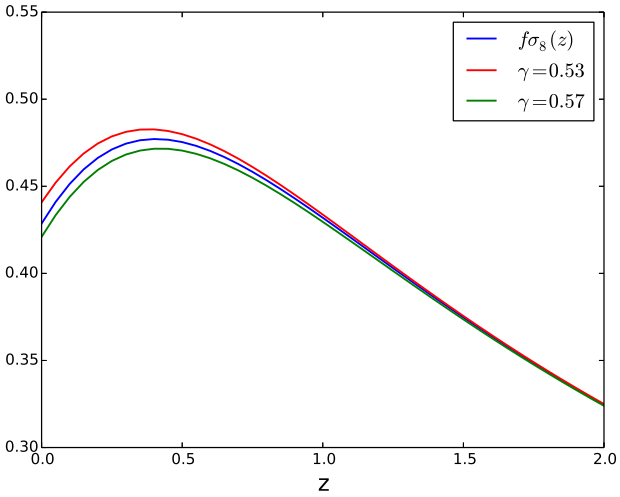


Figure 24: Theoretical $f(z)\sigma_8(z)$ corresponding to the Planck 2018 best-fit ΛCDM model ($\Omega_m = 0.3158$, $\sigma_8 = 0.8120$), compared to what it would be with growth indices $\gamma = 0.53$ and $\gamma = 0.57$. Euclid should be able to distinguish these from each other.

Currently (2021) the most important results on this are from the SDSS BOSS (Baryon Oscillation Spectroscopic Survey) [35] and eBOSS (extended BOSS) [61] surveys. See Fig. 25 and Fig. 27.

7.6 Alcock–Paczyński effect

Consider a spherical group of objects (e.g. galaxies) far away, with negligible peculiar velocities (to avoid the redshift-space distortion), so that it expands with the universe, i.e., forms a comoving sphere with some comoving radius r. The transverse and line-of-sight extents of the sphere are thus equal also in redshift space,

$$s_\parallel = \frac{\Delta z}{H(z)} \;=\; s_\perp = D_M(z)\Delta\vartheta \tag{7.82}$$

so that

$$\frac{\Delta z}{\Delta\vartheta} \;=\; D_M(z)H(z) \;\equiv F_{AP}(z)\,. \tag{7.83}$$

Observing Δz and $\Delta\vartheta$ one can thus determine the Alcock–Paczyński parameter $F_{AP}(z)$ of the universe. One does not need to know the size r of the sphere, only that it has a spherical shape. Alcock and Paczyński[36] proposed this cosmological test in 1979 especially for testing for the presence of a non-zero cosmological constant. Different cosmologies have different $F_{AP}(z)$. It is not necessary to have a sphere (it was used here just for pedacogical reasons); the Alcock–Paczyński test requires just an isotropic distribution, e.g., an isotropic $\xi(r)$. The powerful feature of this test is that it requires no correction for evolution effects.

The weakness of the Alcock–Paczyński test is that it will be contaminated by peculiar velocities, since the isotropy is required in redshift space. A reverse consideration is that if we try to measure the anisotropy (the multipoles) of the redshift-space correlation function or power spectrum, and our fiducial cosmology has a different $F_{AP}(z)$ than the true one, we get the anisotropy wrong. There are thus two contributions to the observed anisotropy of an isotropic distribution: the true redshift-space distortion due to peculiar velocities and this Alcock–Paczyński effect due to using a wrong fiducial cosmology.

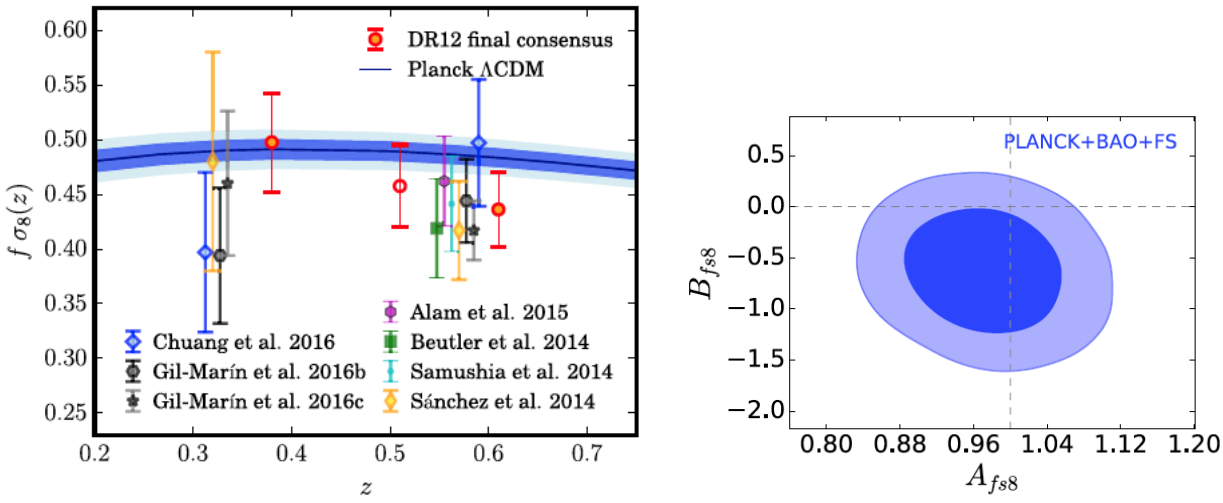


Figure 25: Left: Measurements of $f\sigma_8$ at different redshifts from the BOSS survey. The latest results are the 'DR12 final consensus' from [35]. This is left panel of Fig. 15 in [35]. The curve with blue error band is the prediction of the ΛCDM model with parameters determined from Planck data. This plot can be compared to $D(z)f(z)$ in Fig. 16, except that $D(z)$ has different normalization than $\sigma_8(z)$. Right: Constraints (68% and 95% confidence levels) from BOSS data on a phenomenological two-parameter modification of the growth function: $f\sigma_8 \to f\sigma_8\,[A + B(z - z_p)]$ where the pivot redshift $z_p = 0.51$ was chosen to be at the center of the survey. The data is consistent with no modification, $A = 1$ and $B = 0$, i.e., the prediction from General Relativity. This is Fig. 20 of [35]. See Fig. 27 for an update.

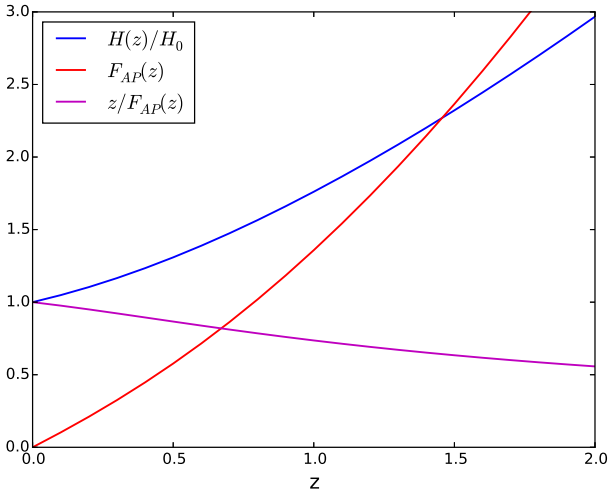


Figure 26: The Hubble parameter $H(z)$ and the Alcock–Paczyński parameter $F_{AP}(z)$ as a function of redshift for the ΛCDM model with $\Omega_m = 0.3$.

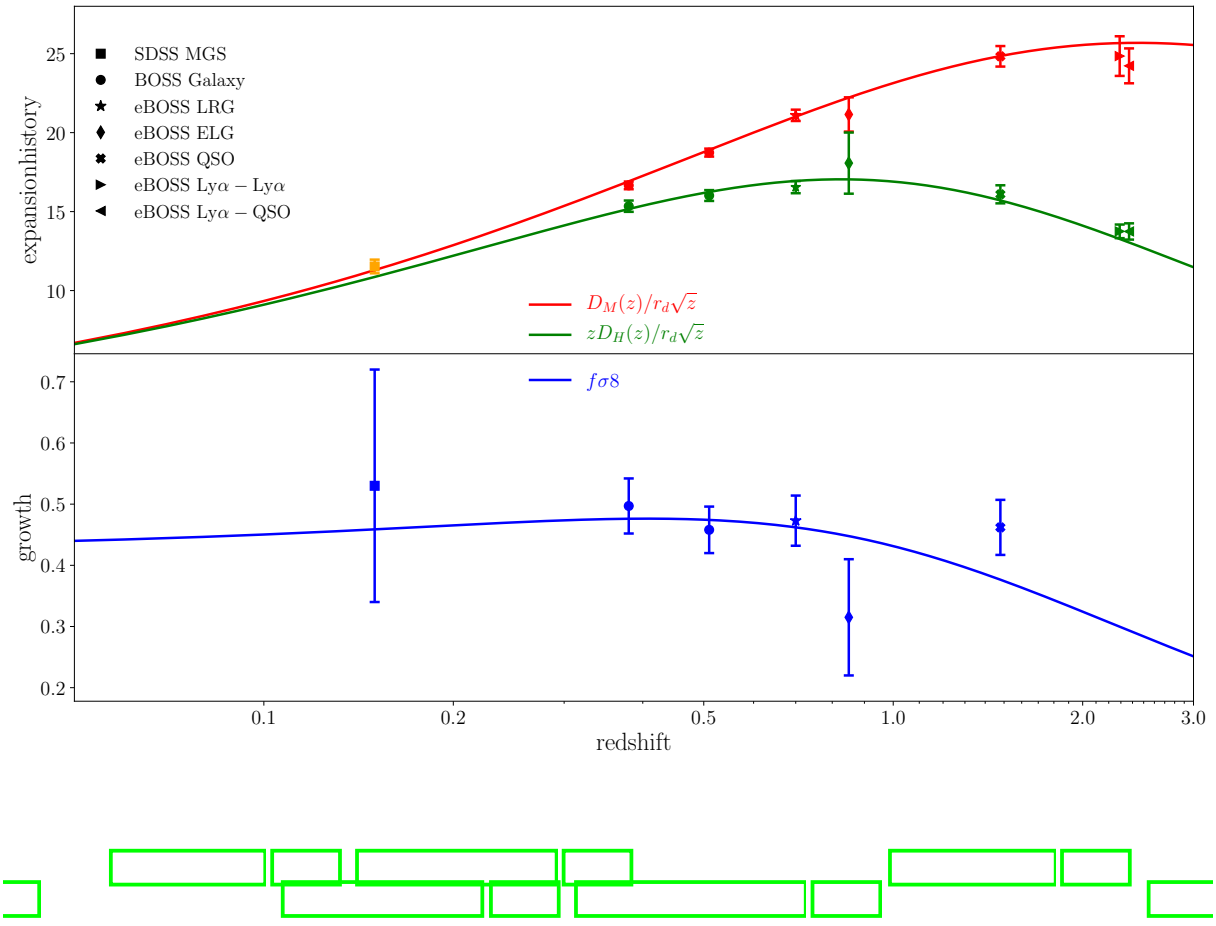


Figure 27: **Bottom panel:** Measurements of $f\sigma_8$ at different redshifts from the eBOSS survey. The fourth data point from the left (effective redshift $z \approx 0.7$) is from the $0.6 < z < 1$ LRG sample of Fig. 21. The blue curve is the theoretical prediction from the Planck 2018 best-fit ΛCDM model (see Fig. 24). Note that here the z axis is logarithmic, whereas in Fig. 24 it was linear. **Top panel:** Distance, i.e., expansion history, measurements from eBOSS based on BAO as standard ruler – this is discussed in Sec. 10. This is Fig. 1 of [61].

8 Measuring the correlation function

Methods to estimate the correlation function from data are based on its definition as excess probability of finding a galaxy pair. One counts from the data catalog the number $DD(r)$ of pairs of galaxies with separation $|\mathbf{x}_2 - \mathbf{x}_1| \in r \pm \frac{1}{2}\Delta r$, where Δr is the *bin width*. (From here on "separation r" will mean that the separation falls in this bin.) What part of this is excess? It should be compared to the number of pairs we would get for $\xi = 0$, i.e., for a Poisson process. For this we generate a random catalog by throwing points randomly (a Poisson process) into the survey volume and count the corresponding number $RR(r)$ of random pairs. This gives us an estimator

$$1 + \widehat{\xi}(r) \;\equiv\; \frac{DD(r)}{RR(r)}\,. \tag{8.1}$$

This assumes that both the data catalog and random catalog were drawn from an ensemble with the same $\langle\rho\rangle$.

An unavoidable problem is that we do not know the true expectation value $\langle\rho\rangle$. We have to estimate it from the data, replacing it by N_D/V where N_D is the number of data points (observed galaxies) and V the survey volume. The estimator (8.1) assumes then an equal number of random points $N_R = N_D$.

The inherent randomness in both the data catalog and the random catalog introduce a random error in the estimator. The larger the number of points the smaller is the expected error. Since it is much cheaper to increase the number of random points (you need computer time) than the number of data points (you need telescope time), one usually uses $N_R \gg N_D$, e.g., $N_R = 50N_D$, to minimize random errors due to the finite number or random points. One then needs to scale the estimator with the ratio of the total number of random pairs $\frac{1}{2}N_R(N_R-1)$ to that of the data pairs $\frac{1}{2}N_D(N_D-1)$. Thus we have our first estimator

$$1 + \widehat{\xi}_1(r) \;\equiv\; \frac{N_R(N_R-1)}{N_D(N_D-1)}\frac{DD(r)}{RR(r)}\,. \tag{8.2}$$

Generalization to anisotropic $\xi(\mathbf{r})$. For simplicity, this chapter is written for $\xi(r)$, as if we could assume that the correlation function $\xi(\mathbf{r})$ is isotropic. In reality, for a galaxy redshift survey, the observations are in redshift space, where $\xi^{(z)}(\mathbf{z})$ is anisotropic. In fact, the whole discussion of this chapter generalizes trivially to a 3D function $\xi^{(z)}(\mathbf{z})$ or the 2D function $\xi^{(z)}(s,\mu)$, by just replacing r everywhere with $\mathbf{r}$ (or $\mathbf{z}$) and taking $\mathbf{r}$ to refer to some bin in the 3D space, which we need not specify before we apply the results. Thus the bin could represent, e.g., a 2D bin $s \pm \Delta s$, $\mu \pm \Delta\mu$.

Monte Carlo integration. The use of the random catalog can be thought of as Monte Carlo integration. The probability of finding a galaxy pair (see Eq. 2.10) with separation r is

$$d\mathbf{P} \;\equiv\; \langle\rho\rangle^2[1+\xi(r)]dV_1dV_2\,. \tag{8.3}$$

The expected number $DD(r)$ of pairs of galaxies with separation r is thus

$$\langle DD(r)\rangle \;=\; \tfrac{1}{2}\langle\rho\rangle^2\int[1+\xi(r)]\Theta_{12}(r)dV_1dV_2\,, \tag{8.4}$$

where

$$\Theta_{12}(r) \;\equiv\; 1 \quad \text{if } |\mathbf{x}_2-\mathbf{x}_1| \in r \pm \tfrac{1}{2}\Delta r, \text{ otherwise } 0\,, \tag{8.5}$$

which should be compared to

$$\langle\rho\rangle^2 V_p(r) \;\equiv\; \tfrac{1}{2}\langle\rho\rangle^2\int\Theta_{12}(r)dV_1dV_2\,, \tag{8.6}$$

where $V_p(r)$ is the *volume in pair space* $\{\{\mathbf{x}_1, \mathbf{x}_2\}\}$ corresponding to pairs whose separation falls into the bin. (The factor $\frac{1}{2}$ is there so that pairs are only counted once, i.e., to cancel the effect of both dV_1 and dV_2 going through the same locations.) This gives us an estimator

$$1 + \widehat{\xi}(r) \;\equiv\; \frac{DD(r)}{\langle\rho\rangle^2 V_p(r)}\,. \tag{8.7}$$

How to calculate the volume $V_p(r)$? It is a complicated volume in the $2d$-dimensional pair space. A standard method for multidimensional integration over regions of complicated shape (defined by some condition) is *Monte Carlo* integration: One throws random points (a Poisson process) into an enclosing volume of simpler shape and counts those points who satisfy the condition of belonging in the region. This fraction of all points provides an estimate of the fraction the region is of the enclosing volume. The application here is to throw random points into the survey volume and count the number $RR(r)$ of pairs of these random points whose separation is r. For a large enough number N_R of random points,

$$RR(r) \;\approx\; \rho_R^2 V_p(r) \;\approx\; \frac{N_R(N_R-1)}{V^2} V_p(r) \;\equiv\; G_p(r)\tfrac{1}{2}N_R(N_R-1)\,, \tag{8.8}$$

where

$$G_p(r) \;\equiv\; \frac{V_p(r)}{\frac{1}{2}V^2} \tag{8.9}$$

is the fraction of the pair space corresponding to bin r. Here $\rho_R = N_R/V$ is the density of random points and $\frac{1}{2}N_R(N_R-1)$ is the total number of random pairs. Thus we have the estimator

$$1 + \widehat{\xi}(r) \;\equiv\; \frac{\rho_R^2}{\langle\rho\rangle^2}\frac{DD(r)}{RR(r)}\,, \tag{8.10}$$

which is the same as (8.2) when we replace the ratio of squared densities with the ratio of total numbers of pairs.

Since a galaxy survey surveys just a finite volume of a single realization, the $\widehat{\xi}(r)$ obtained differs from the true $\xi(r)$ for other reasons besides the difference between N_D/V and $\langle\rho\rangle$. This random error can be amplified by edge effects.

To demonstrate edge effects, consider an extreme case: Suppose the clustering is such that there is typically one tight cluster, say, of 100 galaxies, per survey volume, and nothing else. If this cluster happens to lie far from the survey boundaries, we get $\frac{1}{2}\times 100\times 99 = 4450$ close pairs. However, if we instead have two such clusters at the survey edge so that only half of each cluster is included in the survey, we get only $2\times(\frac{1}{2}\times 50\times 49) = 2450$ close pairs. Thus these two different samples of the same galaxy distribution give very different estimates for $\xi(r)$ at small r. We can try to minimize such edge effects by using *data-random pairs* instead of random-random pairs. The total number of data-random pairs is $N_D N_R$. Since the random points are not correlated with the data points, the number of data-random pairs $DR(r)$ with separation r is not affected by the correlation $\xi(r)$. Thus the excess of $DD(r)$ compared to $DR(r)$ measures the correlation and we have our second estimator

$$1 + \widehat{\xi}_2(r) \;=\; 1 + \widehat{\xi}_{\rm DP}(r) \;\equiv\; \frac{2N_R}{N_D-1}\frac{DD(r)}{DR(r)} \tag{8.11}$$

(Davis&Peebles [37]). Consider now the earlier example. Whereas the factor $RR(r)$ measured just the volume of an r-shell around an average location, the factor $DR(r)$ measures the volume of an r-shell around an average galaxy. In the case where the galaxies lie close to the edge of the survey these volumes are smaller as the part of the shell that falls outside the survey volume is not included. Thus the estimator $\widehat{\xi}_{\rm DP}(r)$ gives results closer to each other for the two samples.

Other combinations of $DD(r)$, $DR(r)$ and $RR(r)$ can also be used as ξ estimators. Hamilton [38] proposed

$$1 + \widehat{\xi}_3(r) \;=\; 1 + \widehat{\xi}_{\rm H}(r) \;\equiv\; \frac{4N_D N_R}{(N_D-1)(N_R-1)}\frac{DD(r)RR(r)}{DR(r)^2}\,. \tag{8.12}$$

Landy&Szalay [39] proposed

$$\widehat{\xi}_4(r) \;=\; \widehat{\xi}_{\rm LS}(r) \;\equiv\; \frac{N_R(N_R-1)}{N_D(N_D-1)}\frac{DD(r)}{RR(r)} - \frac{N_R-1}{N_D}\frac{DR(r)}{RR(r)} + 1\,. \tag{8.13}$$

Symbolically these four estimators are often denoted (leaving out the normalization factors from the different total numbers of pairs)[46]

$$\frac{DD}{RR}\,,\quad \frac{DD}{DR}\,,\quad \frac{DD\cdot RR}{DR^2}\,,\quad \frac{DD-2DR+RR}{RR}\,, \tag{8.14}$$

but note that here the three first represent estimators for $1+\xi$, the last one just for ξ. We follow now Landy&Szalay [39] to estimate their bias and variance in the limit where the correlations are small and $N_R \gg N_D$.. (Although Landy&Szalay refer to "angular correlation functions", the derivation is independent of the dimension and geometry of space. Of course 3D correlations have the additional issues of redshift distortions and selection function that we do not consider in this section.)

8.1 Bias and variance of different estimators

This section follows Landy&Szalay [39] to estimate the bias and variance of the four ξ estimators. The variance estimate applies only in the limit where correlations are small. The square root of the variance provides our estimate for the statistical error. The bias is a systematic error due to the estimator (but if we can estimate the bias we can try to correct for it).

Define α, β, γ to represent the relative deviation of the $DD(r)$, $DR(r)$, $RR(r)$ counts from their expectation values:

$$\begin{aligned} DD(r) &\equiv \langle DD(r)\rangle[1+\alpha(r)] \\ DR(r) &\equiv \langle DR(r)\rangle[1+\beta(r)] \\ RR(r) &\equiv \langle RR(r)\rangle[1+\gamma(r)] \end{aligned} \tag{8.15}$$

(we will mostly suppress the notation of the dependence on r in the following). By definition

$$\langle\alpha\rangle \;=\; \langle\beta\rangle \;=\; \langle\gamma\rangle \;=\; 0\,. \tag{8.16}$$

The more pairs we have in an r-bin the smaller will be the typical relative deviation from the expectation value. Thus we assume α and β will be small and calculate only up to second order in them. Moreover, we assume that N_R is sufficiently large that we can approximate $\gamma \approx 0$. (I kept the notation for γ here, in case we later want to study the effect of N_R.)

For example, for the estimator (8.11) we get the expectation value and variance (**exercise**)

$$\begin{aligned} \langle 1+\widehat{\xi}_{\rm DP}(r)\rangle &= \frac{2N_R}{N_D-1}\frac{\langle DD\rangle}{\langle DR\rangle}\left\langle\frac{1+\alpha}{1+\beta}\right\rangle \;\approx\; \frac{2N_R}{N_D-1}\frac{\langle DD\rangle}{\langle DR\rangle}\left(1-\langle\alpha\beta\rangle+\langle\beta^2\rangle\right) \\ {\rm var}[\widehat{\xi}_{\rm DP}(r)] &\equiv \left\langle\left(\widehat{\xi}_{\rm DP}-\langle\widehat{\xi}_{\rm DP}\rangle\right)^2\right\rangle \;=\; \langle\widehat{\xi}^2_{\rm DP}\rangle - \langle\widehat{\xi}_{\rm DP}\rangle^2 \\ &\approx \left(\frac{2N_R}{N_D-1}\frac{\langle DD\rangle}{\langle DR\rangle}\right)^2\left(\langle\alpha^2\rangle - 2\langle\alpha\beta\rangle+\langle\beta^2\rangle\right)\,, \end{aligned} \tag{8.17}$$

[46] The last one can be written symbolically as $(D-R)^2/RR$, which suggest an alternative implementation: From the D and R catalogs one constructs a "$D-R$" catalog, which contains all the D and R points, but the random points are assigned negative weight. One then counts the $(D-R)(D-R)$ pairs, where those pairs that are DR make a negative contribution.

where we get $\langle\alpha^2\rangle$, $\langle\alpha\beta\rangle$, and $\langle\beta^2\rangle$ from

$$\begin{aligned}
\langle DD\cdot DD\rangle &= \langle DD\rangle^2(1+\langle\alpha^2\rangle) \quad\Rightarrow\quad \langle\alpha^2\rangle = \frac{\langle DD\cdot DD\rangle - \langle DD\rangle^2}{\langle DD\rangle^2} \\
\langle DR\cdot DR\rangle &= \langle DR\rangle^2(1+\langle\beta^2\rangle) \quad\Rightarrow\quad \langle\beta^2\rangle = \frac{\langle DR\cdot DR\rangle - \langle DR\rangle^2}{\langle DR\rangle^2} \\
\langle DD\cdot DR\rangle &= \langle DD\rangle\langle DR\rangle(1+\langle\alpha\beta\rangle) \quad\Rightarrow\quad \langle\alpha\beta\rangle = \frac{\langle DD\cdot DR\rangle - \langle DD\rangle\langle DR\rangle}{\langle DD\rangle\langle DR\rangle}\,. \qquad (8.18)
\end{aligned}$$

Thus we will need the expectation values $\langle DD\rangle$, $\langle DR\rangle$, $\langle RR\rangle$, $\langle DD\cdot DD\rangle$, $\langle DR\cdot DR\rangle$, $\langle DD\cdot DR\rangle$ (for other estimators also $\langle RR\rangle$ occurs).

Consider N_D data points and N_R random points in survey volume V. Expectation values $\langle\cdot\rangle$ in this subsection refer to *conditional expectation values with a fixed value of* N_D. Divide the volume into K microcells, so that the number of galaxies n_i in microcell i is 0 or 1. The expectation value is $\langle n_i\rangle = N_D/K$.

The number of data pairs with separation r is

$$DD(r) \;=\; \sum_{i<j}^{K} n_i n_j \Theta_{ij}(r)\,, \qquad (8.19)$$

where the sum is over all microcell pairs and

$$\Theta_{ij}(r) \;\equiv\; 1\,, \quad \text{if } |\mathbf{r}_j - \mathbf{r}_i| \in r \pm \tfrac{1}{2}\Delta r, \text{ otherwise } 0\,, \qquad (8.20)$$

picks only those pairs with the correct separation. The sum $\sum\Theta_{ij}(r)$ is thus the number of microcell pairs with separation r, and $n_i n_j$ is 1 exactly when both contain a data point (a galaxy). The expectation value is thus

$$\langle DD(r)\rangle \;=\; \sum_{i<j}^{K} \langle n_i n_j\rangle \Theta_{ij}(r)\,. \qquad (8.21)$$

If there were no correlations, we would have simply

$$\langle n_i n_j\rangle \;=\; \frac{N_D}{K}\frac{N_D-1}{K-1}\,, \qquad (8.22)$$

and we would get

$$\langle DD(r)\rangle \;=\; \frac{N_D(N_D-1)}{2} G_p(r)\,, \qquad (8.23)$$

where

$$G_p(r) \;\equiv\; \frac{2}{K(K-1)}\sum_{i<j}^{K}\Theta_{ij}(r) \qquad (8.24)$$

is the *fraction of microcell pairs* with separation r. For small enough microcells (large K) it is independent of K. By definition, $G_p(r)$ is normalized,

$$\sum_r G_p(r) \;=\; 1\,, \qquad (8.25)$$

when the sum go over r bins that cover all possible separations r that fit in the survey volume.

With correlations, the expectation $\langle n_i n_j\rangle$ is modified by the correlation function $\xi(r)$. Now one has to take into account that here $\xi(r)$ is the true, unknown, correlation function, which

is defined with respect to the unknown $\langle\rho\rangle$, whereas we are considering here a constrained expectation value $\langle DD\rangle$ for a fixed $N_D \neq \langle N_D\rangle$. Thus we need to include a normalization factor C_V to correct for this, and we have

$$\langle n_i n_j\rangle \;=\; \frac{N_D}{K}\frac{N_D-1}{K-1}[1+\tilde{\xi}(r)] \;=\; \frac{N_D}{K}\frac{N_D-1}{K-1}C_V[1+\xi(r)]\,, \tag{8.26}$$

where $\tilde{\xi}(r)$ is the correlation function normalized to the survey density, so that we get

$$\langle DD(r)\rangle \;=\; \frac{N_D(N_D-1)}{2}G_p(r)C_V[1+\xi(r)]\,. \tag{8.27}$$

We can relate the normalization factor to $\xi(r)$ by noting that the total number of pairs is

$$\frac{N_D(N_D-1)}{2} \;=\; \sum_r\langle DD(r)\rangle \;=\; \frac{N_D(N_D-1)}{2}C_V\sum_r G_p(r)[1+\xi(r)] \tag{8.28}$$

so that

$$C_V \;=\; \frac{1}{\sum_r G_p(r)[1+\xi(r)]} \;\equiv\; \frac{1}{1+\xi_V} \tag{8.29}$$

where

$$\xi_V \;\equiv\; \sum_r G_p(r)\xi(r) \tag{8.30}$$

is the *mean expected correlation over the survey pair space* $\{\{\mathbf{x}_1,\mathbf{x}_2\}\}$ *volume* (it is positive, since a finite volume misses some of the large-r negative correlations, and becomes smaller for larger volumes). Thus our final result is

$$\langle DD(r)\rangle \;=\; \frac{N_D(N_D-1)}{2}G_p(r)\frac{1+\xi(r)}{1+\xi_V}\,. \tag{8.31}$$

Note that for $\tilde{\xi}$, we get the mean $\tilde{\xi}_V = 0$, the "integral constraint".

8.1.1 Derivation of $\langle DD\cdot DD\rangle$, $\langle DR\cdot DR\rangle$, and $\langle DD\cdot DR\rangle$

The expectation value for the square of $DD(r)$ is

$$\langle DD\cdot DD\rangle \;=\; \sum_{i<j}^{K}\sum_{k<l}^{K}\langle n_i n_j n_k n_l\rangle\Theta_{ij}(r)\Theta_{kl}(r)\,, \tag{8.32}$$

where both i,j and k,l sum over all pairs. There are three different cases for the terms $\langle n_i n_j n_k n_l\rangle$ depending on how many indices are equal ($i\neq j$ and $k\neq l$ for all of them).

Quadruplets: The first case is when i,j,k,l are all different. We denote by $\sum^*_{ijkl}$ this part of the sum. There are $\frac{1}{2}K(K-1)\times\frac{1}{2}(K-2)(K-3)$ such terms and they have

$$\begin{aligned}
\langle n_i n_j n_k n_l\rangle \;&=\; \frac{N_D}{K}\frac{N_D-1}{K-1}\frac{N_D-2}{K-2}\frac{N_D-3}{K-3}\Big\langle(1+\delta_i)(1+\delta_j)(1+\delta_k)(1+\delta_l)\Big\rangle \\
&=\; \frac{N_D}{K}\frac{N_D-1}{K-1}\frac{N_D-2}{K-2}\frac{N_D-3}{K-3}\Big[1+\langle\delta_i\delta_j\rangle+\langle\delta_k\delta_l\rangle+\langle\delta_i\delta_k\rangle+\langle\delta_i\delta_l\rangle+\langle\delta_j\delta_k\rangle+\langle\delta_j\delta_l\rangle+ \\
&\quad+\langle\delta_i\delta_j\delta_k\rangle+\langle\delta_i\delta_j\delta_l\rangle+\langle\delta_i\delta_k\delta_l\rangle+\langle\delta_j\delta_k\delta_l\rangle+\langle\delta_i\delta_j\delta_k\delta_l\rangle\Big] \\
&=\; \frac{N_D}{K}\frac{N_D-1}{K-1}\frac{N_D-2}{K-2}\frac{N_D-3}{K-3}\Big[1+\tilde{\xi}(r_{ij})+\tilde{\xi}(r_{ik})+\tilde{\xi}(r_{il})+\tilde{\xi}(r_{jk})+\tilde{\xi}(r_{jl})+\tilde{\xi}(r_{kl}) \\
&\quad+\tilde{\zeta}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{x}_k)+\tilde{\zeta}(\mathbf{x}_i,\mathbf{x}_k,\mathbf{x}_l)+\tilde{\zeta}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{x}_l)+\tilde{\zeta}(\mathbf{x}_j,\mathbf{x}_k,\mathbf{x}_l)+ \\
&\quad+\tilde{\xi}(r_{ij})\tilde{\xi}(r_{kl})+\tilde{\xi}(r_{ik})\tilde{\xi}(r_{jl})+\tilde{\xi}(r_{il})\tilde{\xi}(r_{jk})+\tilde{\eta}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{x}_k,\mathbf{x}_l)\Big]
\end{aligned} \tag{8.33}$$

where $\delta_i \equiv \delta(\mathbf{x}_i)$ is the density perturbation relative to the mean galaxy number density of the survey (not relative to its ensemble average, which we do not know),

$$1 + \tilde{\xi}(r) \;\equiv\; \frac{1+\xi(r)}{1+\xi_V} \tag{8.34}$$

represents the "renormalized" correlation, $\tilde{\zeta}$ is the 3-point correlation and the last line is the 4-point correlation, where $\tilde{\eta}$ is the reduced 4-point correlation. The 3-point and 4-point correlations are discussed in Sec. 11. For Gaussian perturbations, $\tilde{\xi}$ and $\tilde{\eta}$ vanish. We shall now work in the small-ξ and Gaussian approximation (these are related, since the Gaussian approximation requires that ξ is small; and on the other hand, if the primordial perturbations were Gaussian, which is a natural assumption supported by cosmological data, they would have remained Gaussian while the perturbations were still small), so that only the first line of the last form in (8.33) survives.

The fraction of microcell quadruplets (pairs of pairs) that satisfy $r_{ij} \in r \pm \frac{1}{2}\Delta r$ and $r_{kl} \in r \pm \frac{1}{2}\Delta r$ is $G_p(r)^2$. In the limit of large K[47] the number of other index quadruplets is negligible compared to those where all indices have different values, so we have

$$\sum_{ijkl}^{*} \Theta_{ij}(r)\Theta_{kl}(r) \;=\; \frac{K(K-1)(K-2)(K-3)}{4} G_p(r)^2 \,. \tag{8.35}$$

Due to the Θ constraints in (8.35),

$$\tilde{\xi}(r_{ij}) \;=\; \tilde{\xi}(r_{kl}) = \tilde{\xi}(r) \,, \tag{8.36}$$

whereas the other 4 terms range over different values of r, and we approximate them with the mean over survey $\tilde{\xi}_V = 0$. Thus

$$\begin{aligned} \langle n_i n_j n_k n_l \rangle \;&\approx\; \frac{N_D(N_D-1)(N_D-2)(N_D-3)}{K(K-1)(K-2)(K-3)} \left[1 + 2\tilde{\xi}(r)\right] \\ &=\; \frac{N_D(N_D-1)(N_D-2)(N_D-3)}{K(K-1)(K-2)(K-3)} \frac{1+2\xi(r)-\xi_V}{1+\xi_V} \,. \end{aligned} \tag{8.37}$$

Triplets: The second case is where we have three different values for the indices, so that k or l is equal to i or j, i.e., we have microcell triplets (one of the three points is shared with the two pairs, see Fig. 28). How many such terms do we have in the sum? Altogether there are $K(K-1)(K-2)/6$ triplets of index values, where the three values are all different. But the same triplet occurs several times in the sum. Let the values by $1, 2, 3$. There are six different ways they can be assigned to (i, j, k, l) that satisfy the conditions $i < j$ and $k < l$: $(1, 2, 1, 3)$, $(1, 2, 2, 3)$, $(1, 3, 1, 2)$, $(1, 3, 2, 3)$, $(2, 3, 1, 2)$, and $(2, 3, 1, 3)$. Thus there are $K(K-1)(K-2)$ such terms ($6 \times \frac{1}{6} = 1$). Denote by $\sum^*_{ijk}$ this part of the sum, where i is the shared microcell.

The triplet part of (8.32) becomes

$$\sum_{ijk}^{*} \langle n_i n_j n_k \rangle \Theta_{ij}(r)\Theta_{ik}(r) \,. \tag{8.38}$$

The $\Theta_{ij}(r)\Theta_{ik}(r)$ factor selects those terms where the other points are separated by r from the shared point. Let $G_t(r)$ denote the fraction that satisfy this condition. Thus

$$\sum_{ijk}^{*} \Theta_{ij}(r)\Theta_{ik}(r) \;=\; K(K-1)(K-2) G_t(r) \,. \tag{8.39}$$

[47] The microcell approach is essentially the limit $K \to \infty$, so we could write K^4 instead of $K(K-1)(K-2)(K-3)$ etc.

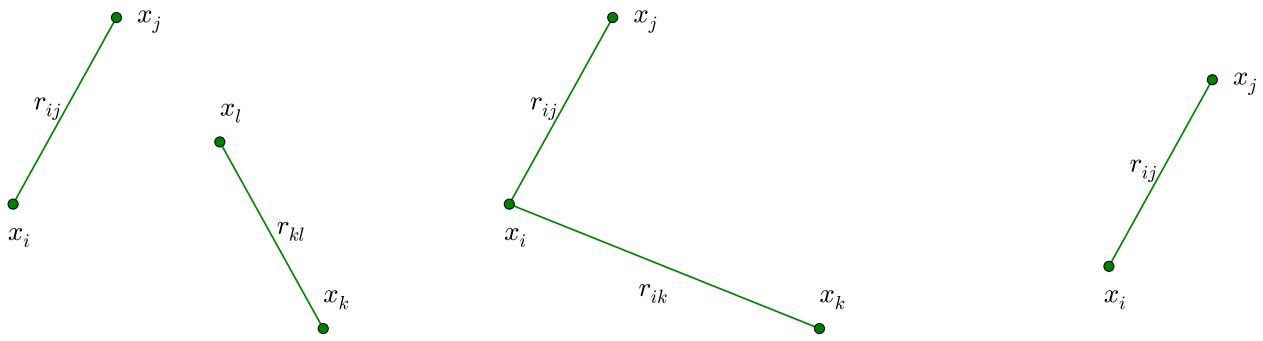


Figure 28: A quadruplet (two pairs), a triplet, and a pair of microcells.

Since most triplets have arms of unequal length, $\sum G_t(r) \ll 1$. It is clear that $G_t(r)$ is of the same magnitude as $G_p(r)^2$, since there are now two conditions to satisfy; but we see later that $G_t > G_p^2$. The difference between G_t and G_p^2 becomes important later. The Θ constraints in (8.39) require that $r_{ij} = r_{ik} = r$ and since cells j and k are connected by two steps of length r, r_{jk} has to lie between 0 and $2r$. We approximate also $\xi(r_{jk})$ with $\xi(r)$. Thus

$$\begin{aligned}\langle n_i n_j n_k\rangle &\approx \frac{N_D(N_D-1)(N_D-2)}{K(K-1)(K-2)}\left[1+3\tilde{\xi}(r)+\tilde{\zeta}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{x}_k)\right] \\ &\approx \frac{N_D(N_D-1)(N_D-2)}{K(K-1)(K-2)}\frac{1+3\xi(r)-2\xi_V}{1+\xi_V}\,. \end{aligned} \tag{8.40}$$

Pairs: The third case is when $i = k$ and $j = l$, i.e., the two pairs are the same, so this part is just a single sum over pairs,

$$\sum_{i<j}\langle n_i n_j\rangle\Theta_{ij}(r)\Theta_{ij}(r) \;=\; \frac{N_D(N_D-1)}{2}G_p(r)\left[1+\tilde{\xi}(r)\right]\,, \tag{8.41}$$

from (8.19) and (8.31), as $\Theta_{ij}(r)\Theta_{ij}(r) = \Theta_{ij}(r)$.

Altogether we have:

$$\begin{aligned}\langle DD\cdot DD\rangle &\approx \frac{N_D(N_D-1)(N_D-2)(N_D-3)}{4}G_p(r)^2\frac{1+2\xi(r)}{1+2\xi_V} \\ &\quad +N_D(N_D-1)(N_D-2)G_t(r)\frac{1+3\xi(r)}{1+3\xi_V}+\frac{N_D(N_D-1)}{2}G_p(r)\frac{1+\xi(r)}{1+\xi_V}\,. \end{aligned} \tag{8.42}$$

(Since ξ_V is small, $1/(1+\xi_V) \approx 1-\xi_V$. I have moved thus ξ_V up or down.)

Thus we have that (**exercise**)

$$\langle\alpha^2\rangle \;\approx\; \frac{2}{N_D(N_D-1)}\left\{2(N_D-2)\left[\frac{G_t(r)}{G_p(r)^2}\frac{1+\xi(r)}{1+\xi_V}-1\right]+\frac{1}{G_p(r)}\frac{1+\xi_V}{1+\xi(r)}-1\right\}\,. \tag{8.43}$$

The $\langle DR\rangle$, $\langle DR\cdot DR\rangle$, $\langle DD\cdot DR\rangle$ are calculated in a similar way, noting that D and R

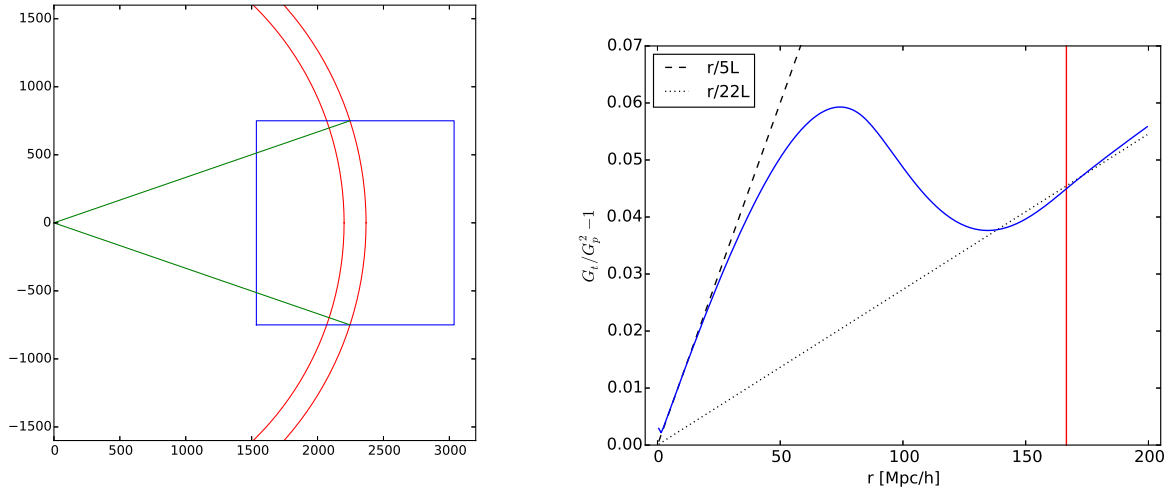


Figure 29: Left: Simple survey geometry (chosen for ease of simulation) corresponding to a redshift bin. The units on the axes are h^{-1}Mpc and the observer is on the left at $(0,0,0)$. The two red curves represent comoving distances to cosmological redshifts $z = 0.95$ and $z = 1.05$ in the ΛCDM model with $\Omega_m = 0.32$. The survey volumes consists of the intersection of the shell between these redshifts with the $(1500\,h^{-1}\mathrm{Mpc})^3$ box illustrated in blue, except points where the line of sight leaves the box before the outer shell are excluded (illustrated in green). The thickness of the shell is $166.6\,h^{-1}$Mpc (this is a bit narrow, giving poor statistics in the radial direction for large r, none beyond $r = 166.6\,h^{-1}$Mpc; so a larger Δz would be better). Right: The difference between G_t and G_p^2 in this geometry, calculated by assigning random points into the survey volume with a Poisson process and counting pairs and triplets. The vertical red line represents the shell thickness. The dashed and dotted lines are simple fits (by hand) to the observed small-r and large-r behavior.

points are uncorrelated with each other. We get

$$\begin{aligned}
\langle DR\rangle &= N_D N_R G_p(r) \\
\langle DR\cdot DR\rangle &= N_D N_R (N_D-1)(N_R-1) G_p(r)^2 + N_D N_R (N_R-1) G_t(r) + \\
&\quad + N_D N_R (N_D-1)\frac{1+\xi(r)}{1+\xi_V} G_t(r) + N_D N_R G_p(r) \qquad (8.44)\\
\langle DD\cdot DR\rangle &= \tfrac{1}{2} N_D (N_D-1)(N_D-2) N_R G_p(r)^2 \frac{1+\xi(r)}{1+\xi_V} + N_R N_D (N_D-1) G_t(r) \frac{1+\xi(r)}{1+\xi_V}\,.
\end{aligned}$$

For the random catalog, we get $\langle RR\rangle$ and $\langle RR\cdot RR\rangle$ by just copying the derivations of $\langle DD\rangle$ and $\langle DD\cdot DD\rangle$, except now $\xi = 0$. However, since we made the approximation $\gamma = 0$, we have simply

$$RR \;=\; \langle RR\rangle \;=\; \frac{N_R(N_R-1)}{2} G_p(r)\,, \qquad (8.45)$$

so that we obtain the value of $G_p(r)$ from $RR(r)$. To find the value of $G_t(r)$ one has to write a separate piece of code to find all such triplets from the random catalogue that satisfy the r condition. See Fig. 29.

Calculation of $\langle DR\cdot DR\rangle$ and $\langle DD\cdot DR\rangle$. Consider $\langle DR\rangle$ and $\langle DR\cdot DR\rangle$, which contain data-random pairs. We have to decide whether we allow a data point and a random point to share the same microcell. Let's decide we don't, i.e., the microcells are small enough, i.e., K is large enough. This means that K is just huge compared to N_D or N_R, so there is really no difference between K, $K-1$, etc. So let's just write K always. Landy&Szalay [39] seem to work here in the $N_R \to \infty$ approximation which I

do not like to make, so I keep, e.g., the distinction between N_R and $N_R - 1$. For $\langle DR\rangle$ I get

$$\langle DR\rangle \;=\; \sum_{i\neq j}^{K}\langle n_i r_j\rangle\Theta_{ij}(r) \;=\; \frac{N_D N_R}{K^2}\sum_{i\neq j}^{K}\Theta_{ij}(r) \;=\; N_D N_R G_p(r)\,, \tag{8.46}$$

where $r_j = 0$ or 1 is the number of random points in microcell j and each microcell pair is included in the sum twice, since we have to check for both possibilities, either the first cell contains a data point and the second cell a random point, or vice versa.

For $\langle DR\cdot DR\rangle$ we follow the calculation of $\langle DD\cdot DD\rangle$, where the sum is divided into three parts. Now the microcell triplets occur 24 times instead of 6; we have 6 cases were the cell i shared by the pairs contains a data point, 6 cases with a random point, and 12 cases where it would have to contain both, for which we assigned zero probability, so these terms vanish. I get

$$\begin{aligned}\langle DR\cdot DR\rangle \;&=\; N_D N_R(N_D-1)(N_R-1)G_p(r)^2 + N_D N_R(N_R-1)G_t(r) + \\ &\quad + N_D N_R(N_D-1)\frac{1+\xi(r)}{1+\xi_V}G_t(r) + N_D N_R G_p(r)\,.\end{aligned} \tag{8.47}$$

This becomes (8.50b) in the limit of large N_R and negligible correlations; the last term can be dropped when $N_R \gg G_p/G_t$.

In $\langle DD\cdot DR\rangle$ we get again a 3-point correlation but with a different condition on the separations. I get

$$\langle DD\cdot DR\rangle \;=\; \tfrac{1}{2}N_D(N_D-1)(N_D-2)N_R G_p(r)^2\frac{1+\xi(r)}{1+\xi_V} + N_R N_D(N_D-1)G_t(r)\frac{1+\xi(r)}{1+\xi_V}\,. \tag{8.48}$$

8.1.2 The small-correlation limit

In (8.42) and (8.44) the correlation function ξ appears only as $1+n\xi$. Since we have already made approximations based on ξ being small, we may approximate these with 1 for calculating the variance, since it is not needed as accurately as the bias. (We do not make this approximation in $\langle DD\rangle$, which is needed for the bias.) With this approximation and in the limit of $N_R \to \infty$, (8.42) and (8.44) become

$$\begin{aligned}\langle DD\cdot DD\rangle \;&=\; \frac{N_D(N_D-1)(N_D-2)(N_D-3)}{4}G_p(r)^2 + N_D(N_D-1)(N_D-2)G_t(r) + \frac{N_D(N_D-1)}{2}G_p(r)\\ \langle DR\rangle \;&=\; N_D N_R G_p(r)\\ \langle DR\cdot DR\rangle \;&=\; N_D N_R^2\left[G_p^2(r)(N_D-1) + G_t(r)\right]\\ \langle DD\cdot DR\rangle \;&=\; \tfrac{1}{2}N_D(N_D-1)N_R\left[(N_D-2)G_p(r)^2 + 2G_t(r)\right]\end{aligned} \tag{8.49}$$

Then we have

$$\begin{aligned}\langle\alpha^2\rangle \;&=\; \frac{2}{N_D(N_D-1)}\left\{2(N_D-2)\left[\frac{G_t(r)}{G_p(r)^2}-1\right] + \frac{1}{G_p(r)} - 1\right\}\\ \langle\beta^2\rangle \;&=\; \frac{1}{N_D}\left[\frac{G_t(r)}{G_p(r)^2}-1\right] \;\equiv\; t\\ \langle\alpha\beta\rangle \;&=\; \frac{2}{N_D}\left[\frac{G_t(r)}{G_p(r)^2}-1\right] \;=\; 2\langle\beta^2\rangle \;=\; 2t\,.\end{aligned} \tag{8.50}$$

Here t is a small quantity that arises from the difference between G_t and G_p^2. Since it is equal to $\langle\beta^2\rangle$, it has to be positive, so that $G_t > G_p^2$. This difference arises from edge effects.

8.1.3 Bias

We define normalized pair counts

$$\begin{aligned} d(r) &\equiv \frac{DD(r)}{\frac{1}{2}N_D(N_D-1)G_p(r)} \quad \Rightarrow \quad \langle d\rangle = \frac{1+\xi(r)}{1+\xi_V} \\ x(r) &\equiv \frac{DR(r)}{N_D N_R G_p(r)} \quad \Rightarrow \quad \langle x\rangle = 1 \end{aligned} \tag{8.51}$$

so that we can write the four estimators as

$$\begin{aligned} 1+\widehat{\xi}_1(r) &\equiv d(r) \\ 1+\widehat{\xi}_{\rm DP}(r) &\equiv d(r)/x(r) \\ 1+\widehat{\xi}_{\rm H}(r) &\equiv d(r)/x(r)^2 \\ 1+\widehat{\xi}_{\rm LS}(r) &\equiv d(r)-2x(r)+2\,. \end{aligned} \tag{8.52}$$

We immediately se that the first and last estimators are unbiased, except for the "integral constraint" factor $1+\xi_V$,

$$1+\langle\widehat{\xi}_1\rangle = 1+\langle\widehat{\xi}_{\rm LS}\rangle = \langle d\rangle = \frac{1+\xi(r)}{1+\xi_V}\,. \tag{8.53}$$

The ratios in the second and third estimators create a bias. In the $\xi\to 0$ limit (8.17a) becomes

$$1+\langle\widehat{\xi}_{\rm DP}\rangle = \langle d(r)/x(r)\rangle = \langle d\rangle\left(1-\langle\alpha\beta\rangle+\langle\beta^2\rangle\right) = \langle d\rangle(1-t) \tag{8.54}$$

and a similar calculation (**exercise**) gives

$$1+\langle\widehat{\xi}_{\rm H}\rangle = \left\langle d(r)/x(r)^2\right\rangle = \langle d\rangle\left(1-2\langle\alpha\beta\rangle+3\langle\beta^2\rangle\right) = \langle d\rangle(1-t) \tag{8.55}$$

8.1.4 Variance in the small-correlation limit

In the limit of small correlations

$$\begin{aligned} \langle\alpha^2\rangle &\approx \frac{2}{N_D(N_D-1)}\left\{2(N_D-2)\left[\frac{G_t(r)}{G_p(r)^2}-1\right]+\frac{1}{G_p(r)}-1\right\} \\ &= 4t+p\,, \end{aligned} \tag{8.56}$$

where

$$p \equiv \frac{2}{N_D(N_D-1)}\left[\frac{1}{G_p}-2\frac{G_t}{G_p^2}+1\right] \approx \frac{2}{N_D(N_D-1)G_p(r)} \tag{8.57}$$

(the other terms inside the brackets are of order 1, whereas $1/G_p$ is large). The approximate form for p is the inverse of the expected number of pairs in the bin (in the case of no correlations); which is the relative variance for a Poisson distribution.

Calculate now the variances

$$\mathrm{var}(\widehat{\xi}) \equiv \left\langle\left(\widehat{\xi}-\langle\widehat{\xi}\rangle\right)^2\right\rangle = \langle\widehat{\xi}^2\rangle-\langle\widehat{\xi}\rangle^2 \tag{8.58}$$

for the four estimators. We use from the above

$$d=\langle d\rangle(1+\alpha)\,,\quad x=\langle x\rangle(1+\beta)\,,\quad \langle\alpha^2\rangle=4t+p\,,\quad \langle\beta^2\rangle=t\,,\quad \langle\alpha\beta\rangle=2t\,. \tag{8.59}$$

For the first estimator

$$\begin{aligned}
\widehat{\xi}_1 &= d - 1 = \langle d\rangle(1+\alpha) - 1 \quad \Rightarrow \quad \widehat{\xi}_1^2 = \langle d\rangle^2(1+\alpha)^2 - 2\langle d\rangle(1+\alpha) + 1 \\
\langle\widehat{\xi}_1^2\rangle &= \langle d\rangle^2(1+\langle\alpha^2\rangle) - 2\langle d\rangle + 1 \\
\langle\widehat{\xi}_1\rangle^2 &= \langle d\rangle^2 - 2\langle d\rangle + 1 \\
\mathrm{var}(\widehat{\xi}_1) &= \langle d\rangle^2\langle\alpha^2\rangle \approx \langle d\rangle^2(4t+p)\,.
\end{aligned} \tag{8.60}$$

and for the Davis-Peebles estimator

$$\begin{aligned}
\widehat{\xi}_{\rm DP} &= \frac{d}{x} - 1 = \langle d\rangle\frac{1+\alpha}{1+\beta} - 1 \approx \langle d\rangle(1+\alpha-\beta-\beta\alpha+\beta^2) - 1 \\
\langle\widehat{\xi}_{\rm DP}\rangle &\approx \langle d\rangle(1-\langle\beta\alpha\rangle+\langle\beta^2\rangle) - 1 = \langle d\rangle(1-t) - 1 \\
\widehat{\xi}_{\rm DP}^2 &= \left(\frac{d}{x}\right)^2 - 2\frac{d}{x} + 1 = \langle d\rangle^2\frac{(1+\alpha)^2}{(1+\beta)^2} - 2\langle d\rangle\frac{1+\alpha}{1+\beta} + 1 \\
\langle\widehat{\xi}_{\rm DP}^2\rangle &= \langle d\rangle^2\left\langle\frac{(1+\alpha)^2}{(1+\beta)^2}\right\rangle - 2\langle d\rangle\left\langle\frac{1+\alpha}{1+\beta}\right\rangle + 1 \approx \langle d\rangle^2(1-t+p) - 2\langle d\rangle(1-t) + 1 \\
\langle\widehat{\xi}_{\rm DP}\rangle^2 &\approx \langle d\rangle^2(1-2t) - 2\langle d\rangle(1-t) + 1 \\
\mathrm{var}(\widehat{\xi}_{\rm DP}) &\approx \langle d\rangle^2(t+p)\,.
\end{aligned} \tag{8.61}$$

Similarly we obtain (**exercise**)

$$\mathrm{var}(\widehat{\xi}_{\rm H}) \approx \mathrm{var}(\widehat{\xi}_{\rm LS}) \approx \langle d\rangle^2 p\,, \tag{8.62}$$

showing that these two have a smaller variance. (I got $\mathrm{var}(\widehat{\xi}_{\rm LS}) \approx \langle d\rangle^2 p + 4t(\langle d\rangle - 1)^2$, but since we are assuming small correlations, $(\langle d\rangle - 1)^2$ is small and we can drop $t(\langle d\rangle - 1)^2$.) Since $\widehat{\xi}_{\rm LS}$ is also unbiased (except for the integral constraint), it is usually preferred.

Note that the result that $\widehat{\xi}_1$ and $\widehat{\xi}_{\rm LS}$ are unbiased, was obtained in the limit $N_R \to \infty$. A finite random catalog introduces a bias, which, however, is small if the random catalog is sufficiently large.

Note that these results for the variances were obtained in the limit of small $\xi(r)$. It is not easy to estimate the variance of the estimators when $\xi(r)$ is large. For this there are methods (*bootstrap* and *jackknife*) based on resampling the data itself. (We won't discuss these now.) In fact, the large $\xi(r)$ for small separations affects the variance even at those larger separations where $\xi(r)$ is small – for example, it hurts the approximation $\xi(r_{jk}) \approx \xi(r)$ before (8.40). Thus the above comparison of estimator variances is indicative only and the *true variances are larger*. The variance of the correlation function estimate is needed for the use of the estimated correlation function for the estimation of cosmological parameters. Even the data resampling methods mentioned above are unreliable and for serious work the variances are estimated by producing a large number of simulated data realizations and calculating the correlation function from all of them, and then calculating the variance over this set of realizations. For large galaxy surveys this is a major computational effort.

Note that the result that the Landy–Szalay estimator is unbiased in the $N_R \to \infty$ limit holds also for large $\xi(r)$; this argument for preferring it is thus on more solid foundation than the small variance.

Exercise: Pair volume fraction. Find $G_p(r)$ for a cubic volume $V = L^d$ in $d = 1, 2,$ and 3 dimensions, in the approximation where one ignores edge effects (good for $r \ll L$) and Δr is treated like dr (good for $\Delta r \ll r$). What is $r_{\rm max}$ (the maximum possible separation r) in these three cases? What do you get for $\sum_r G_p(r)$ when using these approximate formulae? At what value of r the sum already reaches 1?

Exercise: Triplet volume fraction. The difference between $G_t(r)$ and $G_p(r)^2$ is due to edge effects. Now take these properly into account (but you can still treat Δr like dr), and find $G_p(r)$ and $G_t(r)$ in the 1D case (survey volume is L). Show that $G_t(r) > G_p(r)^2$. Are there special values of r where they are equal?

Project: For large surveys it may become computationally expensive to have $N_R \gg N_D$. For example, for Euclid, $N_D \sim 10^9$. With the usually recommended $N_R \sim 10^2 N_D$, there would be $\sim 10^{22}$ random pairs. To study the effect of using a smaller N_R, redo this section without assuming $\gamma \approx 0$. This was done in [40].

---

9 Power spectrum estimation

For large scales, estimating $P(\mathbf{k})$ directly has a number of advantages compared to $\xi(\mathbf{x})$. Uncertainty about the true value of $\langle\rho\rangle$ is just a multiplicative factor for $P(\mathbf{k})$, whereas it may cause us to get even the sign of ξ wrong when ξ is smaller than the difference $(\bar{\rho}^2 - \langle\rho\rangle^2)/\langle\rho\rangle^2$. For large scales, where the perturbations are still in the linear regime, different Fourier modes $\delta_{\mathbf{k}}$ evolve independently, and thus the low-$\mathbf{k}$ modes are not contaminated by the small-scale non-linear behavior[48]; whereas $\xi(r)$ for large r contains a mixture of linear and nonlinear modes.

9.1 Shot noise

We saw in Sec. 2 that for a Poisson distribution $P(\mathbf{k}) = const = 1/\langle\rho\rangle$ and $\xi(\mathbf{r}) = 0$ for $r \neq 0$, but $\xi(0)$ has a nonzero value that increases with resolution, i.e., $\xi(\mathbf{r})$ is a delta function. (Constant and delta function form a Fourier transform pair.) We take the view that the galaxies sample an underlying continuous density field, and we want the $P(\mathbf{k})$ and $\xi(\mathbf{r})$ of that density field. This constant contribution to $P(\mathbf{k})$ and the associated delta-function spike in $\xi(\mathbf{r})$ are the result of this discrete sampling, called *shot noise* (or Poisson noise), and not part of the underlying density field. Thus we do not want to include it in the estimated $P(\mathbf{k})$. This is consistent with the methods to estimate $\xi(r)$ discussed in Sec. 8, since there we estimated ξ by comparing it to a Poisson distribution, so that the expected result for ξ was 0 in the case of Poisson distribution. Thus for a Poisson distribution $P(\mathbf{k}) = \xi(\mathbf{r}) = 0$ now.

Let us redo the power spectrum analysis for a discrete set of objects, done in Sec. 2.5.1 for a Poisson distribution, for a case where there are correlations. The Fourier coefficients of density and its perturbation are related simply

$$\delta_{\mathbf{k}} \;=\; \frac{1}{\langle\rho\rangle}\rho_{\mathbf{k}}\,, \qquad \text{except for } \mathbf{k} = 0. \tag{9.1}$$

Thus

$$P(\mathbf{k}) \;\equiv\; V\langle|\delta_{\mathbf{k}}|^2\rangle \;=\; \frac{V}{\langle\rho\rangle^2}\langle|\rho_{\mathbf{k}}|^2\rangle\,, \tag{9.2}$$

except for $k = 0$, and we are not asking for $P(k = 0)$ here. Thus we shall work with ρ instead of δ as it is conceptually easier.

We again divide the reference volume V into microcells of volume δV. The density of microcell j is

$$\rho(\mathbf{x}_j) \;=\; \frac{n_j}{\delta V}\,. \tag{9.3}$$

where the occupation number is $n_j = 0$ or 1 and its expectation value is $\langle n_j\rangle = \langle\rho\rangle\delta V \ll 1$. For $\rho_{\mathbf{k}}$ and its expectation value we get

$$\begin{aligned}
\rho_{\mathbf{k}} \;&=\; \frac{1}{V}\int_V \rho(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}d^dx \;=\; \frac{1}{V}\sum_j \frac{n_j}{\delta V}e^{-i\mathbf{k}\cdot\mathbf{x}_j}\delta V \;=\; \frac{1}{V}\sum_j n_j e^{-i\mathbf{k}\cdot\mathbf{x}_j} \qquad (9.4)\\
\langle\rho_{\mathbf{k}}\rangle \;&=\; \frac{1}{V}\sum_j \langle n_j\rangle e^{-i\mathbf{k}\cdot\mathbf{x}_j} \;=\; \frac{\langle\rho\rangle}{V}\int_V e^{-i\mathbf{k}\cdot\mathbf{x}}d^dx \;=\; 0 \qquad \text{for } \mathbf{k} \neq 0\,.
\end{aligned}$$

For their correlations we get

$$\langle\rho^*_{\mathbf{k}}\rho_{\mathbf{k}'}\rangle \;=\; \frac{1}{V^2}\sum_{ij}\langle n_i n_j\rangle e^{i\mathbf{k}\cdot\mathbf{x}_i}e^{-i\mathbf{k}'\cdot\mathbf{x}_j} \tag{9.5}$$

[48] To be exact, the Fourier modes are strictly independent only when all scales are in the linear regime; the question of how much nonlinear small scales may affect the large scales in the linear regime is the difficult *backreaction* question, not fully solved; but it is reasonable to expect these effects are small.

Now the first key step here is to separate from the sum over microcell pairs the part where $i = j$. One might think that this makes a negligible contribution since the number of microcells is semi-infinite. However, this is precisely where we have the δ-function spike of ξ, and therefore the contribution is not negligible. But because the volume of the $i = j$ part is negligible in pair space, we can then still treat the $i \neq j$ part of the sum as a full double integral over the volume, we have just taken out the δ-spike of ξ. Thus[49]

$$\begin{aligned}
\langle \rho^*_{\mathbf{k}} \rho_{\mathbf{k}'} \rangle &= \frac{1}{V^2} \sum_j \langle n_j \rangle e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}_j} + \frac{1}{V^2} \sum_{i\neq j} \langle n_i n_j \rangle e^{i\mathbf{k}\cdot\mathbf{x}_i} e^{-i\mathbf{k}'\cdot\mathbf{x}_j} \\
&= \frac{1}{V^2} \sum_j \langle \rho \rangle \delta V e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}_j} + \frac{1}{V^2} \sum_{i\neq j} \langle \rho \rangle^2 \delta V^2 \left[1 + \xi(\mathbf{x}_j - \mathbf{x}_i)\right] e^{i\mathbf{k}\cdot\mathbf{x}_i} e^{-i\mathbf{k}'\cdot\mathbf{x}_j} \\
&= \frac{\langle \rho \rangle}{V^2} \int_V e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} d^d x + \frac{\langle \rho \rangle^2}{V^2} \int_V d^d x e^{i\mathbf{k}\cdot\mathbf{x}} \int_V d^d x' e^{-i\mathbf{k}'\cdot\mathbf{x}'} \\
&\quad + \frac{\langle \rho \rangle^2}{V^2} \int_V d^d x d^d x' \xi(\mathbf{x}' - \mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}} e^{-i\mathbf{k}'\cdot\mathbf{x}'} \\
&= \frac{\langle \rho \rangle}{V} \delta_{\mathbf{k}\mathbf{k}'} + 0 + \langle \rho \rangle^2 \frac{1}{V} \delta_{\mathbf{k}\mathbf{k}'} P(\mathbf{k}) \,, \qquad (9.6)
\end{aligned}$$

where we used statistical homogeneity to get $P(\mathbf{k})$ from $\xi(\mathbf{x}' - \mathbf{x})$ just like in Eq. (1.74).

We can write the result as

$$\tilde{P}(\mathbf{k}) \equiv \frac{V}{\langle \rho \rangle^2} \langle |\rho_{\mathbf{k}}|^2 \rangle = \frac{1}{\langle \rho \rangle} + P(\mathbf{k}) \,, \qquad (9.7)$$

where $\tilde{P}(\mathbf{k})$ is the power spectrum of the point (galaxy) distribution, $1/\langle \rho \rangle$ is the shot noise, and

$$P(\mathbf{k}) = \frac{V}{\langle \rho \rangle^2} \langle |\rho_{\mathbf{k}}|^2 \rangle - \frac{1}{\langle \rho \rangle} \,, \qquad \text{where} \qquad \rho_{\mathbf{k}} = \frac{1}{V} \sum_j n_j e^{-i\mathbf{k}\cdot\mathbf{x}_j} \qquad (9.8)$$

is the power spectrum we want.

To turn this into a power spectrum estimator we replace the expectation values $\langle n_j \rangle$ with the actual observed numbers $n_j = 0$ or 1, which changes the sum over microcells into a sum over galaxies, and $\langle \rho \rangle$ with $\bar{\rho} = N/V$,

$$\begin{aligned}
\widehat{P}(\mathbf{k}) &= V \left(\frac{V^2}{N^2} |\rho_{\mathbf{k}}|^2 - \frac{1}{N} \right) \,, \qquad \text{where} \qquad \rho_{\mathbf{k}} = \frac{1}{V} \sum_i e^{-i\mathbf{k}\cdot\mathbf{x}_i} \\
&= V \left| \frac{1}{N} \sum_i e^{-i\mathbf{k}\cdot\mathbf{x}_i} \right|^2 - \frac{V}{N} \,. \qquad (9.9)
\end{aligned}$$

In this section I use the convention that sums over j are over microcells, and single sums over i are over galaxies – double sums over ij are over microcells.

In practice one groups the $\mathbf{k}$ values into bins and averages over the $\widehat{P}(\mathbf{k})$ obtained from (9.9) to get the $\widehat{P}(\mathbf{k})$ or $\widehat{P}(k)$ for the bin.

[49] In hindsight, it was unnecessary to include the $\mathbf{k} \neq \mathbf{k}'$ case here; with $\mathbf{k} = \mathbf{k}'$ (9.6) becomes

$$\langle |\rho_{\mathbf{k}}|^2 \rangle = \frac{\langle \rho \rangle}{V} + 0 + \frac{\langle \rho \rangle^2}{V^2} \int_V d^d x d^d x' \xi(\mathbf{x}' - \mathbf{x}) e^{-i\mathbf{k}\cdot(\mathbf{x}'-\mathbf{x})} \,,$$

where the last term is $(\langle \rho \rangle^2 / V) P(\mathbf{k})$. We don't even have to assume statistical homogeneity if we take

$$P(\mathbf{k}) = \frac{1}{V} \int_V d^d x d^d x' \xi(\mathbf{x}' - \mathbf{x}) e^{-i\mathbf{k}\cdot(\mathbf{x}'-\mathbf{x})}$$

as the definition of the power spectrum.

9.2 Mask

In reality, the survey volume is not likely to be a perfect cube. Let thus V be a large cubic volume that contains the survey volume $V_S < V$; we assume periodicity in V for Fourier purposes. We account for the survey volume by introducing a mask

$$W(\mathbf{x}) = 1 \quad \text{if } \mathbf{x} \in V_S, \text{ 0 otherwise.} \quad \Rightarrow \quad \int_V W(\mathbf{x}) d^d x = V_S \,. \tag{9.10}$$

We can also use the mask to assign different weights to different parts of the survey (reasons for this will be discussed later), so that

$$0 \le W(\mathbf{x}) \le 1 \,; \tag{9.11}$$

regions with smaller W will be weighted less. We can still define V_S according to (9.10); this weighted volume will then be smaller than the actual volume. As will be seen below, we have to require that W changes slowly, so that the typical wavelengths of the shape of W are much larger than the wavelengths of the Fourier modes $\mathbf{k}$ we try to estimate $P(\mathbf{k})$ for. The weights can be used also to smooth the edge of the survey, so that W does not fall sharply from 1 to 0 there.

The mask multiplies the density field, $\rho(\mathbf{x}) \to \rho^w(\mathbf{x}) \equiv W(\mathbf{x})\rho(\mathbf{x})$, which I call here the weighted density. This has a number of effects: It breaks the statistical homogeneity; true density perturbations will now correspond to $W(\mathbf{x})\rho(\mathbf{x})$ differing from $W(\mathbf{x})\langle\rho\rangle$, which is already inhomogeneous by itself; and the density perturbation obtained this way is the true density perturbation multiplied by the mask, which becomes convolution in Fourier space.

We define the weighted number of galaxies in a microcell as $n_j^w \equiv W(\mathbf{x}_j) n_j = W_j n_j$, where $n_j = 0$ or 1 is the true number of galaxies in the microcell, so that $0 \le n_j^w \le 1$. Now $n_j^2 = n_j$, but $(n_j^w)^2 \ne n_j^w$ in general for a weighted mask. The weighted density of a microcell is

$$\rho^w(\mathbf{x}_j) \;\equiv\; \frac{n_j^w}{\delta V} \;\equiv\; \frac{W(\mathbf{x}_j) n_j}{\delta V} = \frac{W_j n_j}{\delta V} \tag{9.12}$$

For the Fourier components of the weighted density we have

$$\rho_{\mathbf{k}}^w \;=\; \frac{1}{V} \sum_j n_j W_j e^{-i\mathbf{k}\cdot\mathbf{x}_j} \tag{9.13}$$

$$\langle \rho_{\mathbf{k}}^w \rangle \;=\; \frac{1}{V} \sum_j \langle n_j \rangle W_j e^{-i\mathbf{k}\cdot\mathbf{x}_j} \;=\; \frac{\langle\rho\rangle}{V} \int_V W(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d^d x \;=\; \langle\rho\rangle W_{\mathbf{k}} \;\ne\; 0 \,.$$

We now proceed like in (9.6). Since the mask has broken statistical homogeneity, different $\rho_{\mathbf{k}}^w$ will now be correlated. For the power spectrum we need only $\langle |\rho_{\mathbf{k}}^w|^2 \rangle$, so we put $\mathbf{k}' = \mathbf{k}$ from

the start:

$$\begin{aligned}
\langle \rho^{w*}_{\mathbf{k}} \rho^{w}_{\mathbf{k}} \rangle &= \frac{1}{V^2}\sum_j \langle n_j \rangle W_j^2 + \frac{1}{V^2}\sum_{i\neq j}\langle n_i n_j\rangle W_i W_j e^{i\mathbf{k}\cdot(\mathbf{x}_i-\mathbf{x}_j)} \\
&= \frac{1}{V^2}\sum_j \langle\rho\rangle W_j^2 \delta V + \frac{1}{V^2}\sum_{i\neq j}\langle\rho\rangle^2 \delta V^2 \left[1+\xi(\mathbf{x}_j-\mathbf{x}_i)\right] W_i W_j e^{i\mathbf{k}\cdot(\mathbf{x}_i-\mathbf{x}_j)} \\
&= \frac{\langle\rho\rangle}{V}\frac{1}{V}\int_V W(\mathbf{x})^2 d^dx + \langle\rho\rangle^2 \left|\frac{1}{V}\int_V d^dx W(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}\right|^2 \\
&\quad + \frac{\langle\rho\rangle^2}{V^2}\int_V d^dx d^dx' W(\mathbf{x}) W(\mathbf{x}') \xi(\mathbf{x}'-\mathbf{x}) e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} \\
&= \frac{\langle\rho\rangle}{V}\sum_{\mathbf{q}} |W_{\mathbf{q}}|^2 + \langle\rho\rangle^2 |W_{\mathbf{k}}|^2 \qquad (9.14) \\
&\quad + \langle\rho\rangle^2 \frac{1}{V}\sum_{\mathbf{k}'} P(\mathbf{k}')\frac{1}{V}\int_V d^dx W(\mathbf{x}) e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} \frac{1}{V}\int_V d^dx' W(\mathbf{x}') e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}'} ,
\end{aligned}$$

where we used the Parseval formula

$$\frac{1}{V}\int_V W(\mathbf{x})^2 d^dx \;=\; \sum_{\mathbf{q}} |W_{\mathbf{q}}|^2 \qquad (9.15)$$

for the first term and

$$\xi(\mathbf{r}) \;=\; \frac{1}{V}\sum_{\mathbf{k}'} e^{i\mathbf{k}'\cdot\mathbf{r}} P(\mathbf{k}') \qquad (9.16)$$

for the last term.[50] We have now

$$\langle \rho^{w*}_{\mathbf{k}} \rho^{w}_{\mathbf{k}} \rangle - \langle\rho\rangle^2 |W_{\mathbf{k}}|^2 \;=\; \frac{\langle\rho\rangle}{V}\sum_{\mathbf{q}} |W_{\mathbf{q}}|^2 + \frac{\langle\rho\rangle^2}{V}\sum_{\mathbf{k}'} |W_{\mathbf{k}-\mathbf{k}'}|^2 P(\mathbf{k}') . \qquad (9.17)$$

Without the mask, the second term of (9.14), which we now moved to the lhs, was zero. From (9.13) we see that it is $|\langle\rho^w_{\mathbf{k}}\rangle|^2$. Thus we can write the lhs of (9.17) as $\langle|\rho^w_{\mathbf{k}} - \langle\rho^w_{\mathbf{k}}\rangle|^2\rangle$, i.e. we are comparing $W\rho$ to $W\langle\rho\rangle$ to find the density perturbation.

The last term in (9.17) is a convolution of the power spectrum with $|W_{\mathbf{k}}|^2$. To solve $P(\mathbf{k})$ we would have to deconvolve the equation. We do this in an approximate manner, which is based on the assumption that $W(\mathbf{x})$ is smooth at the scales we want to estimate $P(\mathbf{k})$ and that $P(\mathbf{k})$ does not change fast as a function of $\mathbf{k}$. (In the extreme case that $W(\mathbf{x}) = const$, its Fourier transform is a delta function times a constant, and convolving with a delta function is equivalent with multiplying by this constant.[51]) If instead $W(\mathbf{x})$ is just smooth, $W(\mathbf{k})$ will be a narrow peak around $\mathbf{k} = 0$, i.e., in the sum

$$\sum_{\mathbf{k}'} |W_{\mathbf{k}-\mathbf{k}'}|^2 P(\mathbf{k}') \;=\; \sum_{\mathbf{q}} |W_{\mathbf{q}}|^2 P(\mathbf{k}+\mathbf{q}) \qquad (\mathbf{q}\equiv\mathbf{k}'-\mathbf{k} \text{ and } W_{-\mathbf{q}} = W^*_{\mathbf{q}}) , \qquad (9.18)$$

only terms where $|\mathbf{q}| = |\mathbf{k}-\mathbf{k}'| \ll |\mathbf{k}|$ are significant (where $\mathbf{k}$ represents the scale we are interest in). We assume that $P(\mathbf{k}+\mathbf{q})$ does not change much over this range of $\mathbf{q}$ and approximate the sum with

$$P(\mathbf{k})\sum_{\mathbf{q}} |W_{\mathbf{q}}|^2 , \qquad (9.19)$$

[50] The reason I write $P(\mathbf{k})$, not $P_{\mathbf{k}}$, here is that in my conventions $P_{\mathbf{k}} = P(\mathbf{k})/V$ and I want to keep focus on the quantity $P(\mathbf{k})$, which we are trying to estimate and which is independent of the reference volume V.

[51] The reason I didn't write this in equations is that I didn't bother to think about how to deal with the square.

turning the convolution into multiplication, and thus deconvolution into division.

Now we can solve $P(\mathbf{k})$ from (9.17),

$$P(\mathbf{k}) \;\approx\; \frac{V\langle|\rho^w_{\mathbf{k}} - \langle\rho^w_{\mathbf{k}}\rangle|^2\rangle - \langle\rho\rangle\sum_{\mathbf{q}}|W_{\mathbf{q}}|^2}{\langle\rho\rangle^2\sum_{\mathbf{q}}|W_{\mathbf{q}}|^2} \;\equiv\; \frac{V^2\langle|\rho^w_{\mathbf{k}} - \langle\rho^w_{\mathbf{k}}\rangle|^2\rangle - N_{\rm eff}}{\langle\rho\rangle^2 V\sum_{\mathbf{q}}|W_{\mathbf{q}}|^2}\,, \tag{9.20}$$

where

$$N_{\rm eff} \;=\; \langle\rho\rangle V\sum_{\mathbf{q}}|W_{\mathbf{q}}|^2\,. \tag{9.21}$$

To turn this into an estimator, we write $\rho^w_{\mathbf{k}}$ (9.13) and $\langle\rho\rangle$ in terms of n_j and then replace expectation values with observed values and sums over microcells with sums over observed galaxies. We have

$$\sum\langle n_j\rangle W_j = \langle\rho\rangle\sum W_j\delta V \quad\Rightarrow\quad \langle\rho\rangle = \frac{\sum\langle n_j\rangle W_j}{\sum W_j\delta V}\,, \tag{9.22}$$

which we replace with[52]

$$\bar\rho \equiv \frac{\sum n_j W_j}{\sum W_j\delta V} \;=\; \frac{\sum W_i}{V_S} \equiv \frac{N_S}{V_S}\,, \tag{9.23}$$

where we defined

$$N_S \;\equiv\; \sum_i W_i\,; \tag{9.24}$$

and

$$\rho^w_{\mathbf{k}} - \langle\rho^w_{\mathbf{k}}\rangle \;\to\; \frac{1}{V}\sum_j n_j W_j e^{-i\mathbf{k}\cdot\mathbf{x}_j} - \frac{N_S}{V_S}W_{\mathbf{k}} \;=\; \frac{1}{V}\sum_i W_i e^{-i\mathbf{k}\cdot\mathbf{x}_i} - \frac{N_S}{V_S}W_{\mathbf{k}}\,. \tag{9.25}$$

Thus we have the estimator

$$\begin{aligned}\widehat{P}(\mathbf{k}) &= V\frac{V_S^2}{N_S^2\sum_{\mathbf{q}}|W_{\mathbf{q}}|^2}\left(\left|\frac{1}{V}\sum_i W_i e^{-i\mathbf{k}\cdot\mathbf{x}_i} - \frac{N_S}{V_S}W_{\mathbf{k}}\right|^2 - \frac{1}{V}\frac{N_S}{V_S}\sum_{\mathbf{q}}|W_{\mathbf{q}}|^2\right)\\ &= \frac{V}{\sum_{\mathbf{q}}|W_{\mathbf{q}}|^2}\left|\frac{V_S}{V}\frac{1}{N_S}\sum_i W_i e^{-i\mathbf{k}\cdot\mathbf{x}_i} - W_{\mathbf{k}}\right|^2 - \frac{V_S}{N_S}\,.\end{aligned} \tag{9.26}$$

Here the V_S/N_S term subtracts the shot noise.

Exercise: Show that in the case of an unweighted mask (i.e. $W(\mathbf{x}) = 0$ or 1), (9.20) becomes

$$P(\mathbf{k}) \;\approx\; \frac{V^2\langle|\rho^w_{\mathbf{k}} - \langle\rho^w_{\mathbf{k}}\rangle|^2\rangle - \langle\rho\rangle V_S}{\langle\rho\rangle^2 V_S}\,. \tag{9.27}$$

---

[52]Note that $\bar\rho$ defined this way is not the ratio of the actual (unweighted) number of observed galaxies to the actual (unweighted) survey volume, since in N_S the weighting goes by galaxy and in V_S by microcell. MBW do not define these weighted quantities N_S and V_S (their V_S stands for the actual survey volume). I was not able to get exactly the same form for $N_{\rm eff}$ as they give.

9.3 Selection function

In the preceding we assumed $\langle n_j\rangle = \langle\rho\rangle\delta V$. Because of selection effects this will not be true for observed galaxies. Let $n_j = 0$ or 1 now be the number of observed galaxies in a microcell. Then

$$\langle n_j\rangle \;=\; S(\mathbf{x}_j)\langle\rho\rangle\delta V \quad\Rightarrow\quad \left\langle\frac{n_j}{S_j}\right\rangle \;=\; \langle\rho\rangle\delta V\,, \tag{9.28}$$

where $S(\mathbf{x})$ is the selection function. To correct for this we let each observed galaxy i represent $1/S(\mathbf{x}_i)$ galaxies assumed to exist, and define a corrected number of galaxies in a microcell

$$n_j^c \;\equiv\; A\frac{W(\mathbf{x}_j)}{S(\mathbf{x}_j)}n_j \;=\; A\frac{W_j}{S_j}n_j\,, \tag{9.29}$$

where A is an additional normalization factor.[53] It may appear that we just replaced the W_j of Sec. 9.2 with W_j/S_j; but they behave differently since S_j will be absorbed by expectation values according to (9.28).

We get now

$$\begin{aligned}\rho_{\mathbf{k}}^c &= \frac{A}{V}\sum_j n_j\frac{W_j}{S_j}e^{-i\mathbf{k}\cdot\mathbf{x}_j} \qquad (9.30)\\ \langle\rho_{\mathbf{k}}^c\rangle &= \frac{1}{V}\sum_j\left\langle\frac{n_j}{S_j}\right\rangle W_j e^{-i\mathbf{k}\cdot\mathbf{x}_j} \;=\; \frac{A\langle\rho\rangle}{V}\int_V W(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}d^dx \;=\; A\langle\rho\rangle W_{\mathbf{k}}\,,\end{aligned}$$

and arrive at[54] (**exercise**)

$$P(\mathbf{k}) \;\approx\; \frac{V^2\langle|\rho_{\mathbf{k}}^c - \langle\rho_{\mathbf{k}}^c\rangle|^2\rangle - A^2\langle\rho\rangle\int_V\frac{W(\mathbf{x})^2}{S(\mathbf{x})}d^dx}{\langle\rho\rangle^2A^2V\sum_{\mathbf{q}}|W_{\mathbf{q}}|^2}\,. \tag{9.31}$$

9.4 Weights

In a redshift survey we observe fewer galaxies per unit volume at larger distances (redshifts), so that the more distant volume elements are poorly sampled and their contribution to the estimate has larger uncertainty. When we divide by the selection function to compensate for the low observed density, this puts more weight to an individual observed galaxy, so that all volume elements will be weighted equally if we do not assign weights. We may want to account for the larger uncertainty by assigning weights $W(\mathbf{x}) = S(\mathbf{x})$, so that instead each observed galaxy is weighted equally ($W/S = 1$) in the estimate. However, this is not optimal either, since now we assign most of the weight to a small nearby region of the survey, which is not a good sample of the universe; a smaller volume has a larger cosmic variance. Feldman et al. [41] showed that an optimal compromise (to minimize the expected error in the $P(\mathbf{k})$) is to assign weights[55]

$$W(\mathbf{x}) \;=\; \frac{S(\mathbf{x})}{1+\bar n(\mathbf{x})P(k)} \;=\; \frac{S(\mathbf{x})}{1+S(\mathbf{x})\langle\rho\rangle P(k)}\,, \tag{9.32}$$

(called the FKP weights) where $\bar n \equiv S(\mathbf{x})\langle\rho\rangle$ is the expected observed number density at $\mathbf{x}$. For large distances, where $S \ll 1$, we get $W = S \ll 1$ and each observed galaxy gets equal weight in

[53] We can choose $A = \frac{N}{\sum_j n_j W_j S_j^{-1}}$, to make $\sum n_j^c = N$, the actual number of galaxies in the survey.

[54] MBW include the normalization factor A in $W_{\mathbf{q}}$, i.e., $\tilde W_{\mathbf{q}} = AW_{\mathbf{q}}$, so that A disappears from the denominator.

[55] Feldman et al. [41] and Peacock[1] write just $w(\mathbf{x}) = 1/[1+\bar n(\mathbf{x})P(k)]$, but I think their w must mean our W/S; otherwise this does not do what it says.

the estimate. For more nearby regions where $S \approx 1$, the weights become independent of location and each volume element is weighted equally.

Note that the weights are different for estimating $P(\mathbf{k})$ for different k. Since we are in the process of estimating $P(k)$ the task appears circular; but it is not crucial to use exactly optimal weights. We can use prior knowledge of $P(k)$ to set the weights, and if desired, we can iterate by repeating the estimation using the estimated $P(k)$ from the first round.

We did not discuss weights and selection function in the context of ξ estimation. In that case their application is straightforward: we just apply the selection function to the random catalog and weight both catalogs the same way. In the ξ case the procedure that corresponds to the FKP weights for $P(k)$ is to weight each pair by W^2, where

$$W(r;z) \;=\; \frac{1}{1+4\pi\bar{n}(z)J_3(r)} \;=\; \frac{1}{1+\bar{n}(z)V(r)\bar{\xi}(r)}\,, \tag{9.33}$$

where r is the separation of the pair and z is their mean redshift [41, 2].

10 Baryon acoustic oscillation scale as a standard ruler

The acoustic oscillation of the baryon-photon fluid before the photon decoupling imprints on the matter distribution a particular distance scale, the sound horizon at photon decoupling time

$$r_d \;=\; \int_0^{t_d} \frac{c_s(t)}{a(t)} dt \;=\; \int_{z_d}^{\infty} \frac{c_s(z)}{H(z)} dz\,. \tag{10.1}$$

This scale can be seen prominently in the CMB anisotropy: it corresponds to the separation of the acoustic peaks in the CMB angular power spectrum; from which we can measure directly at what angle (the *sound horizon angle* ϑ_s) we see this scale at the last scattering surface. To convert the observed angle to distance scale we need to multiply it by the angular diameter distance, which depends on comsological parameters. However, the theoretical value (10.1) is determined by early universe physics, and in standard cosmology it depends only on the cosmological parameters ω_m and ω_b, which can be determined accurately from the observed heights of the CMB acoustic peaks. Therefore r_d is known very accurately. To be precise, the distance scale $r_* = 144.57 \pm 0.22\,\mathrm{Mpc}$ seen in CMB is slightly different from the baryon acoustic oscillation (BAO) scale $r_d = 147.21 \pm 0.23\,\mathrm{Mpc}$ seen in the matter distribution. This is because the decoupling was not instantaneous and there are many more photons than baryons. In CMB we see the situation when photons on average interacted with baryons the last time, at $z_* = 1089.80 \pm 0.21$, but those photons that interacted with baryons later than average kept the oscillation going on for a bit longer (the d in r_d stands for 'drag'), and the scale r_d in the matter distribution reflect the time when baryons on average stopped oscillating, at $z_d = 1060.01 \pm 0.29$. The numerical values are the Planck+BAO results from [34].

The BAO scale can be observed as enhanced correlation between galaxy positions at this separation. In the line-of-sight direction this corresponds to a redshift difference

$$\Delta z = H(z) r_d \equiv \frac{r_d}{D_H(z)}\,, \tag{10.2}$$

where $D_H(z) \equiv H(z)^{-1}$ is the Hubble distance at redshift z; and in the transverse direction to an angular separation on the sky

$$\Delta\vartheta = \frac{r_d}{D_M(z)}\,. \tag{10.3}$$

By measuring the statistics of the galaxy distribution and determining at what Δz and $\Delta\vartheta$ this enhanced correlation appears at different redshifts one can determine the functions $H(z)$ and $D_M(z)$, and thus the correct cosmology.

The correlation function $\xi(\mathbf{s}) = \xi(s,\mu) = \xi(s_\parallel, s_\perp)$ measures separations in all directions, and the results are often expressed as the weighted geometric mean

$$D_V(z) \;\equiv\; \left[z D_M(z)^2 D_H(z)\right]^{1/3} \tag{10.4}$$

and the ratio

$$F_{AP}(z) \;\equiv\; \frac{D_M(z)}{D_H(z)}\,. \tag{10.5}$$

Since two of the three components of $\mathbf{s}$ are transverse and only one is radial, the transverse directions carry more weight in the data, and D_M can be determined more accurately than D_H. The F_{AP} is the Alcock–Paczyński parameter from Sec. 7.6. The comoving volume element is given by

$$dV \;=\; D_H(z) dz D_M(z)^2 d\Omega \;=\; D_V(z)^3 \frac{dz}{z} d\Omega\,. \tag{10.6}$$

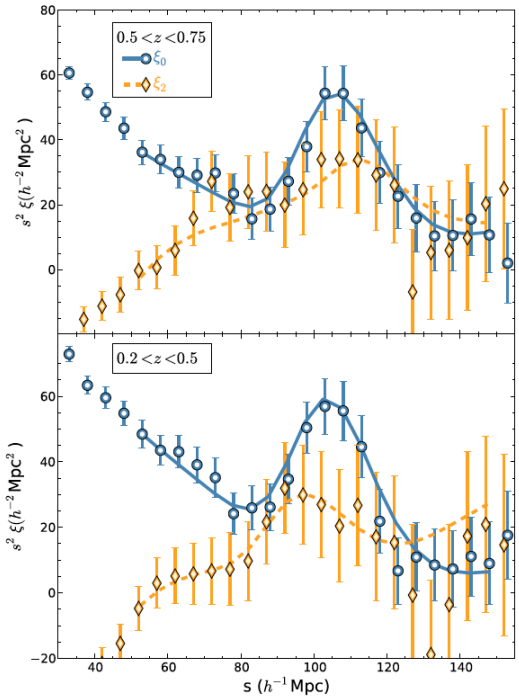


Figure 30: The monopole (blue) and quadrupole (yellow) of the two-point correlation function around the BAO feature, measured by the BOSS survey at two different redshift bins. This is Fig. 11 from [42].

The 'distortion' of the shape of the volume element (ratio of the comoving distances corresponding to Δz and $\Delta\vartheta$) compared to flat space with Hubble law $D(z) = H_0^{-1} z$ is

$$\frac{z D_H(z)}{D_M(z)} = \frac{z}{F_{AP}(z)}; \tag{10.7}$$

this quantity equals 1 at $z = 0$ and decreases with increasing z for ΛCDM (see Fig. 26) since a given Δz corresponds to a shorter comoving separation for higher z.

In reality the determination of these quantities from measured $\xi(\mathbf{s})$ is more complicated, since the value of r_d also depends somewhat on cosmology, so one has to correct for that. Also, the peak of the 2-point correlation function is not exactly at r_d. Rather, one needs to calculate the theoretical $\xi(\mathbf{s})$ and compare it to the observed $\xi(\mathbf{s})$.

Ross et al. [42] describe how the determination was done from the SDSS-III BOSS survey 2-point correlation function. We present here a simplified version of this description.

They start by assuming a fiducial cosmology, using CAMB to calculate the linear power spectrum $P(k)$ for it, adding redshift-space distortions and nonlinear effects to it (we omit the description of how this was done). They then Fourier transform this anisotropic power spectrum (or rather its $\ell = 0$, $\ell = 2$, and $\ell = 4$ multipoles) to obtain the anisotropic two-point correlation function $\xi^{\rm fid}(s, \mu)$ and its multipoles for the fiducial cosmology.

The fiducial cosmology has some $H^{\rm fid}(z)$, $D_M^{\rm fid}(z)$ and $r_d^{\rm fid}$. The observed redshifts are converted to comoving distances and the measured correlation function is calculated from the data assuming the geometry of the fiducial cosmology. Comparing the measured $\xi^{\rm obs}(s, \mu)$ to $\xi^{\rm fid}(s, \mu)$ should find the BAO feature shifted by factors

$$\alpha_\parallel \equiv \frac{H^{\rm fid}(z) r_d^{\rm fid}}{H(z) r_d}, \qquad \alpha_\perp \equiv \frac{D_M^{\rm fid}(z) r_d^{\rm fid}}{D_M(z) r_d} \tag{10.8}$$

in the line-of sight ($\mu = 1$) and transverse ($\mu = 0$) directions.

If we made the, too simplistic, assumption that the effect of the different cosmology is just to strecth/compress $\xi(s,\mu) = \xi(s_\parallel, s_\perp)$ by these factors, we would get that

$$\xi^{\rm obs}(s_\parallel, s_\perp) \;=\; \xi^{\rm fid}(\alpha_\parallel s_\parallel, \alpha_\perp s_\perp)\,, \tag{10.9}$$

or

$$\xi^{\rm obs}(s,\mu) \;=\; \xi^{\rm fid}(s',\mu')\,, \tag{10.10}$$

where (**exercise**)

$$s' \;=\; s\sqrt{\mu^2\alpha_\parallel^2 + (1-\mu^2)\alpha_\perp^2} \qquad \text{and} \qquad \mu' \;=\; \frac{\mu\alpha_\parallel}{\sqrt{\mu^2\alpha_\parallel^2 + (1-\mu^2)\alpha_\perp^2}}\,. \tag{10.11}$$

The observed monopole and quadrupole would then be

$$\begin{aligned}
\xi_0^{\rm obs}(s) \;&\equiv\; \int_0^1 \xi^{\rm obs}(s,\mu)d\mu \;=\; \int_0^1 \xi^{\rm fid}(s',\mu')d\mu \;\equiv\; \xi_0(s,\alpha_\parallel,\alpha_\perp)\,,\\
\xi_2^{\rm obs}(s) \;&\equiv\; \tfrac{5}{2}\int_0^1 \xi^{\rm obs}(s,\mu)(3\mu^2-1)d\mu \;=\; \tfrac{5}{2}\int_0^1 \xi^{\rm fid}(s',\mu')(3\mu^2-1)d\mu\\
&\equiv\; \tfrac{5}{2}\left[\xi_{\mu 2}(s,\alpha_\parallel,\alpha_\perp) - \xi_0(s,\alpha_\parallel,\alpha_\perp)\right]\,,
\end{aligned} \tag{10.12}$$

where the last line defined

$$\xi_{\mu 2}(s,\alpha_\parallel,\alpha_\perp) \;\equiv\; \int_0^1 \xi^{\rm fid}(s',\mu')3\mu^2 d\mu \tag{10.13}$$

(to be used at this step instead of defining the whole thing on the second line to be $\xi_2(s,\alpha_\parallel,\alpha_\perp)$).

We now have to account for the fact that changing the cosmology will change $\xi^{\rm obs}$ also in other ways than this stretching. Our focus is in fitting the BAO feature around r_d, which is a relatively sharp peak. The peak occurs near $s = 100\,h^{-1}$Mpc and therefore we only fit the part of $\xi^{\rm obs}$ at $50 < s < 150\,h^{-1}$Mpc. We allow for the change of cosmology to change the amplitude of the BAO feature but otherwise we expect the changes to be relatively smooth in the vicinity of r_d. Thus we introduce additional fitting parameters B_0 and B_2 to represent rescaling of the BAO feature in monopole and quadrupole and then allow smooth changes in the form of arbitrary second-order polynomials $A_{\ell 1} + A_{\ell 2}s + A_{\ell 3}s^2$ added to $s^2\xi_\ell(s)$. We arrive at the models

$$\begin{aligned}
\xi_0^{\rm mod}(s) \;&=\; B_0\xi_0(s,\alpha_\parallel,\alpha_\perp) + A_{01}s^{-2} + A_{02}s^{-1} + A_{03}\\
\xi_2^{\rm mod}(s) \;&=\; \tfrac{5}{2}\left[B_2\xi_{\mu 2}(s,\alpha_\parallel,\alpha_\perp) - B_0\xi_0(s,\alpha_\parallel,\alpha_\perp)\right] + A_{21}s^{-2} + A_{22}s^{-1} + A_{23}
\end{aligned} \tag{10.14}$$

to be fit to $\xi_0^{\rm obs}(s)$ and $\xi_2^{\rm obs}(s)$. Thus 10 parameters, $\alpha_\parallel$, $\alpha_\perp$, B_0, B_2, A_{01}, A_{02}, A_{03}, A_{21}, A_{22}, and A_{23} are fit to the data, but we only care about what we get for $\alpha_\parallel$ and $\alpha_\perp$. The other eight are 'nuisance' parameters allowing for changes in the shape of $\xi(s,\mu)$ that we did not want to calculate. Thus this procedure is aimed at fitting the BAO scale only, it is not about fitting the full theoretical $\xi(s,\mu)$ to the observed one.

Figure 30 shows ξ_0 and ξ_2 from the BOSS survey; the results from BOSS[35] on $D_M(z)$ and $H(z)$ are shown in Fig. 31, and a "world compilation" of BAO results on $D_M(z)$, $D_V(z)$, and $D_H(z)$ is shown in Fig. 32.

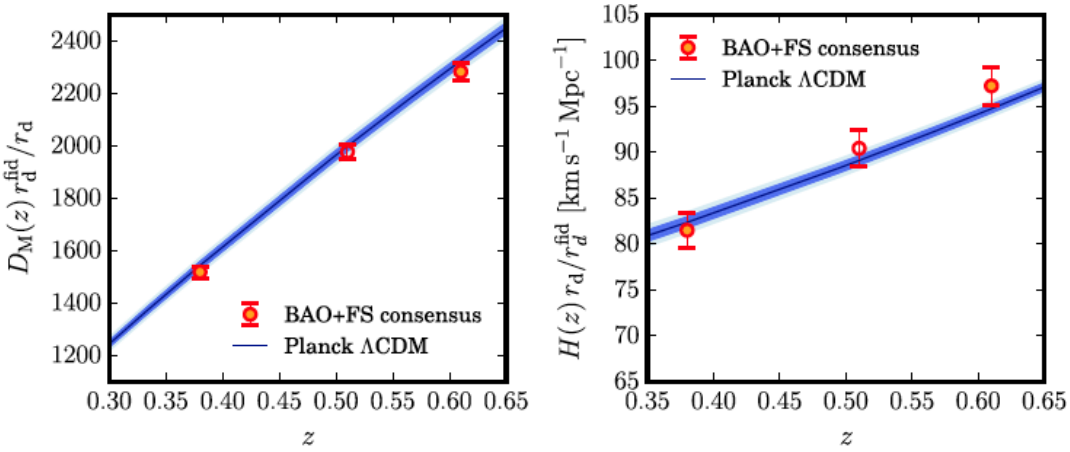


Figure 31: Comoving angular diameter distance $D_M(z)$ and Hubble parameter $H(z)$ measured from BAO by the BOSS survey. The blue lines are ΛCDM predictions using cosmological parameters from Planck[19]. Solid and dashed lines are the best-fit models. (FS stands for 'full shape', i.e., the BOSS team determined these quantities independently from the location of the BAO feature and from the full shape of the correlation function and power spectrum.) This is part of Fig. 12 from [35].

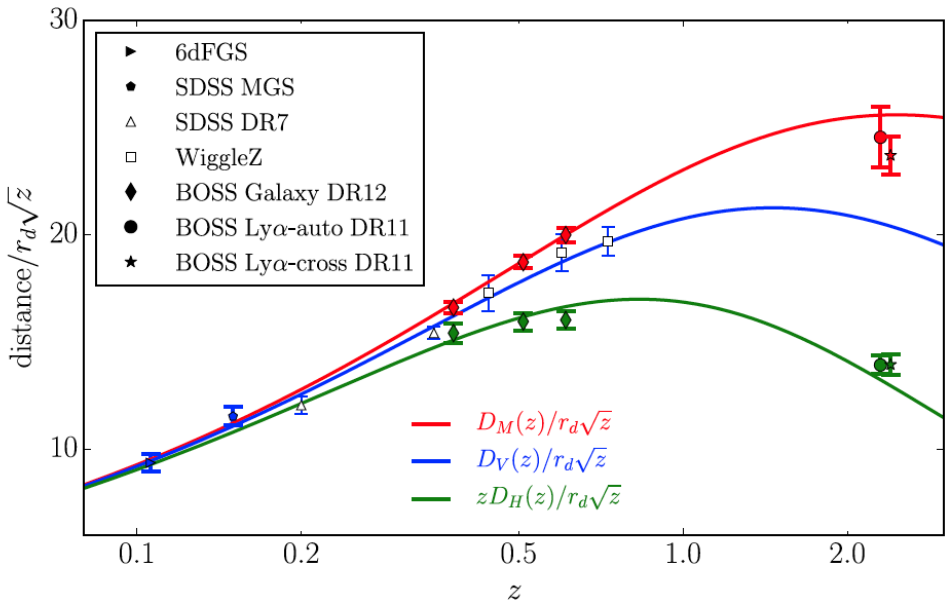


Figure 32: Compilation of BAO determinations of $D_M(z)$, $D_V(z)$, and $D_H(z)$ from different surveys. The scaling by $\sqrt{z}$ is just to compress the curves so that they fit better in the plot with error bars clearly visible. The solid lines are the ΛCDM prediction with $\Omega_m = 0.3156$, $h = 0.6727$. This is Fig.14 from [35]. See Fig. 27 for an updated result from eBOSS [61].

11 Higher-order statistics

11.1 N-point correlation function and (N-1)-spectrum

In analogy with the 2-point correlation function we define the N-point correlation function

$$\xi^{(N)}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) \equiv \langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2)\ldots\delta(\mathbf{x}_N)\rangle \tag{11.1}$$

for a continuous density perturbation field $\delta(\mathbf{x})$. For a distribution of discrete objects

$$d\mathbf{P} = \langle\rho\rangle^N \left\langle \left[1 + \xi^{(N)}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)\right]\right\rangle dV_1 dV_2 \ldots dV_N \tag{11.2}$$

is the probability for finding N objects located at $\mathbf{x}_1$, $\mathbf{x}_2$, ...,$\mathbf{x}_N$. Note that these quantities are invariant under permutation of $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$.

The *connected* N-point correlation function[56] is the remaining part of the N-point correlation function when we subtract from it all the contributions from the lower-order correlations:

$$\zeta^{(N)}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) \equiv \xi^{(N)}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) - \sum_P \zeta^{(N_1)} \ldots \zeta^{(N_n)} \tag{11.3}$$

where $\sum N_i = N$ and $\sum_P$ sums over all partitions of the set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ into smaller subsets. In particular,

$$\begin{aligned}
\zeta^{(1)}(\mathbf{x}) &\equiv \xi^{(1)}(\mathbf{x}) = \langle\delta(\mathbf{x})\rangle = 0 \\
\zeta^{(2)}(\mathbf{x}_1, \mathbf{x}_2) &\equiv \xi^{(2)}(\mathbf{x}_1, \mathbf{x}_2) - \zeta^{(1)}(\mathbf{x}_1)\zeta^{(1)}(\mathbf{x}_2) = \xi^{(2)}(\mathbf{x}_1, \mathbf{x}_2) \\
\zeta^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) &\equiv \xi^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) - \zeta^{(2)}(\mathbf{x}_1, \mathbf{x}_2)\zeta^{(1)}(\mathbf{x}_3) - \zeta^{(2)}(\mathbf{x}_1, \mathbf{x}_3)\zeta^{(1)}(\mathbf{x}_2) \\
&\quad -\zeta^{(2)}(\mathbf{x}_2, \mathbf{x}_3)\zeta^{(1)}(\mathbf{x}_1) - \zeta^{(1)}(\mathbf{x}_1)\zeta^{(1)}(\mathbf{x}_2)\zeta^{(1)}(\mathbf{x}_2) \\
&= \xi^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \\
\zeta^{(4)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) &\equiv \xi^{(4)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) - \zeta^{(2)}(\mathbf{x}_1, \mathbf{x}_2)\zeta^{(2)}(\mathbf{x}_3, \mathbf{x}_4) - \zeta^{(2)}(\mathbf{x}_1, \mathbf{x}_3)\zeta^{(2)}(\mathbf{x}_2, \mathbf{x}_4) \\
&\quad -\zeta^{(2)}(\mathbf{x}_1, \mathbf{x}_4)\zeta^{(2)}(\mathbf{x}_2, \mathbf{x}_3)\,,
\end{aligned} \tag{11.4}$$

so that the connected correlation function differs from the full correlation function only for $N \geq 4$.

For Gaussian density perturbations $\zeta^{(N)} = 0$ for $N > 2$, so that the statistics of a Gaussian field is fully determined by $\xi^{(2)}(\mathbf{x}_1, \mathbf{x}_2)$, and the connected higher-order statistics measure non-Gaussianity.

Assuming statistical homogeneity, $\xi^{(N)}$ depends only on the separations between the points. Picking one point, say $\mathbf{x}_N$, as the reference point and writing $\mathbf{x}_i = \mathbf{x}_N + \mathbf{r}_i$,

$$\xi^{(N)}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = \xi^{(N)}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N-1})\,. \tag{11.5}$$

Because of the permutation invariance, we could order the $\mathbf{r}_i$ by length, so that $|\mathbf{r}_1| \leq |\mathbf{r}_2| \leq \ldots \leq |\mathbf{r}_{N-1}|$, and the values of $\xi^{(N)}$ over this subset of the $d^{(N-1)}$-dimensional space define it completely. There is, however, and additional symmetry since we are free to pick the reference point. For $\xi^{(2)}$, this symmetry is $\xi^{(2)}(\mathbf{r}) = \xi^{(2)}(-\mathbf{r})$, and we discuss $\xi^{(3)}$ in the next subsection.

Assuming also statistical isotropy, $\xi^{(N)}$ is invariant under the rotation of the full N-point configuration. We can rotate it (or the coordinate system) so that $\mathbf{r}_1$ becomes parallel to the x-axis, $\mathbf{r}_2$ will lie in the xy-plane (and so on, if $d > 3$), reducing the number of components $\xi^{(N)}$ depends on to $1 + 2 + \ldots + d + \ldots + d$, where there are $N - 1$ terms in the sum, so that

[56] I don't know if there's a standard notation for this; ζ is the usual notation for the 3-point correlation function (connected or not, for it they are the same).

$\xi^{(2)}$ depends on one component (r), $\xi^{(3)}$ depends on $1+2=3$ components (for $d \geq 2$), and $\xi^{(4)}$ depends on $1+2+3=6$ components for $d \geq 3$. (When we go to redshift space, we lose statistical isotropy, and $\xi^{(N)}$ will depend on more components.)

In analogy with the power spectrum we define the $(N-1)$-spectrum[57]

$$P^{(N-1)}(\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_N) \;\equiv\; V^{N-1}\langle \delta_{\mathbf{k}_1}\delta_{\mathbf{k}_2}\ldots\delta_{\mathbf{k}_N}\rangle\,. \tag{11.6}$$

where

$$\delta_{\mathbf{k}} = \frac{1}{V}\int d^d x\, \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}\,. \tag{11.7}$$

Using periodic boundary conditions and statistical isotropy,

$$\begin{aligned}
\langle \delta_{\mathbf{k}_1}\ldots\delta_{\mathbf{k}_N}\rangle &= \frac{1}{V^N}\int d^d x_1 \ldots d^d x_N e^{-i\sum^N \mathbf{k}_i\cdot\mathbf{x}_i}\langle\delta(\mathbf{x}_1)\delta(\mathbf{x}_2)\ldots\delta(\mathbf{x}_N)\rangle \\
&= \frac{1}{V^N}\int d^d r_1 \ldots d^d r_{N-1} d^d x e^{-i\sum^{N-1}\mathbf{k}_i\cdot(\mathbf{x}+\mathbf{r}_i)} e^{-i\mathbf{k}_N\cdot\mathbf{x}}\langle\delta(\mathbf{x}+\mathbf{r}_1)\ldots\delta(\mathbf{x}+\mathbf{r}_{N-1})\delta(\mathbf{x})\rangle \\
&= \frac{1}{V^N}\int d^d r_1 \ldots d^d r_{N-1} e^{-i\sum^{N-1}\mathbf{k}_i\cdot\mathbf{r}_i}\xi^{(N)}(\mathbf{r}_1,\ldots,\mathbf{r}_{N-1})\int d^d x e^{-i(\sum^N \mathbf{k}_i)\cdot\mathbf{x}}\,,
\end{aligned} \tag{11.8}$$

where the last integral gives V, if $\sum_{i=1}^N \mathbf{k}_i = 0$, otherwise zero.

Thus we have that

$$P^{(N-1)}(\mathbf{k}_1,\ldots,\mathbf{k}_N) \;=\; \int d^d r_1 \ldots d^d r_{N-1} e^{-i\sum^{N-1}\mathbf{k}_i\cdot\mathbf{r}_i}\xi^{(N)}(\mathbf{r}_1,\ldots,\mathbf{r}_{N-1}) \tag{11.9}$$

for $\mathbf{k}_N = -\sum_{i=1}^{N-1}\mathbf{k}_i$ and otherwise $P^{(N-1)}$ is zero. Thus, because of statistical homogeneity, the nonzero part of $P^{(N-1)}$ depends only on the first $N-1$ wave vectors and we can write it as

$$P^{(N-1)}(\mathbf{k}_1,\ldots,\mathbf{k}_{N-1}) \tag{11.10}$$

The correlation functions vanishes at large separations, so that we can extend the integrals in (11.9) to infinity. Thus the N-point correlation function and the $(N-1)$-spectrum form an $(N-1)d$-dimensional Fourier transform pair and

$$\xi^{(N)}(\mathbf{r}_1,\ldots,\mathbf{r}_{N-1}) \;=\; \frac{1}{(2\pi)^{(N-1)d}}\int d^d k_1 \ldots d^d k_{N-1} e^{+i\sum^{N-1}\mathbf{k}_i\cdot\mathbf{r}_i} P^{(N-1)}(\mathbf{k}_1,\ldots,\mathbf{k}_{N-1})\,. \tag{11.11}$$

In the Fourier integral notation, $\delta(\mathbf{k})$ in place of $V\delta_{\mathbf{k}}$, we write

$$\langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\ldots\delta(\mathbf{k}_N)\rangle \;=\; (2\pi)^d\delta_D^d\left(\sum_{i=1}^N \mathbf{k}_i\right)P^{(N-1)}(\mathbf{k}_1,\ldots,\mathbf{k}_{N-1}) \tag{11.12}$$

11.2 Three-point correlation function and bispectrum

For $N=3$, assuming statistical homogeneity, we have the 3-point correlation function

$$\zeta(\mathbf{r}_1,\mathbf{r}_2) \;\equiv\; \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r}_1)\delta(\mathbf{x}+\mathbf{r}_2)\rangle \tag{11.13}$$

[57] Again, I do not know if there is a standard notation or what is the standard term for this. The expression "$(N-1)$-spectrum" and notation $P^{(N-1)}$ are my own. For $N=3$ and $N=4$ the standard notations and terms are B, the bispectrum, and T, the trispectrum. $P^{(1)}(\mathbf{k}_1,\mathbf{k}_2) = V\langle\delta_{\mathbf{k}_1}\delta_{\mathbf{k}_2}\rangle$ is the same as the power spectrum we defined earlier as $P(\mathbf{k}) = \langle|\delta_{\mathbf{k}}|^2\rangle$. We get below that $\mathbf{k}_2 = -\mathbf{k}_1$. Because of the reality of $\delta(\mathbf{x})$, $\delta_{\mathbf{k}_2} = \delta_{-\mathbf{k}_1}$ is equal to $\delta^*_{\mathbf{k}_1}$, so indeed $\delta_{\mathbf{k}_1}\delta_{\mathbf{k}_2} = |\delta_{\mathbf{k}_1}|^2$.

and its Fourier transform, the *bispectrum*

$$\begin{aligned} B(\mathbf{k}_1,\mathbf{k}_2) &= \int d^d r_1 d^d r_2 e^{-i\mathbf{k}_1\cdot\mathbf{r}_1 - i\mathbf{k}_2\cdot\mathbf{r}_2}\zeta(\mathbf{r}_1,\mathbf{r}_2) \\ \zeta(\mathbf{r}_1,\mathbf{r}_2) &= \frac{1}{(2\pi)^{2d}}\int d^d k_1 d^d k_2 e^{i\mathbf{k}_1\cdot\mathbf{r}_1 + i\mathbf{k}_2\cdot\mathbf{r}_2} B(\mathbf{k}_1,\mathbf{k}_2)\,, \end{aligned} \tag{11.14}$$

with

$$\langle \delta_{\mathbf{k}_1}\delta_{\mathbf{k}_2}\delta_{\mathbf{k}_3}\rangle = \frac{1}{V^2}\delta_{\mathbf{k}_1+\mathbf{k}_2,-\mathbf{k}_3} B(\mathbf{k}_1,\mathbf{k}_2) \tag{11.15}$$

or

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle = V^3\langle \delta_{\mathbf{k}_1}\delta_{\mathbf{k}_2}\delta_{\mathbf{k}_3}\rangle \to (2\pi)^d\delta_D^d(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)B(\mathbf{k}_1,\mathbf{k}_2)\,. \tag{11.16}$$

The bispectrum may also be denoted as $B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$, with the condition $\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3=0$ understood.

Assuming statistical isotropy, both $\zeta(\mathbf{r}_1,\mathbf{r}_2)$ and $B(\mathbf{k}_1,\mathbf{k}_2)$ are *functions of triangles.* In case of ζ the triangle is formed by the three points, in case of B, the triangle is formed by the closed loop of the wave vectors $\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3=0$. Both of them depend on three scalar quantities, which are required to define a triangle, e.g., the lengths of the three sides, or two sides and the angle between them.

Functions of three variables are already complicated to study and plot. One may choose to concentrate on some selected triangle shapes, e.g., equilateral triangles.

Another approach is multipole expansion [43]. Express the bispectrum as a function of the side lengths k_1 and k_2, and the angle θ between them, and expand the angular dependence in Legendre polynomials,

$$B(k_1,k_2,\theta) = \sum_\ell B_\ell(k_1,k_2)L_\ell(\cos\theta)\,, \tag{11.17}$$

where

$$B_\ell(k_1,k_2) = \frac{2\ell+1}{2}\int_{-1}^{1} B(k_1,k_2,\theta)L_\ell(\cos\theta)d\cos\theta\,. \tag{11.18}$$

We expand the 3-point correlation function likewise

$$\zeta(r_1,r_2,\theta) = \sum_\ell \zeta_\ell(r_1,r_2)L_\ell(\cos\theta)\,, \tag{11.19}$$

where

$$\zeta_\ell(r_1,r_2) = \frac{2\ell+1}{2}\int_{-1}^{1} \zeta(r_1,r_2,\theta)L_\ell(\cos\theta)d\cos\theta\,. \tag{11.20}$$

Only the lowest multipoles are expected to be important. This allows easier display of results, since each multipole can be plotted as a 2D contour or color plot. Note that this was assuming statistical isotropy, the angle is related to the shape of the triangle, not to direction with respect to line of sight as in Sec. 7; adding redshift-space distortion, ζ and B become functions of more variables.

Example: Relate ζ_ℓ to B_ℓ in 3D. (This is from [43].) We need here (5.30),

$$e^{i\mathbf{k}\cdot\mathbf{x}} = 4\pi\sum_{\ell m} i^\ell j_\ell(kx)Y_{\ell m}(\hat{\mathbf{x}})Y^*_{\ell m}(\hat{\mathbf{k}}) = 4\pi\sum_{\ell m} i^\ell j_\ell(kx)Y^*_{\ell m}(\hat{\mathbf{x}})Y_{\ell m}(\hat{\mathbf{k}}) \tag{11.21}$$

and (5.28),

$$L_\ell(\cos\vartheta_n) = \frac{4\pi}{2\ell+1}\sum_m Y^*_{\ell m}(\hat{\mathbf{n}}_1)Y_{\ell m}(\hat{\mathbf{n}}_2)\,. \tag{11.22}$$

Starting from (11.14) with (11.17),

$$\begin{aligned}
\zeta(\mathbf{r}_1,\mathbf{r}_2) &= \frac{1}{(2\pi)^6}\int d^3k_1 d^3k_2 e^{i\mathbf{k}_1\cdot\mathbf{r}_1}e^{i\mathbf{k}_2\cdot\mathbf{r}_2}\sum_\ell B_\ell(k_1,k_2)L_\ell(\cos\vartheta_k)\\
&= \frac{1}{4\pi^4}\sum_\ell\sum_{\ell_1 m_1}\sum_{\ell_2 m_2}\int k_1^2dk_1\int k_2^2dk_2\, i^{\ell_1}i^{\ell_2}j_{\ell_1}(k_1r_1)j_{\ell_2}(k_2r_2)B_\ell(k_1,k_2)\\
&\quad \frac{4\pi}{2\ell+1}\sum_m\int d\Omega_1 d\Omega_2 Y^*_{\ell_1 m_1}(\hat{\mathbf{r}}_1)Y_{\ell_1 m_1}(\hat{\mathbf{k}}_1)Y_{\ell_2 m_2}(\hat{\mathbf{r}}_2)Y^*_{\ell_2 m_2}(\hat{\mathbf{k}}_2)Y^*_{\ell m}(\hat{\mathbf{k}}_1)Y_{\ell m}(\hat{\mathbf{k}}_2)\\
&= \frac{1}{4\pi^4}\sum_\ell\int k_1^2dk_1\int k_2^2dk_2\,(-1)^\ell j_\ell(k_1r_1)j_\ell(k_2r_2)B_\ell(k_1,k_2)\frac{4\pi}{2\ell+1}\sum_m Y^*_{\ell m}(\hat{\mathbf{r}}_1)Y_{\ell m}(\hat{\mathbf{r}}_2)\\
&= \frac{1}{4\pi^4}\sum_\ell\int k_1^2dk_1\int k_2^2dk_2\,(-1)^\ell j_\ell(k_1r_1)j_\ell(k_2r_2)B_\ell(k_1,k_2)L_\ell(\cos\vartheta_r)\,, \qquad (11.23)
\end{aligned}$$

from which we can pick

$$\zeta_\ell(r_1,r_2) = \frac{1}{4\pi^4}\int k_1^2dk_1\int k_2^2dk_2\,(-1)^\ell j_\ell(k_1r_1)j_\ell(k_2r_2)B_\ell(k_1,k_2)\,. \qquad (11.24)$$

From theoretical grounds, the magnitude of the bispectrum of the cosmological density field is expected to be of the order $P(k)^2$, which motivates the definition of the *reduced bispectrum*

$$Q(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) \equiv \frac{B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)}{P(k_1)P(k_2)+P(k_2)P(k_3)+P(k_3)P(k_1)}\,. \qquad (11.25)$$

11.3 Measuring the three-point correlation function

In Sec. 8 we found that the best estimator (minimizing bias and variance) for the two-point correlation function is the Landy–Szalay estimator, which can be written symbolically as

$$\widehat{\xi}_{\rm LS} = \frac{(D-R)^2}{R^2}\,. \qquad (11.26)$$

This can be generalized to N-point correlation functions as the Szapudi–Szalay estimator [44]

$$\widehat{\xi}^{(N)}_{SS} = \frac{(D-R)^N}{R^N}\,. \qquad (11.27)$$

For the three-point correlation function this becomes

$$\widehat{\zeta}_{SS} = \frac{(D-R)^3}{R^3} = \frac{DDD-3DDR+3DRR-RRR}{RRR}\,, \qquad (11.28)$$

where DDD etc. denote the number of triplets of data and random points corresponding to the given triangle configuration. These configurations are binned, e.g. into bins of $(r_1\pm\frac{1}{2}\Delta r_1, r_2\pm\frac{1}{2}\Delta r_2, \theta\pm\frac{1}{2}\Delta\theta)$, and the triplet counts have to be scaled by the total numbers of the four different triplet types, like we did for pair counts in Sec. 8. (One can also combine the data and random catalogs into a $D-R$ catalog, where the random points have negative weights, and just count $D-R$ triplets.)

The computational cost of this estimator scales as the number of triplets, which goes as N^3, where N is the number of data or random points. One usually uses many more random points than data points, so the relevant N is the number of random points. The correlation function is usually needed only up to some $r_{\rm max}$, so one can save computation by arranging the points into a data structure which helps to skip triplets with larger separations than $r_{\rm max}$, but even then the computational cost is a problem for large surveys.

Slepian&Eisenstein [45] have proposed an approximate method, which scales as N^2, and will be faster for sufficiently large surveys.[58] See also [46].

[58] I haven't yet read their paper, so I am not saying more about it now.

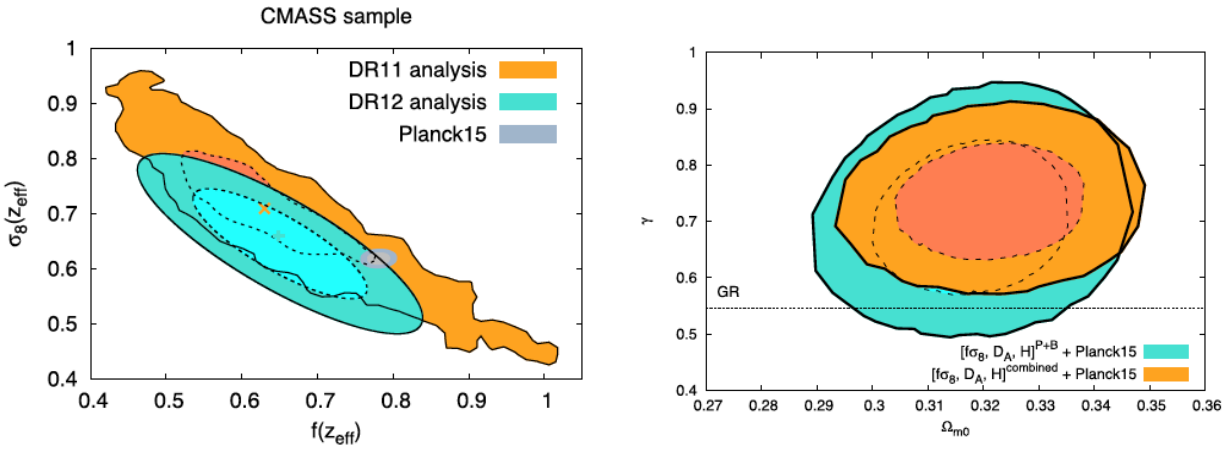


Figure 33: Left: Measurement of f and σ_8 from the CMASS sample of the BOSS survey. The effective redshift for the CMASS sample is $z = 0.57$. 'Planck15' refers to the theoretical prediction of ΛCDM cosmology, using cosmological parameters determined from Planck data (i.e., not a Planck measurement of f and σ_8). This is Fig. 13 of [49]. Right: Constraints on the growth index γ and Ω_m from combined BOSS and Planck data. This is Fig. 18 of [49].

11.4 Higher-order statistics in cosmology

Higher-order statistics (bispectrum, 3-point correlation function, and the reduced (higher than three)-correlation functions) vanish for Gaussian perturbations. Primordial perturbations are thought to be Gaussian or very close to Gaussian (Planck did not find any signal of primordial non-Gaussianity; the simplest inflation models predict very-close-to-Gaussian primordial perturbations). Linear evolution maintains Gaussianity of perturbations. Higher-order statistics measure thus nonlinear effects even at large scales. Therefore they provide tests of modified gravity. Alternatively, they test primordial non-Gaussianity. Combining higher-order statistics results with power spectrum and/or 2-point correlation results can be used to break the degeneracy between f and σ_8 (or bias), i.e., both can be determined separately and not just their product.

(Here should be a section on higher-order perturbation theory to show how higher-order correlations arise.)

11.5 Results from galaxy surveys

The 3-point correlation function measured in the BOSS survey is discussed in [47] and [48] and the bispectrum measured in the BOSS survey is described in [49].

Fig. 33 shows the BOSS determination of the growth rate f, σ_8, and growth index γ. BOSS (combined with Planck data) finds $\gamma = 0.733 \pm 0.069$, a 2.7σ deviation from the prediction $\gamma = 0.55$ of GR. They say that "This tension could be due to (i) a statistical fluctuation in the $f\sigma_8$ measured values; (ii) unaccounted systematic uncertainties in BOSS or Planck15 data; (iii) an indication of a failure of the CDM or the GR gravity model. Future galaxy surveys using more redshift bins and more accurate data may shed light on this tension revealing the origin of this discrepancy."[49].

12 Galaxy surveys

12.1 Historical surveys

Some of the material for this section is from [50].

12.1.1 Catalogs of Nebulae

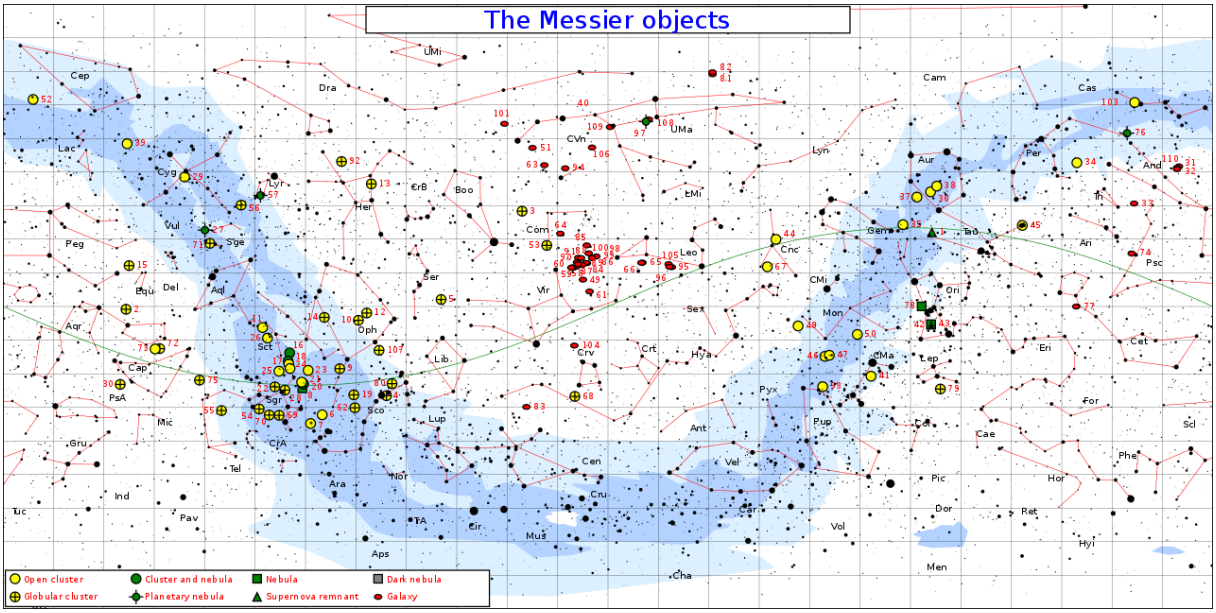


Figure 34: Star chart depicting the Messier objects. From Wikipedia, copied under the terms of the GNU Free Documentation License `https://en.wikipedia.org/wiki/en:GNU_Free_Documentation_License`.

The French astronomer Charles Messier (1730–1817) compiled a catalog of permanent diffuse objects on the sky. His main interest were comets, so the purpose of the catalog was to list objects that looked like comets, but were not, since they were not moving. His first version from 1771 had 45 objects, and by 1781 he had extended it to 103 objects. Based on Messier's notes the list was later extended to 110 objects in the final Messier catalog, *Catalogue des Nébuleuses et des Amas d'Étoiles*. Many of these objects are star clusters (globular or open) or nebulae (planetary or diffuse) in our own galaxy, and M1, the Crab Nebula, is a supernova remnant, but 41 (M31–33, M49, M51, M58–61, M63–66, M74, M77, M81–91, M94–96, M98–102, M104–106, and M108-110) are galaxies. See Fig. 34.

William and Caroline Herschel compiled the *Catalogue of Nebulae and Clusters of Stars* in 1786. By 1810 they had extended it to 2500 objects visible from the northern latitudes. John Herschel extended it to the *General Catalogue of Nebulae and Clusters* in 1864, with 5079 objects, including observations from South Africa. The *New General Catalogue of Nebulae and Clusters of Stars* (NGC), containing 7840 objects, was compiled by John Dreyer in 1888, based on observations by many astronomers. He supplemented it in 1895 and 1908 by two *Index Catalogues* (IC) containing 5386 additional objects. Most of the objects in these catalogs are galaxies. The true nature of galaxies was not known then. These early catalogs included essentially all known diffuse objects, and were not results from a uniform survey, so they were not suitable for measuring the large scale structure of the universe.

12.1.2 2D Surveys

The Lick Survey [21] gave us the first good view of the large-scale structure of the universe. It was already mentioned in Sec. 4, see Fig. 12. It was conducted in 1950–1967 using the 0.5 m refractor at Lick observatory, covered the northern galactic hemisphere and about half of the southern galactic hemisphere, had a limiting magnitude of about $m_* = 19$, and included over 800 000 galaxies. Shane and Wirtanen divided the sky into $10' \times 10'$ cells and counted by eye and hand the number of galaxies in each cell.

Figure 35: A $1° \times 1°$ region of the Palomar Sky Survey (red plate), from the area of the Coma galaxy cluster.

The Palomar Schmidt telescope (aperture 1.22 m or 48 inches; in Southern California) was completed in 1948. (It was renamed the Samuel Oschin telescope in 1986.) It opened the era of Schmidt telescopes, designed to provide a wide field of view with limited aberration, thus suitable for sky surveys. It was used in the 1950s for the Palomar Sky Survey (POSS I). The survey covered the full sky north of declination $-30°$ in 936 pairs (blue- and red-sensitive) of $6° \times 6°$ photographic plates. The limiting magnitude was about 19 for the blue and 20 for the red. Different observatories around the world obtained copies of these images. I remember looking at them (just out of curiosity) at the Helsinki Observatory at around 1980. It can now be accessed digitally, e.g., at `https://archive.stsci.edu/cgi-bin/dss_form`. See Fig. 35 for a sample. The southern sky was covered by similar UK and ESO Schmidt telescopes in Australia and Chile in the 1970s.

The APM (Automated Plate Measuring) survey [51] from 1990, see Fig. 36, covered 4300 squared degrees of the southern galactic hemisphere. It used 185 plates obtained with the UK Schmidt Telescope. Instead of selecting galaxies by eye, it used "the SERC Automated Plate Measuring system in Cambridge" to digitize the survey plates, detect and analyze galaxy images,

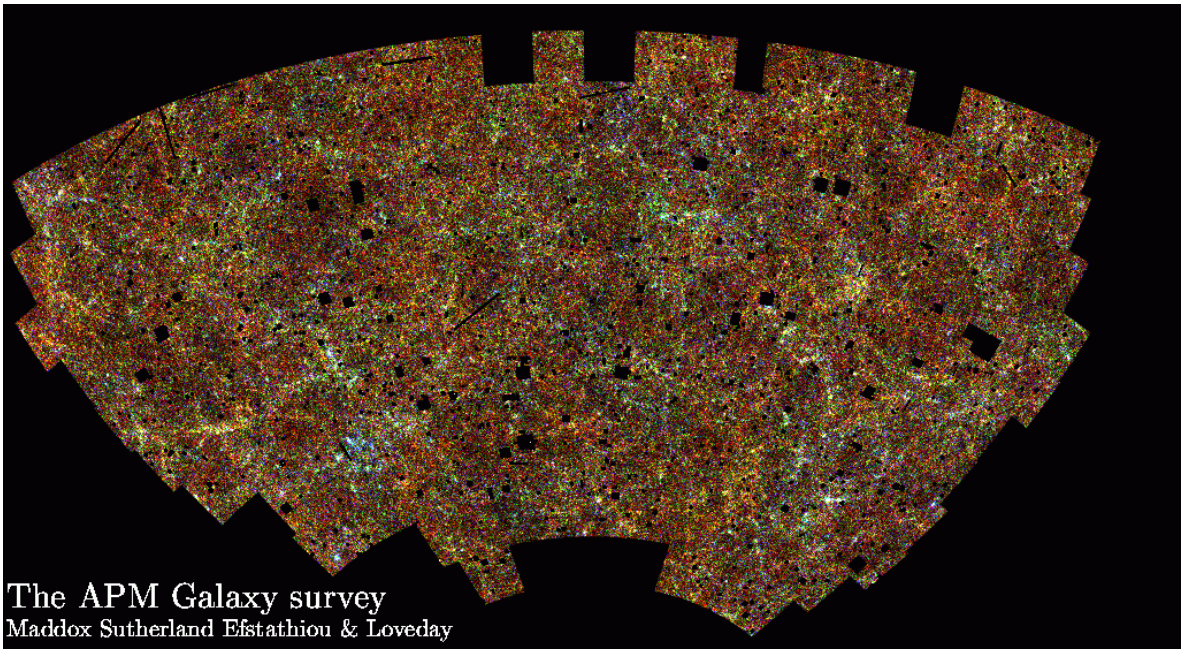


Figure 36: The APM galaxy survey: galaxy number density map.

and produce their position and magnitude. With a limiting magnitude of $m_* = 20.5$ it went deeper than the Lick survey and contained over 2 million galaxies. In addition to speeding up the process and relieving humans from this tedious task, the automated system had the advantage of guaranteeing a uniform treatment of the whole survey area.

12.1.3 Redshift Surveys

For the third dimension, the redshifts of the galaxies need to be measured. This is much more effort than just taking sky images and measuring galaxy positions. To get a good spectrum requires collecting much more light than is needed for an image position.

The first redshift survey of sufficient extent to map the large-scale structure was the Harvard-Smithsonian Center for Astrophysics (CfA) redshift survey. There were actually two surveys, CfA1 with limiting magnitude 14.5 in 1977–82 and CfA2 with limiting magnitude 15.5 in 1985–95. CFA2 contained 18 000 galaxies in the northern sky. The CfA2 survey proceeded in slices of 6° thickness in declination. Fig. 37 shows their first slice, published in 1986 [53]. This can be said to be our first view of the 3D structure in the distribution of galaxies, showing, in addition to clusters, large voids and overdense structures (sheets and/or filaments) extending over much larger scales than the volume of a cluster. The CfA1 survey was not deep enough (did not reach faint enough magnitudes) to reveal these structures.

The Las Campanas redshift survey [54] used the Las Campanas telescope in Chile to measure the redshifts of 26 418 galaxies during 1988–94. The survey covered over 700 square degrees. The median redshift of the survey was $z = 0.1$ (30 000 km/s), so it went much deeper than the CfA2 survey. See Fig. 38.

The 2dF Galaxy Redshift Survey [`http://www.2dfgrs.net/`] used the 2dF multifibre spectrograph on the Anglo-Australian Telescope, which is capable of observing 400 objects simultaneously over a 2° diameter field [55]. The source catalog was the APM survey. The survey covered 1500 square degrees, in two strips (plus some random small fields), with a magnitude limit 19.45, see Fig. 39. Reliable redshifts were obtained for 221 414 galaxies. The median redshift was $z = 0.11$. The final data release was in 2003.

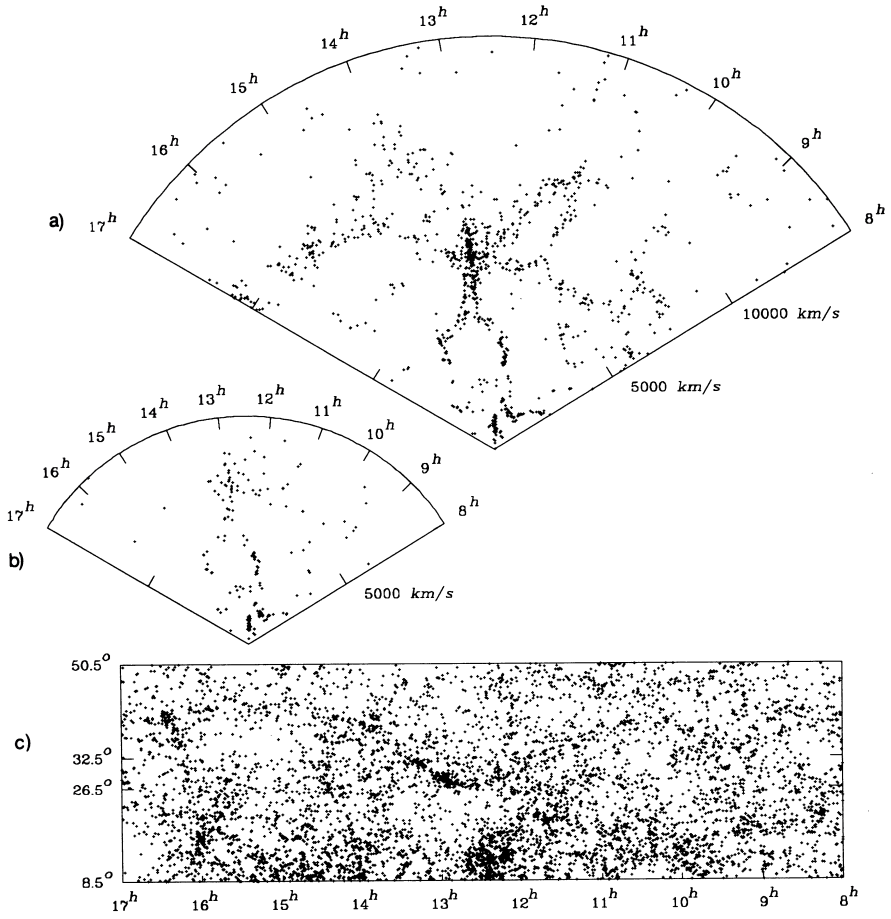


Figure 37: (a) The first slice of the CfA2 survey, declination range $26.5° \leq \delta \leq 32.5°$, containing 1061 galaxies. Redshifts are given in velocity units: the map extends to 15 000 km/s, i.e., to $z = 0.05$. The Coma cluster is at the center, elongated radially by the finger-of-God effect. The narrow overdense structure extending in an arc from 17^h, 10 000 km/s to 8^h, 6000 km/s was named the Great Wall. Extensions of the survey to lower and higher declinations confirmed that the structure indeed extended to this direction and was planar, not a filament. (b) The same as (a), but limited to $m \leq 14.5$ and $v \leq 10\,000$ km/s, containing 182 galaxies. (c) Sky map of the galaxies with $m \leq 15.5$ in the Zwicky catalog that the CfA was using as a starting point for the redshift survey. From [53].

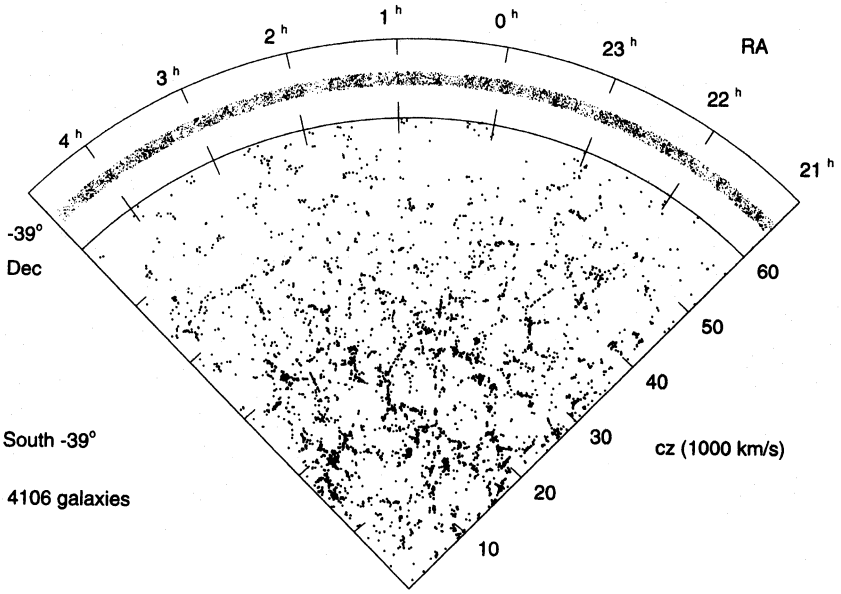


Figure 38: One of the six slices of the Las Campanas redshift survey. The thickness of the slice is 1.5° in declination. From [54].

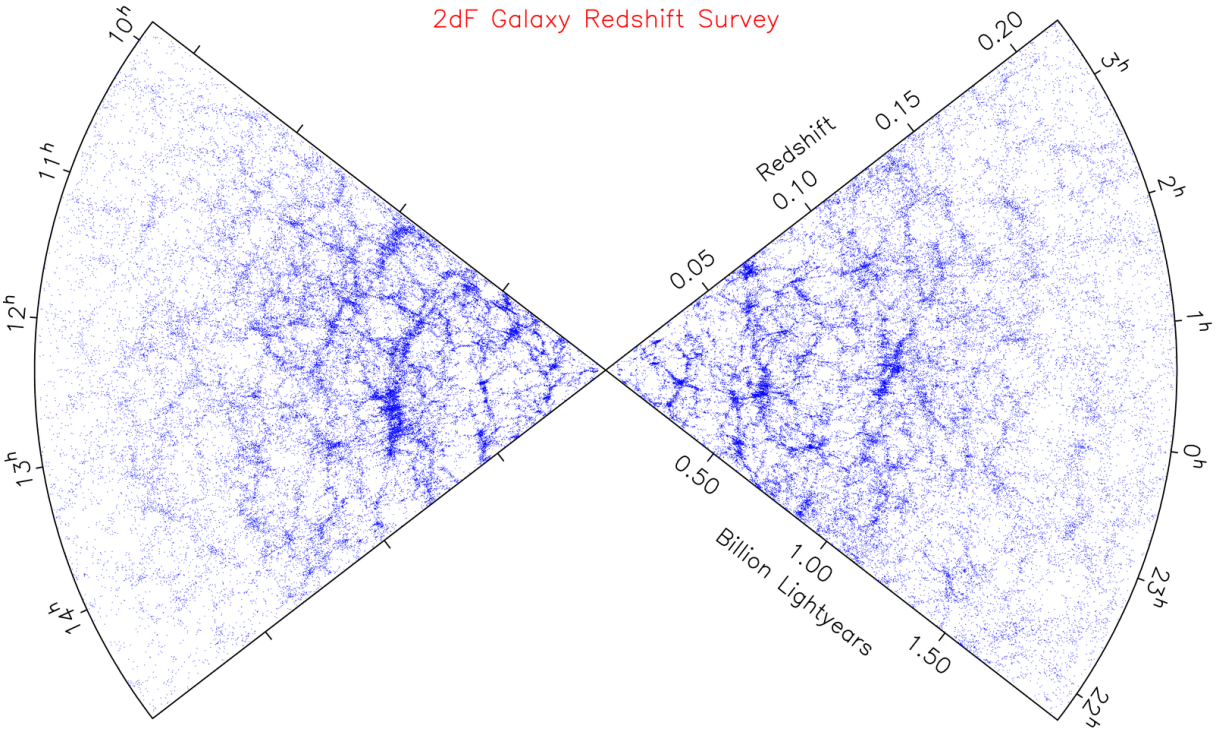


Figure 39: The 2dF Galaxy Redshift Survey. The left cone is an up to 10° wide strip along the celestial equator – compare to Fig. 1. The right cone is an up to 15° wide strip centered at declination $\delta = -30°$.

12.1.4 Weak Lensing Surveys

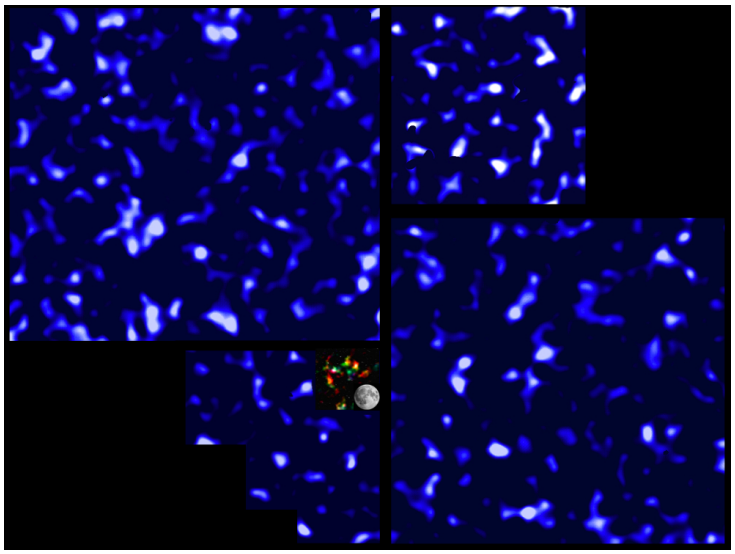


Figure 40: The mass distribution determined by weak lensing in the 4 CFHTLens fields. The central colour inset shows the previous (before CFHTLens) largest weak lensing mass map (COSMOS Dark Matter map, credit: NASA, ESA, P. Simon and T. Schrabback) and the full moon to scale. From `https://www.cfht.hawaii.edu/en/news/CFHTLens/`. Credit: Van Waerbeke, Heymans, and CFHTLens collaboration.

Weak gravitational lensing can be used to map the distribution of all matter, including dark matter. Weak lensing is discussed in the companion course to this, Gravitational Lensing. Very briefly: The gravity of the mass distribution between a distant galaxy and us bends the path of light distorting the image of the galaxy. For most galaxies the effect is small, of the order of 1% distortion, and in this regime of *weak lensing* appears as an elongation of the image in some direction. Since we do not know the undistorted shapes of the images, the distortion can only be determined statistically: how the average shape of many galaxies in a particular small region of sky deviates from circular symmetry. Thus weak lensing surveys require a very large number of galaxy images to be useful. The observed average distortion, called *shear*, provides a measure of integrated (and weighted so that masses half-way along the line of sight carry the heaviest weight) mass between the observer and the source galaxies. The more technical term for this integrated mass measure is *convergence*, denoted κ; and this quantity is actually the deviation of weighted surface mass density (per solid angle) from the background value corresponding to the mean density of the universe, normalized to a critical value, the dividing line between weak and strong lensing.

Thus weak lensing surveys can provide a 2D mass map of the sky. For such a very large number of galaxies it is not practical to obtain accurate spectroscopic redshifts, but for the 2D mass maps they are not needed, we need just a rough redshift estimate provided by the *photometric redshift*: images are taken at several different wavelength bands to provide the "color" of the galaxy, which gives a probability distribution for its redshift. Some level of information about the third dimension of the mass distribution (*tomography*) is provided by dividing the source galaxies into different photometric redshift bins. Typically the width of such a redshift bin is at least $\Delta z = 0.2$.

While for the redshift surveys I took the cut-off for historical surveys to be near year 2000, for the weak lensing surveys I take it to be near 2020. The time of modern weak lensing surveys is just beginning, with Euclid, LSST, and Roman.

The Canada+France Hawaii Telescope Lensing Survey (CFHTLens) covered 154 square de-

grees (in 4 different fields) of sky in five optical bands, with a limiting magnitude 24.5. On average, there were 14 galaxies per square arcmin useful for lensing analysis [56, 57]. See Fig. 40 for the obtained mass maps in the 4 fields. The data was obtained in 2003–2008. There were many analysis papers (see `https://www.cfhtlens.org/astronomers/publications`): e.g., in [58] 4.2 million galaxy images with $m \leq 22$ and redshifts $0.2 \leq z \leq 1.3$ were used for the analysis.

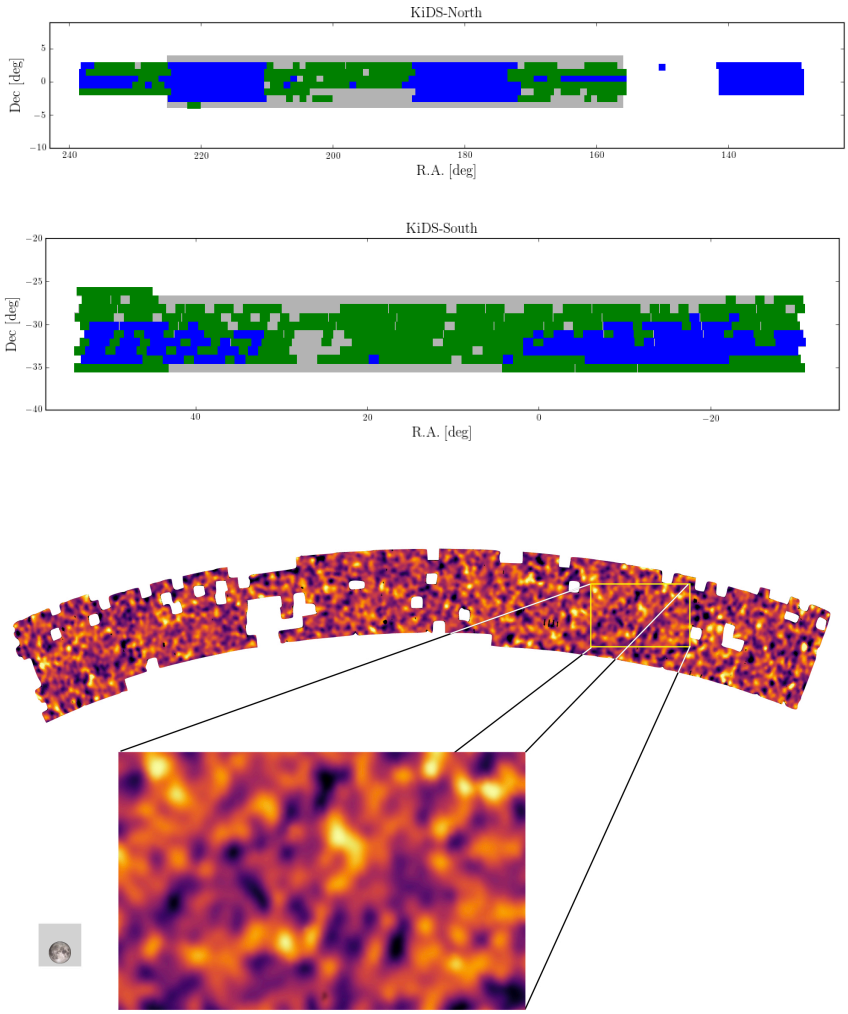


Figure 41: The KiDS survey: Top panel: The area covered in data releases 1–3 is shown in blue, the additional area in DR4 is in green, and the target 1350 deg^2 area is outlined in gray. From `https://kids.strw.leidenuniv.nl/DR4/releasenotes.php`. Bottom panel: Mass map of the KiDS Southern field. High-density regions shown in yellow, low-density regions in dark color. The gray square shows the size of an individual 1 deg^2 KiDS image in the enlarged region, with a full moon for scale. Over 1000 images make up the KiDS map. From `https://kids.strw.leidenuniv.nl/pr_jul2020.php`.

The Kilo-Degree Survey (KiDS) [`https://kids.strw.leidenuniv.nl/`] is a European Southern Observatory (ESO) survey, which began in 2011. The survey uses the VLT Survey Telescope, with mirror diameter 2.61 m and the 268 megapixel OmegaCAM camera, with a 1 deg^2 field of view. The complete survey will cover 1350 deg^2 of sky in two fields, an equatorial and a southern field. See Fig. 41. The fourth data release, in 2019, covered 1006 deg^2 and over 21 million galaxies useful for lensing analysis (reliable shape measurement and photometric redshift). It is difficult to get a good overview of the KiDS results because there are so many KiDS publications (220 papers listed at `https://kids.strw.leidenuniv.nl/papers.php` as of today (May 4, 2023)) without a clear overview paper. See Fig. 41 for a KiDS mass map.

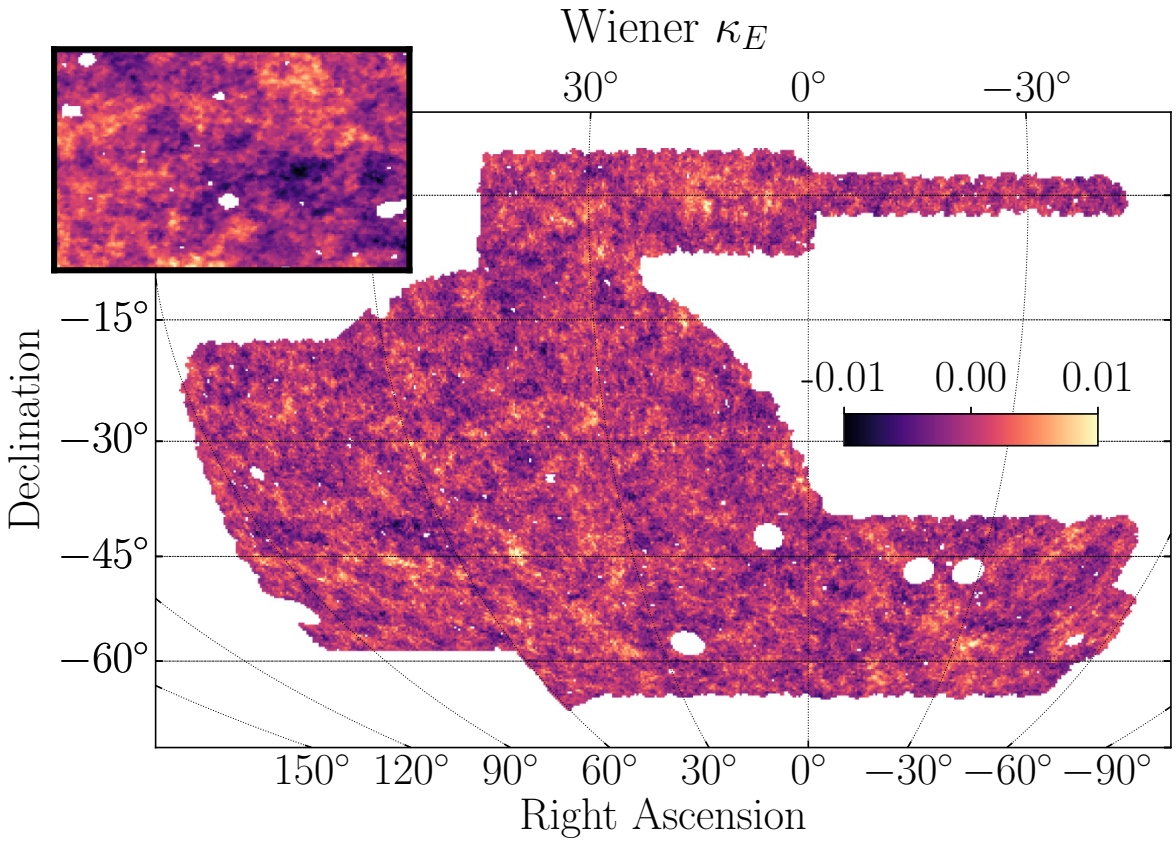


Figure 42: Mass map from the 3-year DES data. From [59], where four different versions of the mass map, from using 4 different map reconstruction methods, are presented (their Fig. 5). The one shown here is from their Wiener filter method. The quantity shown is the convergence κ_E. Since $|\kappa_E| \ll 1$, we are in the weak lensing regime, where the distortions of the galaxy images are small. The inset is an enlarged $15° \times 10°$ region centered at RA$= 70°$, Dec$= -40°$.

The Dark Energy Survey (DES) was a six-year survey that began on August 31, 2013 and ended taking data on January 9, 2019, having surveyed 5000 deg^2 of sky on the southern hemisphere and observed over 300 million galaxies [`https://www.darkenergysurvey.org/`]. It used a 570-megapixel digital camera on a 4-meter telescope in Chile. They have so far released their 3-year results. The 3-year data already covers the full 5000 deg^2 survey area. See Fig. 42 for a DES mass map.

12.2 Sloan Digital Sky Survey

The Sloan Digital Sky Survey (SDSS) is an ongoing galaxy survey that began in 2000. It uses a dedicated 2.5-meter wide-field telescope in New Mexico. Timewise the project has been divided into 5 stages, I (2000–2005), II (2005–2008), III(2008–2014), IV (2014–2020), and V, which began in 2020. Each stage had multiple data releases (DR). The final data release of SDSS-IV, DR17, was in 2021 [60]. The main sources for this section are the SDSS home page (`https://www.sdss.org/`) and Wikipedia (`https://en.wikipedia.org/wiki/Sloan_Digital_Sky_Survey`). SDSS has produced an enormous number of publications, which makes it difficult to obtain a good overview of SDSS.

At first (until 2009) SDSS operated both in photometric and spectroscopic mode. In photometric mode it took images in 5 different bandpasses (wavelength ranges) from near ultraviolet through optical to near infrared wavelengths. The 5 bandpasses are called u (near ultraviolet),

g (green), r (red), i, and z (both near infrared). DR8 (2009) is the final photometric release. The images cover 14 555 square degrees (35% of the full sky) and almost a billion objects. After that, SDSS has concentrated on spectroscopy. For cosmology, the main use of spectroscopy is to determine redshifts of galaxies, which can be used as proxy for their distances. Cosmological spectroscopic surveys are thus called *redshift surveys* and from them we get the 3D distribution of galaxies.

12.2.1 SDSS-I

SDSS-I took images of more than 8000 square degrees of sky and measured spectra of galaxies and quasars selected from 5700 square degrees of those images. The later stages contained several subsurveys, some of which were not relevant for cosmology, but focused on the Milky Way and nearby galaxies. We mention below only the cosmological subsurveys.

12.2.2 SDSS-II

Sloan Legacy Survey: This redshift survey contains spectra of over 800 000 galaxies and 100 000 quasars, covering over 7500 square degrees. (Most of the data is from SDSS-I.)

Sloan Supernova Survey: Focused on Type Ia supernova, which are standardizable candles and can thus be used to determine the redshift-distance relation. The final release (in 2014) contains 4607 confirmed or likely supernovae.

12.2.3 SDSS-III

Baryon Oscillation Spectroscopic Survey (BOSS): This redshift survey was aimed at measuring the baryon acoustic oscillation (BAO) scale, which is a statistical standard ruler, at different redshifts; to constrain the expansion history of the universe. It focused on luminous objects, Luminous Red Galaxies (LRG) and quasars, which can be detected at high redshifts.

12.2.4 SDSS-IV

Extended Baryon Oscillation Spectroscopic Survey (eBOSS): eBOSS was "the last use of the Sloan Telescope for galaxy redshift surveys designed to measure cosmological parameters using BAO and RSD techniques, with SDSS now focusing on other exciting astronomical questions" [61]. In [61] the galaxies and quasars selected from all the SDSS redshift surveys for BAO and RSD studies were grouped into 6 samples:

- **Main Galaxy Sample (MGS)**: 63 163 galaxies from SDSS-I and -II with $0.07 < z < 0.2$ and $M < -21.2$ from a contiguous 6813 deg^2 area of the sky, selected to probe the full redshift range roughly homogeneously.
- **BOSS DR12 galaxies:** 1 372 737 galaxies with $0.2 < z < 0.75$ from 9 376 deg^2 of sky.
- **eBOSS Luminous Red Galaxies (LRG):** 377 458 LRGs with $0.6 < z < 1.0$ from the same 9 376 deg^2 of sky.
- **eBOSS Emission Line Galaxies (ELG):** 173 736 ELGs with $0.6 < z < 1.1$ from 1 170 deg^2 of sky.
- **eBOSS Quasars:** 343 708 quasars with $0.8 < z < 2.2$ from 4 699 deg^2 of sky.
- **Lyman-α Forest Samples:** 210 005 quasars with $2.1 < z < 3.5$ from BOSS and eBOSS, whose spectra were suitable for analyzing fluctuations in neutral hydrogen along the line of sight to the quasar, using the Lyman-α absorption of the quasar light by this hydrogen.

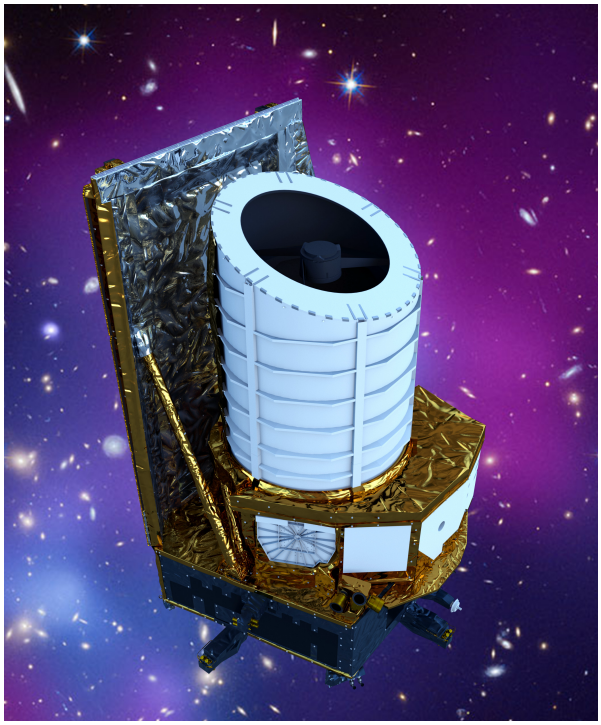


Figure 43: Left: Artist view of Euclid. Credit: ESA/ATG medialab (spacecraft); NASA, ESA, CXC, C. Ma, H. Ebeling and E. Barrett (University of Hawaii/IfA), et al. and STScI (background). Right: Euclid in a clean room of Thales Alenia Space in March 2023 before shipping to the launch site. Euclid is 4.7 m tall and 3.7 m in diameter and has a mass of 2 tonnes. Credit: Thales Alenia Space.

12.3 Euclid

Euclid is a wide-field European Space Agency (ESA) space telescope [https://www.euclid-ec.org/]. It is an ESA medium-class mission, M2 of the Cosmic Vision 2015–25 planning cycle for ESA's space science missions (see Table 6). The planned launch is in July 2023 from Cape Canaveral, Florida, with a SpaceX Falcon 9 rocket. Euclid will operate on an orbit around the second Lagrange point (L2) of the Sun–Earth system, 1.5 million km from Earth in the anti-Sun direction.

Euclid has a 1.2 m diameter on-axis 3-mirror Korsch telescope and two instruments, for visible and near infared light. The telescope has a dichroic plate that splits the visible light from the infrared, allowing both instruments to observe the same field of view simultaneously.

The visible light (VIS) instrument has a broad-band filter for wavelengths 550–900 nm, a 6×6 matrix of 4096×4132 pixel CCDs (Charge Coupled Devices), and 0.57 deg^2 field of view. The pixel size corresponds to 0.1×0.1 arcsec. [https://www.euclid-ec.org/?page_id=2485]. See Fig. 44.

The Near Infrared Spectrometer and Photometer (NISP) has both a photometric channel for taking images at three near infrared bands: Y (900–1192 nm), J (1192–1544 nm), and H (1544–2000 nm); and a slitless spectrometer with four grisms: 3 "red" (1250–1850 nm) and one "blue" (920–1250 nm). NISP has a 4×4 matrix of 2040×2040 infrared detectors (H2RGs, Hawaii-2RG integrated circuits, not CCDs) covering a field of view of 0.53 deg^2. The pixel size corresponds to 0.3×0.3 arcsec. "Slitless" means that the spectrometer will provide an image of each object in the field of view, spread out in a spectrum. This will mean that spectra of different objects will partially overlap, if they lie along the same line as the direction into which the spectrum is spread. The three different red "grisms" have different orientations, so that an overlap which occurs with one grism does not occur with another. [https://www.euclid-ec.org/?page_id=2490].

	Mission	launch	science
M1	Solar Orbiter	February 10, 2020	Sun
L1	JUICE	April 14, 2023	Jupiter's moons
M2	Euclid	July 2023	cosmology
M3	PLATO	2026	exoplanets
M4	ARIEL	2029	exoplanets
M5	EnVision	?	Venus orbiter
L2	ATHENA	2035	X-rays
L3	LISA	2037	gravitational waves

Table 6: Current ESA Science Programme: L and M missions. The ESA cost of a Large (L) mission is of the order of 1 billion euro, that of a medium (M) mission about half a billion. In addition there are Small (S) and Fast (F) missions, and Missions of Opportunity (missions already approved by some other space agency, where ESA may take a role). Some of the information is from `https://sci.esa.int/web/home/-/51459-missions`.

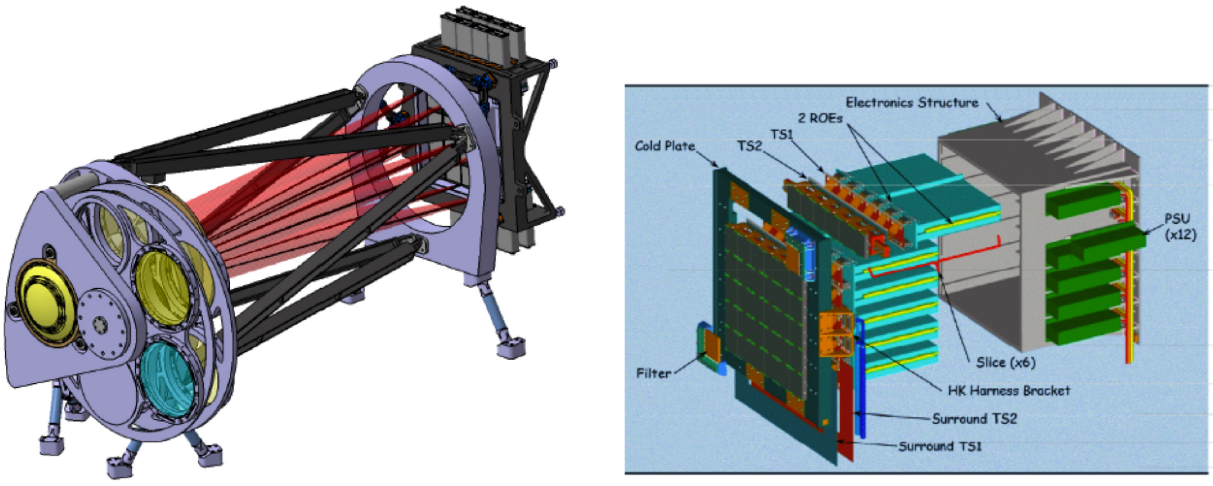


Figure 44: Left: NISP. Right: VIS. From [20].

The exposure time for each field of view will be divided between the three imaging bands and four grisms. See Fig. 44.

ESA provides the spacecraft and the telescope, has made a contract with SpaceX for the launch, and is responsible for the spacecraft operation and the transmission of data to Earth. The Euclid Consortium (EC) of 17 nations (Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, Netherlands, Norway, Portugal, Romania, Spain, Switzerland, United Kingdom, and United States) provides the NISP and VIS instruments and the Science Ground Segment of nine national Science Data Centers (SDC-CH, SDC-DE, SDC-ES, SDC-FI, SDC-FR, SDC-IT, SDC-NL, SDC-UK, and SDC-US), and is responsible for the analysis of Euclid data. The ESA cost of the Euclid mission is over 700 million euro, and the EC cost is about half of that, so this is an over-a-billion-euro project.

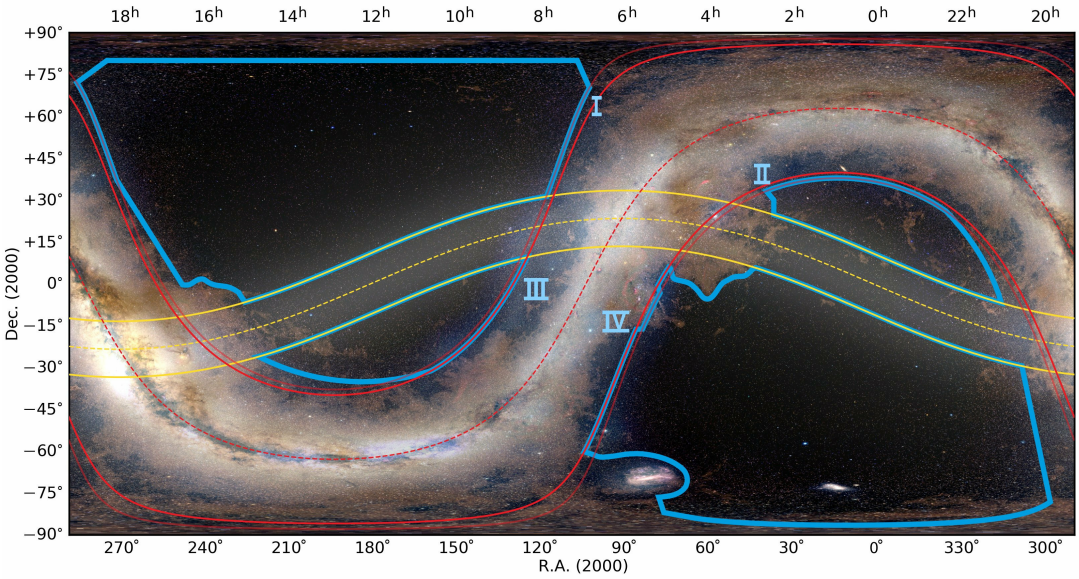


Figure 45: The Euclid Region of Interest. The Milky Way and the Zodiac divide it into four separate regions. From [62].

The main purpose of Euclid is to do the Euclid Wide Survey, covering about 15 000 deg^2, over a third of the sky. Euclid will operate in a step-and-stare mode, spending about 70 min for each field of view. To fill the gaps between the CCDs and detectors in the matrix, this time is divided into four slightly different , "dithered", pointings. The Wide Survey area is selected from a 17 000 deg^2 "region of interest" (RoI), avoiding the Milky Way plane and the Zodiac to allow a good view of the extragalactic universe, see Fig. 45.

The Euclid Wide Survey is both a redshift survey and a weak lensing survey. The redshift survey is based on the spectra from NISP, from which we expect accurate redshifts for 30 million galaxies. Only the red grism is used for the Wide Survey. The most important spectral line for the survey is H_α, the first line in the Balmer series, corresponding to the electron transition from energy level $n = 3$ to $n = 2$ in a hydrogen atom. The wavelength of H_α is 656 nm, so it falls in the red grism range (1250–1850 nm) for redshifts $z = 0.9$–1.8. For the weak lensing survey, accurate shapes of galaxies are determined from the sharp VIS images, while their photometric redshifts are estimated from the NISP images at the three infrared bands together with ground-based observations at five other bands. The weak lensing survey has a limiting magnitude of 24.5 and is expected to contain 1.5 billion galaxies.

Sending a telescope to L2 is much more expensive than observing from the ground. There are a number of reasons for going into space:

1. The observation conditions are much more stable in an orbit around L2 than on ground, which is important to get unbiased statistical measures.

2. The whole sky is visible from L2. Again, it is important for correct statistics that the same telescope and instrument is used for all of the survey.

3. The lack of atmosphere makes sharper and undistorted images possible, which is critical for shear (weak lensing) measurements.

4. The atmosphere absorbs infrared radiation. Although some of the infrared radiation gets through, the absorption is wavelength dependent, which distorts the statistics of a redshift survey and the infrared color needed for photometric redshifts.

These reasons are more critical for image shapes and infrared observations, and therefore Euclid is designed to do only that part of the survey. For visible wavelength color the advantage of going to space is not that large, and it is much more economical to make the visible band photometric observations on ground. In addition to the three near infrared bands from NISP, Euclid relies on ground observations (see the subsections on LSST and Northern Surveys below) of five visible bands for the determination of photometric redshifts.

Ground-based redshift surveys like SDSS have obtained extensive information on the large-scale structure at redshifts $z \lesssim 1$. It appears that the expansion of the universe began to accelerate near $z = 1$; to study the effect of dark energy, we need measurements from both sides of this redshift, up to $z \sim 2$. For $z \gtrsim 2$ the effect of dark energy is expected to be negligible. Thus the Euclid redshift survey focus on $z = 0.9$–1.8 is exactly where more information is the most needed.

Three Euclid Deep Fields (within the Wide Survey area) covering $40\,\mathrm{deg}^2$ will be observed to 2 magnitudes deeper than the Wide Survey (and also with the blue grism). In addition there will be calibration fields and auxiliary fields. These fields will be used for calibration of the Euclid instruments and analysis software and are thus needed for the quality of the Wide Survey, but they will also be used for additional science of their own, especially the Deep Fields.

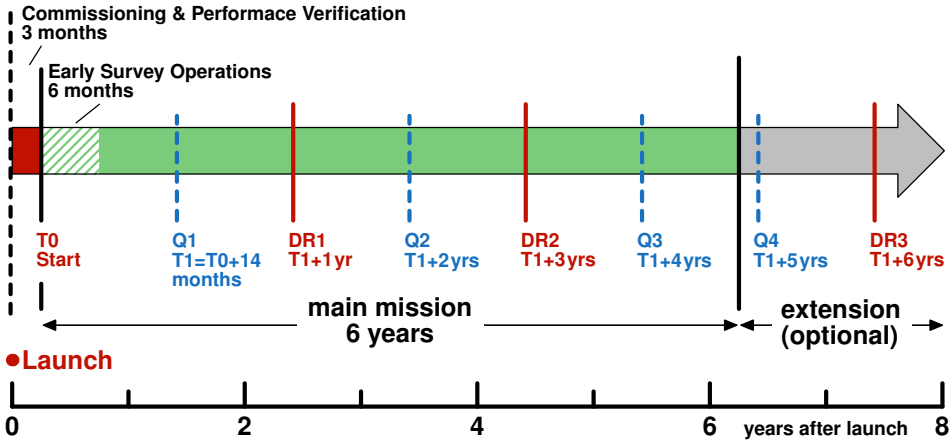


Figure 46: Euclid timeline. From [62].

It takes about a month for Euclid to reach the orbit around L2. The orbit is a Lissajous orbit[59] and it is unstable in the direction towards or away from Earth, so that an orbital correction is required once a month [62]. The survey begins 3 months after launch, after a

[59]There are two kinds of orbits in use around the Lagrangian points L1 and L2: Lissajous orbits and halo orbits. Halo orbits are periodic while Lissajous orbits are not.

one-month commissioning phase and a 2-month performance verification phase. There will be three data releases, DR1 for the first year of data (2500 deg^2), DR2 for three years (7500 deg^2) and DR3 for the full 6 years (15 000 deg^2) of data. After collecting the data, there is a 14 month proprietary period, during which only EC members (about 2000 scientists) have access to the data, before the release. In addition, there will be quick data releases of small areas of the sky. See Fig. 46.

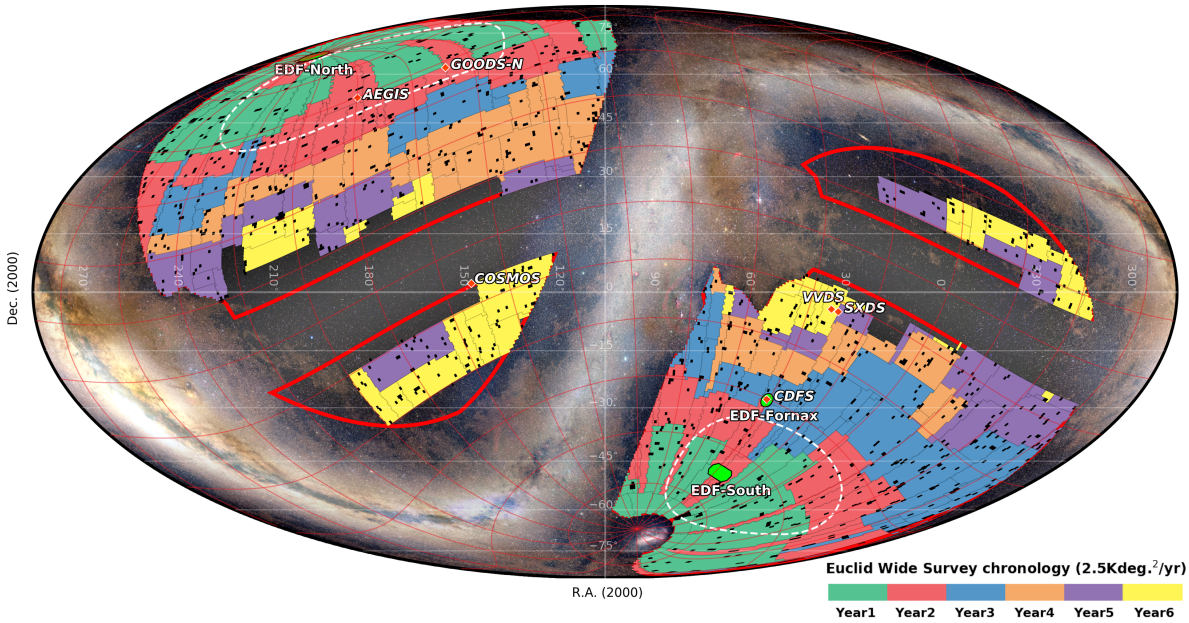


Figure 47: Euclid reference survey RSD 2021A. Celestial coordinates are shown in white and ecliptic coordinates in red. The different colors show the areas of the sky covered by the wide survey in each of the six years. The red lines outlines the RoI. The white dashed lines outline the galactic caps where the extragalactic visibility is the best. The reference survey covers 14 514 deg^2. The small holes in the survey are due to "blinding stars", which are so bright that they would degrade the quality of a large part of the field of view. There are 809 such holes totalling 635 deg^2. The 3 Euclid deep fields are shown in light green. The six auxiliary fields are marked with red diamonds. Calibration fields are not marked. From [62].

The sequence of pointings comprising the Euclid survey requires careful planning to make optimal use of the time available. There are are a number of constraints that need to be satisfied, e.g., the angle between the pointing direction and the Sun must not vary too much to keep the thermal conditions stable: it will be kept between 87° and 110°; thus at any given time of the year only a 23° wide strip of the sky is observable. The final survey sequence has not been yet decided, but Fig. 47 shows the Euclid reference survey RSD 2021A (RSD = reference survey definition) that satisfies the constraints and is a result of optimization. The regions furthest from the Milky Way, where the extragalactic visibility is the best, are observed in the first two years, and the two smaller "island" regions of the ROI are observed in the last two years of the 6-year survey.

Figure 48 shows a schematic view of the Euclid data analysis pipeline. Euclid data is transmitted from space to ground stations, which transmit the data to the Mission Operations Center (MOC) located at European Space Operations Centre (ESOC) in Darmstadt. From there the data is sent to the Science Operations Center (SOC), located at the European Space Astronomy Centre (ESAC) near Madrid, which performs the lowest level (Level-1) of the data analysis: telemetry checking and handling, real time assessment of "housekeeping " (status of spacecraft

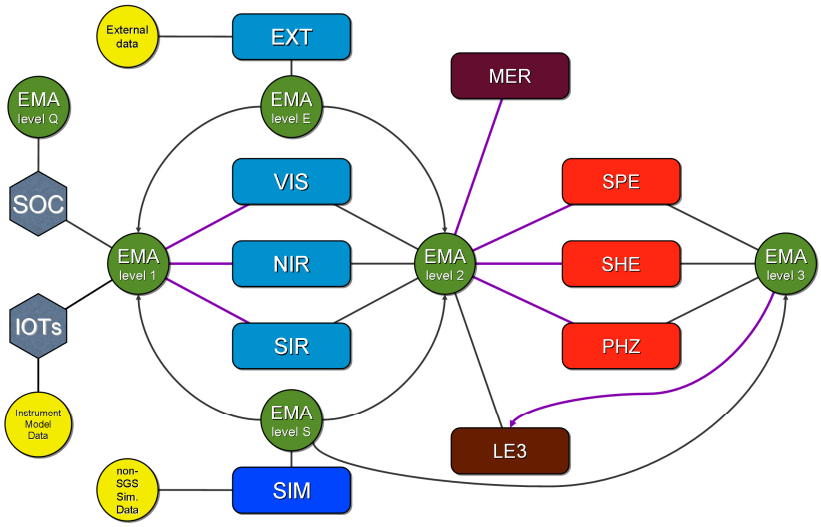


Figure 48: Schematic view of the Euclid data processing pipeline. See the main text for explanations. From [20].

and instruments) data, unpacking and decompression.

The 9 Euclid SDCs get the data from the SOC. IOTs are the Instrument Operation Teams that provide information about the instrument properties. The data products are distributed among the SDCs in the Euclid Mission Archive (EMA). VIS, NIR, and SIR refer to the reduction of the raw images from the instruments to provide calibrated images. The reduction contains such steps as the subtraction of "darks", i.e., what kind of output we get from the instrument when it receives no light, and scaling by "flats", i.e., images produced by the instrument when it is looking at uniform illumination, and removal of artifacts such as tracks of cosmic rays. EXT refers to processing external data received from ground-based surveys that has to be "Euclidized" to be in the same form as Euclid data. MER refers to merging the Euclid and external data to produce stacked images and source catalogs where the information from the different instruments and surveys is combined.

SPE refers to extraction of the redshifts from the spectra of the calibrated spectral images produced by SIR. SHE refers to determination of the shapes of galaxy images in the calibrated images produced by VIS. This step has to correct for the additional distortion produced by the optics of the telescope and requires knowledge of the *point spread function* (PSF), i.e., how does the image of a point source appear on the different parts of the VIS CCDs. PHZ determines the photometric redshifts estimated from the images at different wavelength bands from NIR and external data. SIM provides simulated data, which is needed for testing, validating, and qualifying the pipeline, and also for calibration and error estimation of the Euclid data products. The output of SPE, SHE, PHZ is in the form of final galaxy catalogs, containing their positions, magnitudes, redshifts and shapes.

LE3 refers to the last level, "Level-3", of the data analysis, where the final cosmology products, in the form of mass maps and various statistics, such as two- and three-point correlation functions, power spectra and bispectra, and their covariance matrices are calculated from these catalogs.

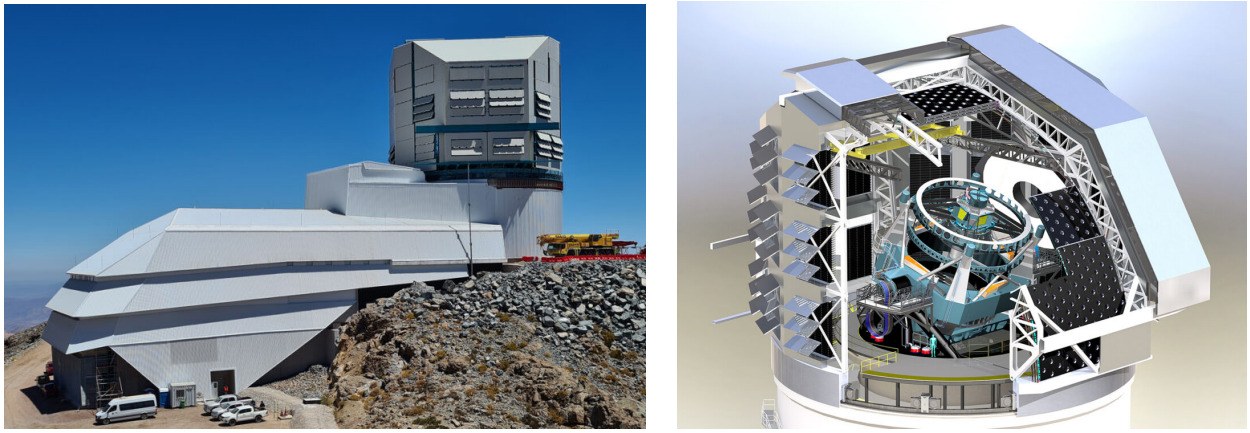


Figure 49: Vera C. Rubin Observatory on top of Cerro Pachón, 30°15′ S 70°45′ W, at 2647 m altitude, in Northern Chile. From `https://rubinobservatory.org/slideshows/eyes-on-asteroids`.

12.4 Ground-based surveys

12.4.1 LSST

In the planning stage, LSST stood for Large Synoptic Survey Telescope. The observatory has now been named Vera C. Rubin Observatory [63] and the telescope Simonyi Survey Telescope. See Fig. 49. The original acronym has been repurposed to stand for Legacy Survey of Space and Time, a ten-year survey using the 8.4 m telescope and 3200 megapixel camera (the largest digital camera ever constructed) of the observatory, with 6 different optical filters (u, g, r, i, z, and y (see Fig. 50), covering wavelengths 320–1050 nm) [64]. The telescope has been designed for repeated observation of the entire visible sky in a short time scale (synoptic = see all). The field of view is 9.6 deg^2 (so that pixel size is 0.2 arcsec) and the telescope can be repointed in less than five seconds. Its etendue, the product of light collecting area and field of view, $A\Omega = 319\,\mathrm{m^2 deg^2}$, will be significantly larger than any other optical telescope. The entire visible sky will be observed every 3–4 nights. This makes the survey excellent for spotting supernovae, asteroids, and other transient (changing in brightness or position) objects. Rubin Observatory is expected to discover 10–100 times more Solar System objects than known before and provide advance warning of objects that may be on collision course with Earth. First light is expected in August 2024.

According to the baseline survey strategy [65], 85% of observing time will be taken by the main survey (WFD = Wide, Fast, Deep), which covers 18 000 deg^2 uniformly. The declination range is $\delta = -65°$–$+5°$ (the total sky area in this range is 20 500 deg^2, but a region aligned with the Galactic Plane is not included in WFD). Two back-to-back 15-second exposures (this pair is called a *visit*) are taken of each field of view (two observations are needed to detect and remove tracks of cosmic rays from the images). On average, there will be 816 visits per night, and there will be 3026 observing nights in ten years. There are 2293 fields in the main survey. The same field of view will be observed every 3–4 nights, resulting in 910 visits (62 u, 88 g, 199 r, 201 i, 180 z, and 180 with the y filter) on average during the 10-year LSST. As often as possible there will be a pair of visits of the same field with 15–60 minute separation to help catch moving solar system bodies [64]. As the repeated observations are coadded, the view of the sky becomes gradually deeper, the limiting magnitude improving from (23.14, 24.47, 24.16, 23.40, 22.23, 21.57) of a single image to (25.4, 27.0, 27.1, 26.4, 25.2, 24.4) for the full stacked image after 10 years.[60]

[60] These numbers are from [65], and may yet change. The design specifications were an 18 000 deg^2 main survey area and 825 visits per field (56, 80, 184, 184, 160, and 160 per filter). On `https://www.lsst.org/scientists/keynumbers` the limiting magnitudes are given as (23.9, 25.0, 24.7, 24.0,

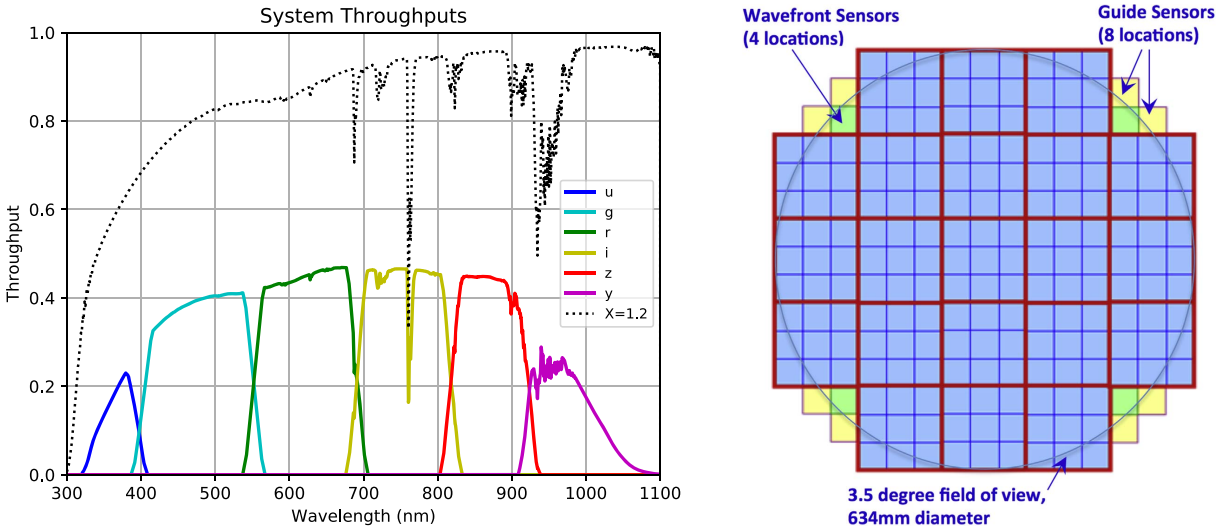


Figure 50: Left: LSST bandpasses. The dotted line is atmospheric transmission. Right: LSST Camera focal plane array. Each small square is a 4K×4K pixel sensor. From [64].

Approximately 10–20% of observing time will be spent on other projects such as "deep drilling fields" (single pointings observed in extended sequences). LSST is expected to observe 20 billion galaxies, millions of Type Ia supernovae [64], and 17 billion stars, and determine orbits of 6 million solar system bodies [`https://www.lsst.org/scientists/keynumbers`]. The Deep Drilling Fields will yield well-sampled light curves of tens of thousands of SNe to redshifts peaking around $z \sim 0.7$ and reaching beyond $z = 1$ [64].

23.3, 22.1) for a single exposure and (26.1, 27.4, 27.5, 26.8, 26.1, 24.9) after 10 years.

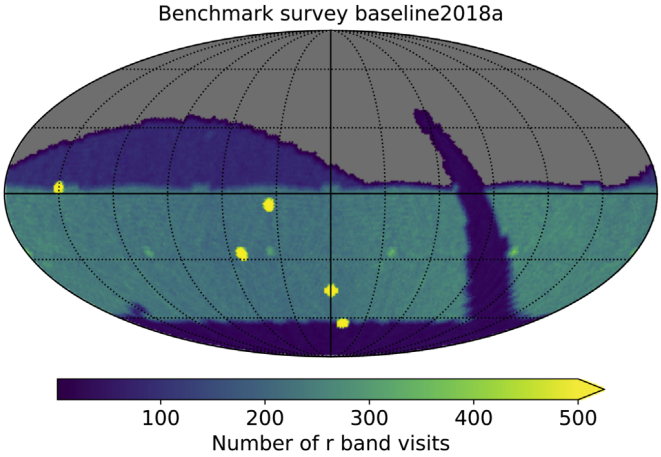


Figure 51: LSST baseline survey. The color scale shows number of visits using the r filter. For the main survey area this number is about 200. There are three surveys with lower number of visits to cover the rest of the southern ecliptic and equatorial hemispheres: The "northern ecliptic region" on the upper left is included to increase completeness of solar system objects. The arc on the right is the Galactic plane region that was excluded from the main survey since the high stellar density would lead to confusion of objects in such deep exposure. The yellow regions are Deep Drilling Fields, 2500–4500 r band visits. From [64].

12.4.2 Northern Surveys

For photometric redshift determination Euclid needs to supplement its three NIR imaging bands with images at visible light bands; especially the *griz* bands are critical [20, 62]. These can be more economically obtained by ground-based surveys. For the southern sky, Euclid can initially use data from DES and later from LSST. For the Northern sky there is no comparable survey with sufficiently many wavelength bands. Thus data from several different surveys will be used. Euclid is negotiating with LSST to extend it north up to $\delta = +30°$. For further north there are several surveys conducted from the northern hemisphere. Some of them cannot effectively observe (from Hawaii) further north than $\delta = 80°$, and therefore the Euclid survey will also use this as the northern limit. [62]

Three wide-field telescopes in Hawaii have joined forces to conduct the Ultraviolet Near Infrared Optical Northern Survey (UNIONS) [`https://www.skysurvey.cc`]: CFHT and Subaru Telescope on Maunakea (19°50′ N, 155°29′ W) and Pan-STARRS on Haleakala (20°42′ N, 156°15′ W). CFHT will provide the u and r bands, Pan-STARRS the i band, and Subaru the g and z bands; for the declination range $30° \leq \delta \leq 80°$ and more than 25° away from the galactic plane (an area of 4861 deg^2). In the u band LSST will not survey north of $\delta = 12°$ due to the high airmass resulting from the required low elevation angle of the telescope. Therefore UNIONS is aiming to extend the survey in u down to $\delta = 18°$ or 12°. An original motivation for UNIONS was to provide ground-based data for Euclid, but UNIONS is a separate entity from Euclid, and has also goals of its own.

Pan-STARRS (Panoramic Survey Telescope and Rapid Response System) [66] is like LSST for the north, although smaller, in that it continuously monitors the sky, to detect moving or variable objects. It uses two 1.8 m telescopes equipped with Gigapixel Cameras (1400 Mpix) at the Haleakala Observatory on Maui, Hawaii. It has been in operation since 2008 (then with just one telescope) and is mainly funded by the NASA Near Earth Object (NEO) observation program, and thus its main purpose is to search for potentially hazardous asteroids that may impact Earth in the future. In 2017 Pan-STARRS discovered the first interstellar object known to visit the solar system, 'Oumuamua. Euclid is counting on Pan-STARSS to provide the i-band north of $\delta = +30°$.

The Canada-France Imaging Survey (CFIS) [https://www.cfht.hawaii.edu/Science/CFIS/] is using the 3.6 m Canada-France-Hawaii Telescope (CFHT) on Maunakea with its 380 Mpix MegaCam camera to provide the u and r bands.

The Subaru telescope [https://subarutelescope.org/en/] is an 8.2 meter telescope on Maunakea, Hawaii, operated by the National Astronomical Observatory of Japan. It will use its 870 Mpix Hyper Suprime-Cam (HSC) with a 1.5° diameter field of view to conduct the WISHES (Wide Imaging with Subaru HSC of the Euclid Sky) survey to provide the z band. The z band is the most demanding of the bands to be obtained from the ground, so it requires an 8 m class telescope. The Waterloo Hawaii IfA G-band Survey (WHIGS) uses Canadian and University of Hawaii access by UNIONS co-leads to Subaru HSC to provide the r band.

The Javalambre Survey Telescope (JST250) [https://oajweb.cefca.es/telescopes/jst250] at 40°02′ N, 1°01′ W in Teruel, Spain, with a 2.55 m diameter mirror will use its Javalambre Panoramic Camera (JPCam) with a 4.3 deg^2 field of view to conduct the Javalambre-Euclid Deep Imaging Survey in g-band (JEDIS-g). Both Subaru and JEDIS are mentioned in connection of providing the g band in the north for Euclid. According to [62], the Subaru g-band is an effort launched in 2019 to complement the Spanish effort.

12.4.3 DESI

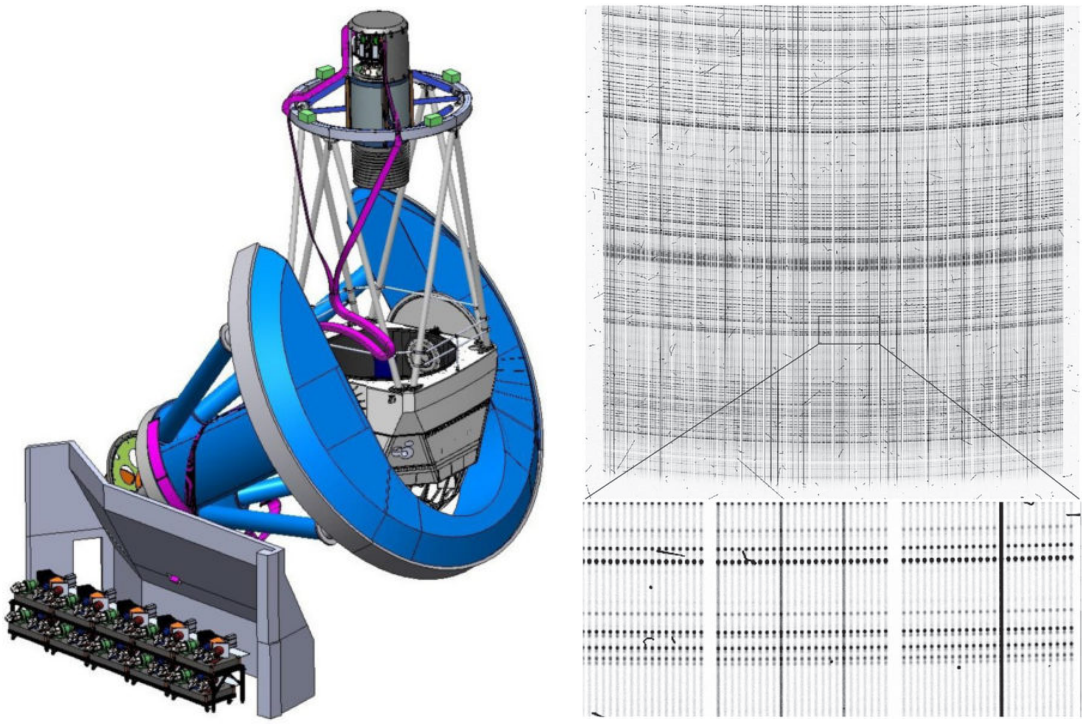


Figure 52: Left: The Mayall telescope and DESI. The focal plane with the fiber positioners is at the top. The fibers going to the spectrographs (bottom left) are shown in purple. Right: Example NIR CCD image from a DESI spectrograph showing 500 spectra as parallel vertical lines. The horizontal curved lines are sky lines from the atmosphere. Also cosmic ray tracks are visible. From [68].

The Dark Energy Spectroscopic Instrument (DESI) [67, 68] is a multiobject spectroscopic system installed at the Mayall 4 m telescope at Kitt Peak in Arizona. The field of view is 8.0 deg^2. The focal plane is divided into 10 petals, each equipped with 500 fiber positioning robots. For each exposure the robots position the fibers to look at the selected galaxies, whose spectra we want. Fiber bundles from the petals lead to 10 spectrographs. Each spectrograph has three arms, corresponding to blue (360–593 nm), red (560–772 nm) and near infrared (747–

980 nm), with a CCD (4096×4096 pixels for blue, 4114×4124 for red and NIR). The fibers are arranged to produce the 500 spectra as parallel lines on the CCDs. See Fig. 52. Thus DESI can take 5000 spectra simultaneously.

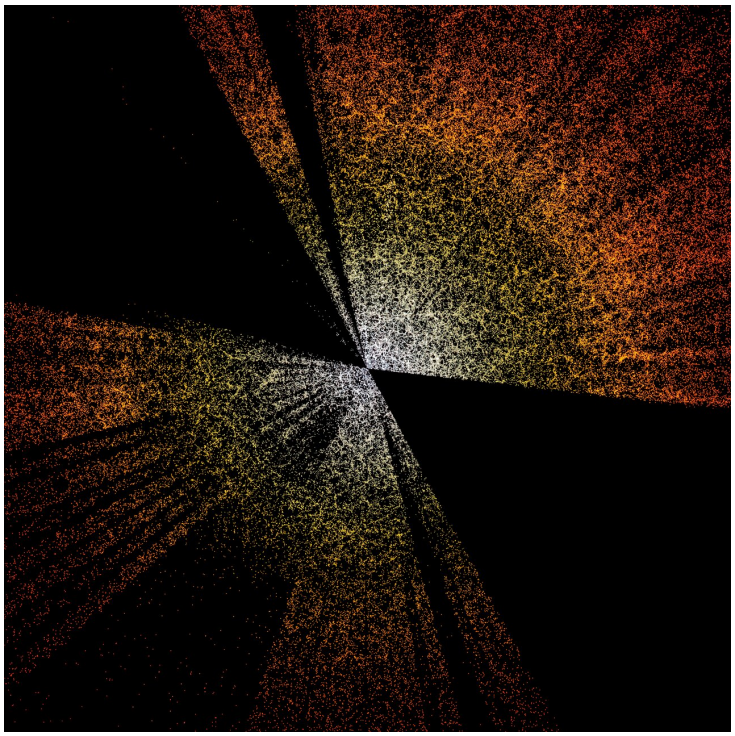


Figure 53: A slice through the universe containing 400 000 galaxies from the first few months of DESI operation. The furthest galaxies are 10 billion light years away. Credit: D. Schlegel/Berkeley Lab using data from DESI. Acknowledgment: M. Zamani (NSF's NOIRLab).

The DESI survey [`https://www.desi.lbl.gov/the-desi-survey/`] is a 5-year redshift survey to take spectra of 40 million pre-selected galaxies over one-third of the sky, 10 times as many as SDSS. The survey is based on four classes of galaxies:

1. bright galaxies out to $z = 0.4$ with limiting magnitude 20
2. luminous red galaxies (LRG) out to $z = 1.0$
3. emission line galaxies (ELG), which have strong emission lines DESI can pick out to $z = 1.6$
4. quasars, which DESI can detect out to $z = 3.5$ and beyond

The survey is a stage IV (see Sec. 12.6) dark energy survey based on measuring the BAO scale as a function of redshift.

The survey began with survey validation observations from December 2020 through June 2021, when the main survey began. The first data releases will be the Early Data Release from the survey validation and DR1 covering data until September 14, 2022 [69]. No data has been released yet (as of May 19, 2023), but a publicity image showing 400 000 galaxies was published in January 2022 [`https://noirlab.edu/public/news/noirlab2203/`], see Fig. 53.

12.5 Roman

The Decadal Survey of Astronomy and Astrophysics is conducted by the National Academy of Sciences in the United States every 10 years to make recommendations on what major projects

Figure 54: Roman Observatory. Credit: GSFC/SVS

should be invested on. The 2010 survey [70] identified as the top priority space project the Wide-Field Infrared Survey Telescope (WFIRST) to be sent to an orbit around L2. The original design had a 1.3 m diameter mirror but in 2012 the National Reconnaissance Office offered to donate to NASA two surplus telescopes they had, with 2.4 m diameter mirrors. It was decided to use one of them for WFIRST. See Fig. 54. The larger telescope meant launch mass (4166 kg) almost twice as large as Euclid (2160 kg) and therefore a geosynchronous orbit was considered instead of L2; this would also help accommodate the larger data rater the larger telescope made possible. In 2020 the mission was renamed Nancy Grace Roman Space Telescope. It has now been decided to send it to a halo orbit around L2 using the SpaceX Falcon Heavy launcher. The mission was delayed by the delays and cost overruns of the James Webb Space Telescope (which was the top space priority of the previous Decadal Survey) as NASA did not have sufficient funding for both simultaneosly. With the JWST launch in December 2021, Roman has then been progressing swiftly and the launch is expected in late 2026. The planned mission duration is 5 years, with a possible 5-year extension. [71].

Roman will have two instruments: the Wide Field Instrument (WFI), and the Coronagraph Instrument. WFI is the primary instrument, and will be used for a galaxy survey covering a billion galaxies. It will also perform a microlensing survey of the inner Milky Way, and is expected to find 2 600 exoplanets. WFI has a 300-megapixel camera and a $0.281\,\mathrm{deg}^2$ ($0.8 \times 0.4\,\mathrm{deg}$) field of view. It has 8 different filters, a grism and a prism [72]. The Coronagraph Instrument is for taking images and spectra of nearby exoplanets.

Roman will perform a number of different surveys. There will be three "core community surveys" using WFI: 1) 24 months of observing time will be devoted to the High Latitude Wide Area Survey, covering $2000\,\mathrm{deg}^2$ of extragalactic sky, taking images at different bands and spectra. 2) 6 months for the High Latitude Time Domain survey, covering 5–$20\,\mathrm{deg}^2$, revisited with a 5-day cadence, the goal being to measure light curves of Type Ia supernovae. 3) 13 months for Galactic Bulge Time Domain survey, covering $2\,\mathrm{deg}^2$ near the galactic center to find exoplanets. In addition, 3 months is reserved for Coronagraph Instrument Tech Demo Observations of exoplanets. The remaining 15 months of the nominal 5-year mission is reserved for "General Astrophysics Surveys" for which astronomers can make proposals. All Roman observations will be public with no proprietary period.

The High Latitude Wide Area Survey is similar to the Euclid survey, but complementary to it, as it covers a much smaller region of sky (at least $1700\,\mathrm{deg}^2$) but to greater depth. The imaging part covers the NIR range 930–2000 nm with four different bands (927–1192 nm, 1131–1454 nm, 1380–1774 nm, and 1683–2000 nm), leading to weak lensing shape measurements of hundreds of millions of galaxies. The spectroscopic part covers the same region of sky with slitless grism spectroscopy at 1000–1930 nm. The primary target spectral lines are H_α (656.3 nm) for redshifts $0.53 < z < 1.88$ and O-III (500.7 nm) up to $z = 2.77$. (The survey is not yet fully defined; the

above is a possible survey design).

12.6 Dark Energy Survey Stages

In 2006 the Dark Energy Task Force (DETF) [73] defined different stages (I–IV) of observational projects according to how accurately they can determine the dark energy equation of state. They defined the DETF Figure of Merit (FoM) as the reciprocal of the area of the error ellipse enclosing the 95% confidence limit in the w_0, w_a plane, assuming the two-parameter equation of state

$$w(a) = w_0 + (1 - a)w_a \,. \tag{12.1}$$

So roughly FoM $\propto 1/(\Delta w_0 \Delta w_a)$, where Δw_0 and Δw_a are the uncertainties of the two parameter values. Stage I corresponded to what was known in 2006, Stage II represent projects ongoing then, and Stages III-IV referred to future projects, Stage IV defined as projects that could increase FoM by an order of magnitude from Stage II, Stage III providing a more modest improvement.

DETF identified four main observational techniques:

1. Measurements of the Baryon Acoustic Oscillation (BAO) scale in galaxy redshift surveys
2. Galaxy Cluster (CL) surveys measuring the density and distribution of galaxy clusters
3. Supernova (SN) surveys using type Ia supernovae to determine the luminosity distance vs. redshift relation
4. Weak Lensing (WL) surveys, which measure the angular-diameter distance vs. redshift relation and the growth rate of structure

DETF found that no single technique is sufficient alone, so that a combination of methods and surveys is needed.

The situation in 2023 is that the observations of Stage III projects have been mostly completed, although all results have not yet been published, whereas Stage IV projects are starting, see Table 7. The two Stage III projects in the table that we have not discussed are the Subaru Measurements of Images and Redshifts (SuMIRe) and Hobby-Eberly Telescope Dark Energy Experiment (HETDEX) [`https://hetdex.org/`].

project	dates	area [deg^2]	type of data	z	method
BOSS	2008–2014	10 000	opt-S	0.3–0.7	BAO/RSD
				2–3.5 (Ly-α)	
KiDS	2011–2010	1350	opt-I		WL/CL
DES	2013–2019	5000	opt-I		WL/CL
					SN/BAO
eBOSS	2014–2018	7500	opt-S	0.6–2.0	BAO/RSD
				2–3.5 (Ly-α)	
SuMIRE	2014–2024	1500	opt-I		WL/CL
			opt/NIR-S	0.8–2.4	BAO/RSD
HETDEX	2017–2023	450	opt-S	1.9–3.5	BAO/RSD
DESI	2021–2026	14 000	opt-S	0–1.7	BAO/RSD
LSST	2025–2035	20 000	opt-I		WL/CL
					SN/BAO
Euclid	2023–2029	15 000	opt-I		WL/CL
			NIR-S	0.7–2.2	BAO/RSD
Roman	2026–2031	2200	NIR-I		WL/CL/SN
			NIR-S	1.0–3.0	BAO/RSD

Table 7: Stage III (above dividing line) and Stage IV (below) surveys. Dates refer to data taking; not all data or results from Stage III surveys has been yet (May 2023) published. opt = visible light, NIR = near infared, S = spectra, I = images. The redshift range (z) is for galaxies, except for eBOSS the upper part of the range $z = 0.6$–2.0 is quasars, and Ly-α indicates structure determined from absorption features in quasar spectra. RSD = redshift space distortions. Copied from the presentation by Josh Frieman (U. Chicago, Fermilab) in the UCLA Dark Matter Conference, March 29, 2023. [74]

12.7 Etendue

Usually the measure of the power of a telescope is taken to be its collecting area A, which is related to (the square of) the diameter of the main mirror or lens. For surveys, equally important is the solid angle Ω of the field of view (FoV). The product of these two measures $A\Omega$ is called the *etendue*. If the sky were uniformly bright, the amount of light collected by the telescope per unit time would be proportional to the etendue. The time it takes to survey a given area of the sky to a given limiting magnitude is inversely proportional to the etendue. From Table 8 we see that to do the Euclid 6-year survey would take the Hubble Space Telescope over 200 years.

telescope (instrument)	diameter [m]	A [m^2]	FoV diam.	Ω [deg^2]	$A\Omega$ [$m^2 deg^2$]
Palomar Hale (WaSP)	5.1	~ 20		0.095	~ 2
Palomar Schmidt (ZTF)	1.26			47	~ 50
Pan-STARRS (1 telescope)	1.8		3°		
JST 250 (JPCAM)	2.55	3.89	3°	4.3	26.5
Subaru (HSC)	8.2		1.5°		
Rubin	8	33		9.6	319
Hubble (ACS/WFC)	2.4	4.0		0.0031	0.0126
Webb (NIRCam)	6.5	25.4		0.0027	0.068
Euclid	1.2	1.006		0.56	0.56
Roman (WFI)	2.4	~ 4		0.281	~ 1.1

Table 8: The etendue of various telescopes (and instruments). We list first ground-based telescopes and then space telescopes. These are further grouped into non-survey and survey telescopes. A survey telescope is characterized by a large etendue. Some of the information was difficult to find, and the different telescopes/instruments may not have been treated consistently (instrument FoV is not necessarily the same as telescope FoV). Often just the aperture diameter is given, not the collecting area. Non-survey telescopes usually have many instruments with different fields of view, none of which necessarily maximizes the field of view possible with the telescope. Modern ground-based non-survey telescopes are not included because it was difficult to find a representative field of view for them.

References

[1] J.A. Peacock: Cosmological Physics (Cambridge University Press 1999), Chapter 16

[2] H. Mo, F. van den Bosch, and S. White: Galaxy Formation and Evolution (Cambridge University Press 2010)

[3] P.J.E. Peebles: The Large-Scale Structure of the Universe (Princeton University Press 1980)

[4] C.M. Baugh, *The real-space correlation function measured from the APM Galaxy Survey*, Mon. Not. R. Astron. Soc. **280**, 267 (1996), astro-ph/9512011

[5] A.J. Benson, R.G. Bower, C.S. Frenk, C.G. Lacey, C.M. Baugh, and S. Cole, *What shapes the luminosity function of galaxies?*, Astrophys. J. **599**, 38 (2003), astro-ph/0302450

[6] A.G. S'anchez et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological implications of the large-scale two-point correlation function*, Mon. Not. R. Astron. Soc. **425**, 415 (2012), arXiv:1203.6616

[7] A.R. Liddle and D.H. Lyth: Cosmological Inflation and Large-Scale Structure (Cambridge University Press 2000)

[8] G.B. Arfken and H.J. Weber: Mathematical Methods for Physicists, 6th edition (Elsevier 2005)

[9] I.S. Gradshteyn and I.M. Ryzhik: Table of Integrals, Series, and Products (Academic Press 1980)

[10] E. Hawkins et al., *The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Universe*, Mon. Not. R. Astron. Soc. **346**, 78 (2003),astro-ph/0212375

[11] S. Dodelson: Modern Cosmology (Academic Press 2003), Chapter 7

[12] J. Richard Gott III et al., *A Map of the Universe*, Astrophys. J. **624**, 463 (2005), astro-ph/0310571

[13] M. Tegmark et al., *Cosmological Constraints from the SDSS Luminous Red Galaxies*, Phys. Rev. **D74**, 123507 (2006), astro-ph/0608632

[14] I. Szapudi, *A new method for calculating counts in cells*, Astrophys. J. **497**, 16 (1998)

[15] J.V. Wall and C.R. Jenkins: Practical Statistics for Astronomers (Cambridge University Press 2012), Chapter 10

[16] T.J. Broadhurst, R.S. Ellis, D.C. Koo, and A.S. Szalay, *Large-scale distribution of galaxies at the galactic poles*, Nature **343**, 726 (1990).

[17] H. Kurki-Suonio, G.J. Mathews, and G.M. Fuller, *Deviation from Periodicity in the Large-Scale Distribution of Galaxies*, Astrophys. J. Lett. **356**, L5-L7 (1990).

[18] C.J. Conselice, A. Wilkinson, K. Duncan, A. Mortlock, *The evolution of galaxy number density at $z < 8$ and its implications*, draft version October 9, 2016, linked from `http://hubblesite.org/news_release/news/2016-39`

[19] Planck Collaboration, Astronomy & Astrophysics **594**, A13 (2016), arXiv:1502.01589

[20] Euclid Definition Study Report, ESA/SRE(2011)12, arXiv:1110.3193

[21] C.D. Shane and C.A. Wirtanen, Publ. Lick Obs. XXII, Pt. 1 (as cited by [22]

[22] M. Seldner, B. Siebers, E.J. Groth, P.J.E. Peebles, *New reduction of the Lick catalog of galaxies*, Astronomical Journal, **82**, 249 (1977)

[23] R.M. Soneira, P.J.E. Peebles, *A computer model universe: Simulation of the nature of the galaxy distribution in the Lick catalog*, Astronomical Journal **83**, 845 (1978)

[24] H. Karttunen, P. Kröger, H. Oja, M. Poutanen, and K.J. Donner: Fundamental Astronomy (Springer 1987)

[25] A.D. Montero-Dorta and F. Prada, *The SDSS DR6 luminosity functions of galaxies*, Mon. Not. R. Astron. Soc. **399**, 1106 (2009), arXiv:0806.4930

[26] D.A. Varshalovich, A.N. Moskalev, V.K. Khersonskii: Quantum Theory of Angular Momentum (World Scientific 1988)

[27] P. Vandergheynst and Y. Wiaux, in B. Forster, P. Massopust (ed.): Four Short Courses on Harmonic Analysis (Birkhäuser/Springer 2009), p.135

[28] E.W. Kolb and M.S. Turner: The Early Universe (Addison-Wesley 1990)

[29] C. Lacey and S. Cole, *Merger rates in hierarchical models of galaxy formation*, MNRAS **262**, 627 (1993), Appenix A: Critical overdensity for collapse for $\Omega_0 < 1$.

[30] T.T. Nakamura and Y. Sato, *Strong gravitational lensing and velocity function as tools to probe cosmological parameters*, Progress of Theoretical Physics **97**, 49 (1997), Appendix C: Spherical collapse in $\Omega + \lambda = 1$ universe

[31] D.H. Lyth and A. R. Liddle: The Primordial Density Perturbation (Cambridge University Press 2009)

[32] J.A. Peacock et al., *A measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey*, Nature **410**, 169 (2001)

[33] J.E. Bautista et al., *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic correlation function between redshifts 0.6 and 1*, MNRAS **500**, 736 (2021), arXiv:2007.08993

[34] Planck Collaboration, *Planck 2018 results. VI. Cosmological parameters*, submitted to Astronomy & Astrophysics (2018), arXiv:1807.06209v1

[35] S. Alam et al., *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample*, MNRAS **470**, 2617 (2017)

[36] C. Alcock and B. Paczyński, *An evolution free test for non-zero cosmological constant*, Nature **281**, 358 (1979)

[37] M. Davis and P.J.E. Peebles, *A Survey of galaxy redshifts. V. The two-point position and velocity correlations*, ApJ **267**, 465 (1983)

[38] A.J.S. Hamilton, *Toward better ways to measure the galaxy correlation function*, ApJ **417**, 19 (1993)

[39] S.D. Landy and A.S. Szalay, *Bias and variance of angular correlation functions*, ApJ **412**, 64 (1993)

[40] E. Keihanen et al., *Estimating the galaxy two-point correlation function using a split random catalog*, Astronomy & Astrophysics **631**, A73 (2019), arXiv:1905.01133

[41] H.A. Feldman, N. Kaiser, and J.A. Peacock, *Power-spectrum analysis of three-dimensional redshift surveys*, ApJ **426**, 23 (1994)

[42] A.J. Ross et al., *The clustering of galaxies in the completed SDSS-II Baryon Oscillation Spectroscopic Survey: observational systematics and baryon acoustic oscillations in the correlation function*, MNRAS **464**, 1168 (2017), arXiv:1607.03415

[43] I. Szapudi, *Three-point statistics from a new perspective*, astro-ph/0404476

[44] I. Szapudy and A.S. Szalay, *A new class of estimators for the N-point correlations*, Astrophys. J. **494**, L41 (1998)

[45] Z. Slepian and D.J. Eisenstein, *Computing the three-point correlation function of galaxies in $O(N^2)$ time*, MNRAS **452**, 4142 (2015)

[46] Z. Slepian and D.J. Eisenstein, *A practical computational method for the anisotropic redshift-space three-point correlation function*, MNRAS **478**, 1468 (2018)

[47] Z. Slepian et al., *The large-scale three-point correlation function of the SDSS BOSS DR12 CMASS galaxies*, MNRAS **468**, 1070 (2017)

[48] Z. Slepian et al., *Detection of baryon acoustic oscillation features in the large-scale three-point correlation function of SDSS BOSS DR12 CMASS galaxies*, MNRAS **469**, 1738 (2017)

[49] H. Gil-Marín et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: RSD measurement from the power spectrum and bispectrum of the DR12 BOSS galaxies*, MNRAS **465**, 1757 (2017)

[50] S. Okamura, *A historical overview of galaxy surveys*, in Proceedings of the IAU Symposium No. 341, 2018, https://doi.org/10.1017/S1743921319004800

[51] S. J. Maddox, G. Efstathiou, W. J. Sutherland, and J. Loveday, *Galaxy correlations on large scales*, MNRAS **242**, 43P (1990)

[52] S. J. Maddox, G. Efstathiou, and W. J. Sutherland, *The APM Galaxy Survey - III. An analysis of systematic errors in the angular correlation function and cosmological implications*, MNRAS **283**, 1227 (1996)

[53] V. de Lapparent, M.J. Geller, and J.P. Huchra, *A slice of the Universe*, Astrophys. J. **302**, L1 (1986)

[54] S.A. Shectman et al., *The Las Campanas redshift survey*, Astrophys. J. **470**, 172 (1996)

[55] M. Colless et al., *The 2dF Galaxy Redshift Survey: spectra and redshifts*, MNRAS **328** 1039 (2001)

[56] C. Heymans et al., *CFHTLens: the Canada-France-Hawaii Telescope Lensing Survey*, arXiv:1210.0032, MNRAS **427**, 146 (2012)

[57] T. Erben et al., *CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey - imaging data and catalogue products*, arXiv:1210.8156, MNRAS **433**, 2545 (2013)

[58] M. Kilbinger et al., *CFHTLenS: combined probe cosmological model comparison using 2D weak gravitational lensing*, MNRAS **430**, 2200 (2013)

[59] N. Jeffrey et al., *Dark Energy Survey Year 3 results: Curved-sky weak lensing mass map reconstruction*, arXiv:2105.13539, MNRAS **505**, 4626 (2021)

[60] Abdurro'uf et al., *The Seventeenth Data Release of the Sloan Digital Sky Surveys: Complete Release of MaNGA, MaStar, and APOGEE-2 Data*, arXiv:2112.02026, Astrophys. J. Supp. S. **259**, 35 (2022)

[61] S. Alam et al. (eBOSS), *The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory*, arXiv:2007.08991

[62] Euclid Collaboration, R. Scaramella et al., *Euclid preparation I. The Euclid Wide Survey*, Astronomy & Astrophysics **662**, A112 (2022)

[63] https://rubinobservatory.org/,
https://en.wikipedia.org/wiki/Vera_C._Rubin_Observatory

[64] Ivezić et al., *LSST: From Science Drivers to Reference Design and Anticipated Data Products*, Astrophys. J. **873**, 111 (2019)

[65] LSST Science Collaboration, P. Marshall et al., *Science-Driven Optimization of the LSST Observing Strategy*, arXiv.1708.04058

[66] http://pswww.ifa.hawaii.edu/pswww/ , http://www2.ifa.hawaii.edu/research/Pan-STARRS.shtml

[67] DESI Collaboration, *Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument*, Astron. J. **164**, 207 (2022) and https://www.desi.lbl.gov/

[68] J. Guy et al., *The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument*, Astron. J. **165**, 144 (2023)

[69] https://data.desi.lbl.gov/doc/releases/

[70] National Research Council: New Worlds, New Horizons in Astronomy and Astrophysics (The National Academies Press 2010) https://doi.org/10.17226/12951

[71] https://roman.gsfc.nasa.gov/, https://en.wikipedia.org/wiki/Nancy_Grace_Roman_Space_Telescope

[72] https://roman.gsfc.nasa.gov/science/WFI_technical.html

[73] A. Albrecht et al. *Report of the Dark Energy Task Force*, astro-ph/0609591

[74] UCLA Dark Matter March 29 – April 1, 2023, https://conferences.pa.ucla.edu/dark-matter-2023/, J. Frieman presentation: Dark Energy https://indico.cern.ch/event/1188759/contributions/5043999/attachments/2619045/4529758/DarkEnergy-UCLA%20(1).pdf