Distribution function approach to redshift space distortions

Seljak, U ; McDonald, P

Abstract: We develop a phase space distribution function approach to redshift space distortions (RSD), in which the redshift space density can be written as a sum over velocity moments of the distribution function. These moments are density weighted and have well defined physical interpretation: their lowest orders are density, momentum density, and stress energy density. The series expansion is convergent if k u/aH < 1, where k is the wavevector, H the Hubble parameter, u the typical gravitational velocity and = cos , with being the angle between the Fourier mode and the line of sight. We perform an expansion of these velocity moments into helicity modes, which are eigenmodes under rotation around the axis of Fourier mode direction, generalizing the scalar, vector, tensor decomposition of perturbations to an arbitrary order. We show that only equal helicity moments correlate and derive the angular dependence of the individual contributions to the redshift space power spectrum. We show that the dominant term of 2 dependence on large scales is the cross-correlation between the density and scalar part of momentum density, which can be related to the time derivative of the matter power spectrum. Additional terms contributing to 2 and dominating on small scales are the vector part of momentum density-momentum density correlations, the energy density-density correlations, and the scalar part of anisotropic stress density-density correlations. The second term is what is usually associated with the small scale Fingers-of-God damping and always suppresses power, but the first term comes with the opposite sign and always adds power. Similarly, we identify 7 terms contributing to 4 dependence. Some of the advantages of the distribution function approach are that the series expansion converges on large scales and remains valid in multi-stream situations. We finish with a brief discussion of implications for RSD in galaxies relative to dark matter, highlighting the issue of scale dependent bias of velocity moments correlators.

Distribution function approach to redshift space distortions

Uroš Seljak,^{a,b,c,d} **Patrick McDonald,**^{b,e}

^{a}Institute of Theoretical Physics, University of Zurich, 8057 Zurich, Switzerland

^{b}Lawrence Berkeley National Laboratory, University of California, Berkeley, California 94720, USA

^{c}Department of Physics, Department of Astronomy, University of California, Berkeley, California 94720, USA

^{d}Institute for the Early Universe, Ewha Womans University, Seoul 120-750, S. Korea

^{e}Physics Dept., Brookhaven National Laboratory, Building 510A, Upton, NY 11973-5000, USA

E-mail: useljak@berkeley.edu, pvmcdonald@lbl.gov

Abstract. We develop a phase space distribution function approach to redshift space distortions (RSD), in which the redshift space density can be written as a sum over velocity moments of the distribution function. These moments are density weighted and have well defined physical interpretation: their lowest orders are density, momentum density, and stress energy density. The series expansion is convergent if $k\mu u/aH < 1$, where k is the wavevector, H the Hubble parameter, u the typical gravitational velocity and $\mu = \cos\theta$, with θ being the angle between the Fourier mode and the line of sight. We perform an expansion of these velocity moments into helicity modes, which are eigenmodes under rotation around the axis of Fourier mode direction, generalizing the scalar, vector, tensor decomposition of perturbations to an arbitrary order. We show that only equal helicity moments correlate and derive the angular dependence of the individual contributions to the redshift space power spectrum. We show that the dominant term of μ^2 dependence on large scales is the cross-correlation between the density and scalar part of momentum density, which can be related to the time derivative of the matter power spectrum. Additional terms contributing to μ^2 and dominating on small scales are the vector part of momentum density-momentum density correlations, the energy density-density correlations, and the scalar part of anisotropic stress density-density correlations. The second term is what is usually associated with the small scale Fingers-of-God damping and always suppresses power, but the first term comes with the opposite sign and always adds power. Similarly, we identify 7 terms contributing to μ^4 dependence. Some of the advantages of the distribution function approach are that the series expansion converges on large scales and remains valid in multi-stream situations. We finish with a brief discussion of implications for RSD in galaxies relative to dark matter, highlighting the issue of scale dependent bias of velocity moments correlators.

Contents

1 Introduction

Galaxy clustering has traditionally been one of the most important ways to extract cosmological information. Galaxies are not a faithful tracer of dark matter, as their clustering strength is biased relative to the dark matter. However, they are expected to follow the same gravitational potential as the dark matter and hence have the same velocities. This is not observable through angular clustering, which is only sensitive to correlations transverse to the line of sight. It is however detectable in redshift surveys, because the redshift of the galaxy does not provide information only on the radial distance, but also on the radial velocity through the Doppler shift. This induces anisotropies in the clustering, which are generically called redshift space distortions (RSD) [1]. They provide an opportunity to extract information on the dark matter clustering directly. On large scales clustering of galaxies along the line of sight is enhanced relative to the transverse direction due to peculiar motions and this allows one to determine the ratio of logarithmic rate of growth f to bias b [2]. Combining the statistics from different lines of sight one can eliminate the unknown bias and measure directly the logarithmic rate of growth times the amplitude.

It has been argued that using RSD information could greatly increase our knowledge of cosmological models, including tests of dark energy and general relativity [3–5]. Galaxy clustering has clear advantages over the alternatives such as weak lensing: it is intrinsically 3-dimensional, thus providing better statistics, and it has high signal to noise. While most of the predictions in the literature are model dependent, a generic statement can be made that if systematic effects were perfectly understood RSD would be one of the most powerful techniques for such studies. The main problem with RSD is that nonlinear velocity effects extend to rather large scales and give rise to a scale dependent and angular dependent clustering signal. It is easy to see these effects in any real redshift survey: one sees elongated features along the line of sight, called the fingers-of-god (FoG) effect, which are caused by random velocities inside virialized objects such as clusters, which scatter galaxies along the radial direction in redshift space, even if they have a localized spatial position in real space. This is just an extreme example and other related effects, such as nonlinear infall streaming motions, also cause nonlinear corrections. This means that one needs to understand these and separate them from the nonlinear evolution of the dark matter and from the nonlinear relation between the galaxies and the dark matter, both of which also give rise to a scale dependent bias [6].

Several recent studies have investigated these nonlinear effects [7–13], some limiting the analysis to dark matter only and some also including galaxies or halos. The common denominator of these studies is that they are based on various ansatzes for the scale and angular dependence of RSD,

typically combined with some perturbation theory analysis. This has the advantage of having just a few free parameters, so that if the ansatz is accurate one can model the effects accurately. The reverse is also true and the problem is that it is difficult to make general statements regarding the range of validity for any given model. There is another problem connected to perturbation theory, in that the usual perturbation theory makes a single stream approximation, which we know breaks down on small scales inside the virialized halos (indeed, FoG are a manifestation of multi-streaming on small scales).

In this paper we present a different approach to RSD: we use a distribution function approach to show that one can make a series expansion of RSD, which is convergent on sufficiently large scales and we derive the most general form of RSD correlator allowed by the symmetries. In this paper we present the formal derivations and conceptual implications, reserving all the applications to future work. The structure of this paper is as follows: in section 2 we develop the distribution function approach to RSD and derive the helicity decomposition. In section 3 we discuss the power spectra, and use rotation symmetries to derive the most general form of the RSD correlator. We also discuss the lowest order contributions and connect them to physical quantities such as density, momentum, stress energy tensor etc. This is followed by a discussion in section 4.

2 Redshift-space distortions from the distribution function

The exact evolution of collisionless particles is described by the Vlasov equation [14]. Following the discussion by [15], we start from the distribution function of particles $f(\mathbf{x}, \mathbf{q}, \tau)$ at a phase-space position $(\mathbf{x}, \mathbf{q})$ and at conformal time τ in order to derive the perturbative redshift-space distortions. Here $\mathbf{x}$ is the comoving position and $\mathbf{q} = \mathbf{p}/a = m\mathbf{u}$ is the comoving momentum, where $\mathbf{u} = d\mathbf{x}/d\tau$. In the following we will omit the time dependence, i.e we will write $f(\mathbf{x}, \mathbf{q})$. The density field in real space is obtained by averaging the distribution function over momentum:

$$\rho(\mathbf{x}) \equiv m_p \int d^3\mathbf{q} \, f(\mathbf{x}, \mathbf{q}), \tag{2.1}$$

where m_p is the particle mass and $a = 1/(1+z)$ is the scale factor (z is the redshift). In redshift space the position is distorted by peculiar velocities, thus the comoving redshift-space coordinate for a particle is given by $\mathbf{s} = \mathbf{x} + \hat{r} \, u_\parallel/\mathcal{H}$, where $\hat{r}$ is the unit vector pointing along the observer's line of sight, $u_\parallel$ is the radial velocity, $m_p u_\parallel = q_\parallel = \mathbf{q} \cdot \hat{r}$, and $\mathcal{H} = aH$, where H is the Hubble parameter. Then the mass density in redshift space is given by

$$\rho_s(\mathbf{s}) = m_p \int d^3\mathbf{x} \, d^3\mathbf{q} \, f(\mathbf{x}, \mathbf{q}) \, \delta^D\left(\mathbf{s} - \mathbf{x} - \hat{r}\frac{u_\parallel}{\mathcal{H}}\right) = m_p \int d^3\mathbf{q} \, f\left(\mathbf{s} - \hat{r}\frac{u_\parallel}{\mathcal{H}}, \mathbf{q}\right). \tag{2.2}$$

By Fourier transforming equation 2.2, we find

$$\begin{aligned}
\rho_s(\mathbf{k}) &= m_p \int d^3\mathbf{x} \, d^3\mathbf{q} \, f(\mathbf{x}, \mathbf{q}) \, e^{i\mathbf{k}\cdot\mathbf{x} + ik_\parallel u_\parallel/\mathcal{H}} \\
&= m_p \int d^3\mathbf{x} \, e^{i\mathbf{k}\cdot\mathbf{x}} \int d^3\mathbf{q} \, f(\mathbf{x}, \mathbf{q}) \, e^{ik_\parallel u_\parallel/\mathcal{H}},
\end{aligned} \tag{2.3}$$

where $\mathbf{k}$ is the wavevector in redshift space, corresponding to the redshift-space coordinate $\mathbf{s}$.

Now we expand the second integral in equation 2.3 as a Taylor series in $k_\parallel u_\parallel/\mathcal{H}$,

$$\begin{aligned}
m_p \int d^3\mathbf{q} \, f(\mathbf{x}, \mathbf{q}) \, e^{ik_\parallel u_\parallel/\mathcal{H}} &= m_p \int d^3\mathbf{q} \, f(\mathbf{x}, \mathbf{q}) \sum_{L=0} \frac{1}{L!} \left(ik_\parallel u_\parallel/\mathcal{H}\right)^L \\
&= \bar{\rho}\left[\sum_{L=0} \frac{1}{L!} \left(\frac{ik_\parallel}{\mathcal{H}}\right)^L T_\parallel^L(\mathbf{x})\right],
\end{aligned} \tag{2.4}$$

where $\bar{\rho}$ is the mean mass density and

$$T_\parallel^L(\mathbf{x}) = \frac{m_p}{\bar{\rho}} \int d^3\mathbf{q} \, f(\mathbf{x}, \mathbf{q}) \, u_\parallel^L, \tag{2.5}$$

where in the last expression the integral over phase space assures that the quantity is defined in terms of a sum over all particles at position $\mathbf{x}$. For a single stream at $\mathbf{x}$ this is just $(1 + \delta(\mathbf{x}))u_\parallel^L(\mathbf{x})$. These are thus radial components of the moments of the distribution function and the distribution function description allows for inclusion of both bulk velocities and multi-streamed velocities. Note that these quantities are mass-weighted, and so well-defined for any system: one just sums over all the particles in the system weighting each one by the appropriate power of their radial velocities. If the field needs to be defined on a grid a simple assignment scheme of particles to the grid suffices, and empty grid cells are assigned a value of 0. We note this to contrast it with volume weighted quantities which need to be defined even if there are no particles assigned to a given grid cell, which is often impossible for sparse biased tracers, specially in underdense regions. We return to this issue later.

The Fourier component of the density fluctuation in redshift space is

$$\delta_s(\mathbf{k}) = \sum_{L=0} \frac{1}{L!} \left(\frac{ik_\parallel}{\mathcal{H}} \right)^L T_\parallel^L(\mathbf{k}) \ , \tag{2.6}$$

where $T_\parallel^L(\mathbf{k})$ is the Fourier transform of $T_\parallel^L(\mathbf{x})$.

$$T_\parallel^L(\mathbf{k}) = \int d^3\mathbf{x} \ T_\parallel^L(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{2.7}$$

For $L = 0$ we have $T_\parallel^0(\mathbf{k}) = \delta(\mathbf{k})$, the density fluctuation in real space.

2.1 Angular decomposition of moments of distribution function

The objects $T_\parallel^L(\mathbf{x})$ introduced in equation 2.5 are radial components of moments of the the distribution function, which are rank L tensors,

$$T_{i_1,i_2,..i_L}^L = \frac{m_p}{\bar{\rho}} \int d^3\mathbf{q} \ f(\mathbf{x},\mathbf{q}) \, u_{i_1} u_{i_2} ... u_{i_L}. \tag{2.8}$$

The real-space density field corresponds to $L = 0$, i.e. zeroth moment, the $L = 1$ moment corresponds to the momentum density, $L = 2$ gives the stress energy density tensor etc. These objects are symmetric under exchange of any two indices and have $(L+1)(L+2)/2$ independent components. They can be decomposed into helicity eigenstates under rotation around $\mathbf{k}$, as we do next.

Since translational symmetry guarantees that each Fourier mode is only correlated with itself, we can work with each Fourier mode separately, and add them appropriately in the end when we discuss the power spectra. By symmetry we may take $\mathbf{k}$ to be along z-axis. We can decompose the distribution function into spherical harmonics,

$$f(\mathbf{k}, q, \theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} f_l^m(\mathbf{k}, q)Y_{lm}(\theta, \phi), \tag{2.9}$$

where q is the amplitude of the momentum (often a term $(-i)^l\sqrt{\frac{4\pi}{2l+1}}$ is inserted into this expansion, but we will drop all such terms here). The components $f_l^m(\mathbf{k}, q)$ are helicity eigenmodes (i.e., eigenmodes of angular momentum component in z-direction $L_z = -i\partial/\partial\phi$) and under rotation by angle ψ around the z-axis they transform as

$$f_l^m(\mathbf{k}, q)' = e^{im\psi}f_l^m(\mathbf{k}, q). \tag{2.10}$$

This follows from the transformation properties of spherical harmonics. A quantity which transforms under rotation according to this equation is said to have helicity m. A quantity with helicity 0 is called a scalar, that with helicity $m = \pm 1$ is called a vector and that with $m = \pm 2$ a tensor, but the expansion goes to arbitrary values of m.

Moments of the distribution function are defined in terms of integrals of velocity moments over the distribution function. We can define helicity eigenstates of moments of the distribution function as

$$T_l^{L,m}(\mathbf{k}) = \frac{4\pi m_p}{\bar{\rho}} \int q^2 dq u^L f_l^m(\mathbf{k}, q).$$ (2.11)

Note that each term $T_l^{L,m}$ contains L powers of velocity u ($u \equiv |\mathbf{u}|$, and recall that $\mathbf{q} = m\mathbf{u}$).

A general rank L tensor $T_{i_1, i_2, ..i_L}^L$ can be decomposed into $2L + 1$ helicity eigenmodes $T_L^{L,m}$ ($m = -L, .., L$) and additional components formed by a product of a scalar u^2 with rank $L - 2$ tensors. This gives additional $2(L - 2) + 1$ helicity eigenmodes $T_{L-2}^{L,m}$ ($m = -(L - 2)..(L - 2)$), and additional components formed again by u^2 and rank $L - 4$ tensor, which gives additional $2(L - 4) + 1$ helicity eigenmodes $T_{L-4}^{L,m}$ ($m = -(L - 4)..(L - 4)$) etc.

For the lowest terms we have rank 1 tensor T_i^1, momentum density, which is a 3-vector in the usual geometrical context, and can be decomposed into a $m = 0$ helicity scalar component $T_1^{1,0}$ and two $m = \pm 1$ helicity vector components $T_1^{1,\pm 1}$. A general symmetric rank 2 tensor T_{ij}^2 has 6 independent components. These can be decomposed into an isotropic rank 0 helicity scalar term $T_0^{2,0} = (1 + \delta)u^2$, which corresponds to the energy density, and 5 $l = 2$ components: one helicity scalar part of anisotropic stress tensor $T_2^{2,0}$, two helicity vector part of anisotropic stress tensor $T_2^{2,\pm 1}$ and two helicity tensor part of anisotropic stress tensor $T_2^{2,\pm 2}$. At $L = 3$ we have 10 independent components, of which 7 are $T_3^{3,m}$, i.e. $l = 3$, $m = -3, .., 3$, and 3 are $T_1^{3,m}$, $l = 1$, $m = -1, 0, 1$ tensors, formed by taking isotropic u^2 and multiplying it with a 3-vector u_i, the latter of which can be decomposed into $m = -1, 0, 1$ components.

One can show that $2L + 1 + 2(L - 2) + 1 + 2(L - 4) + 1 + ... = (L + 1)(L + 2)/2$, so this decomposition gives the required number of independent components of a general symmetric tensor of rank L. In analysis of general relativity it is customary to expand the metric and stress energy tensor into scalar ($m = 0$), vector ($m = \pm 1$) and tensor ($m = \pm 2$) helicity modes (SVT decomposition). No higher order helicity modes are needed, since only tensors of rank 0, 1 and 2 enter into the description of the metric and energy momentum tensor. In contrast the moments of distribution function contain tensors of arbitrary rank and the expansion in equations 2.9, 2.11 is the appropriate generalization of the SVT decomposition.

So far we worked in the basis defined by $\mathbf{k}$ pointing in z direction. In general we are interested in computing the components of the moments in the radial direction $\hat{r}$. If $\hat{r}$ is parallel to $\mathbf{k}$ then only $m = 0$ components contribute, while for a general direction all of them do. The angular dependence of the moments is obtained by performing a rotation of the basis from $z||\mathbf{k}$ to $z'||\hat{r}$. We can achieve this by rotating by ϕ around z and then by θ around the axis perpendicular to z, z', so in terms of the general rotation by 3 Euler angles we have $T_l^{L,m'} = \sum_{-m}^m D_{m,m'}^l(\phi, \theta, 0) T_l^{L,m}$, where $D_{m,m'}^l(\phi, \theta, \phi')$ is the general rotation matrix of spin l associated with the 3 Euler angles ϕ, θ, ϕ' (we do not need to perform the rotation around z' by ϕ'). Since $u_\parallel$ is invariant under rotation around $z'||\hat{r}$ only $m' = 0$ survive. The rotation matrix is given by the spherical harmonics, $D_{0,m}^l(\phi, \theta, 0) = \sqrt{4\pi/(2l + 1)} Y_{lm}(\theta, \phi)$. Combining all together we find

$$T_\parallel^L(\mathbf{k}) = \sum_{(l=L, L-2, ..)} \sum_{m=-l}^{m=l} n_l^L T_l^{L,m}(\mathbf{k}) Y_{lm}(\theta, \phi),$$ (2.12)

where N_l^L is a constant independent of angle whose numerical value will not be needed.

3 Power spectra

We will adopt a plane-parallel approximation, where only the angle between the line of sight and the Fourier mode needs to be specified. The redshift-space power spectrum is defined as $\langle \delta_s(\mathbf{k}) \delta_s^*(\mathbf{k}') \rangle = P^{ss}(\mathbf{k}) \delta_D(\mathbf{k} - \mathbf{k}')$. Equation 2.6 gives,

$$P^{ss}(\mathbf{k}) = \sum_{L=0}^\infty \sum_{L'=0}^\infty \frac{(-1)^{L'}}{L! \, L'!} \left(\frac{ik_\parallel}{\mathcal{H}}\right)^{L+L'} P_{LL'}(\mathbf{k}),$$ (3.1)

where $P_{LL'}(\mathbf{k})\delta(\mathbf{k}-\mathbf{k}') = \langle T_{\parallel}^{L}(\mathbf{k})(T_{\parallel}^{*L'}(\mathbf{k}'))\rangle$. Note that $P_{LL'}(\mathbf{k}) = P_{L'L}(\mathbf{k})^*$ so that the total result is real valued, as expected. Thus we only need to consider the terms $P_{LL'}(\mathbf{k})$ with $L \le L'$, each of which comes with a factor of 2 if $L \ne L'$ and 1 if $L = L'$. We can also write $k_{\parallel}/k = \cos\theta = \mu$,

$$P^{ss}(\mathbf{k}) = \sum_{L=0}^{\infty} \frac{1}{L!^2}\left(\frac{k\mu}{\mathcal{H}}\right)^{2L} P_{LL}(\mathbf{k}) + 2\sum_{L=0}^{\infty}\sum_{L'>L} \frac{(-1)^{L'}}{L!\,L'!}\left(\frac{ik\mu}{\mathcal{H}}\right)^{L+L'} P_{LL'}(\mathbf{k})\,. \tag{3.2}$$

Next we want to insert the helicity decomposition of equation 2.12 and consider the implications of rotational symmetry on the power spectrum. Each term $P_{LL'}(\mathbf{k})$ contains products of multipole moments

$$T_{l}^{L,m}Y_{lm}(\theta,\phi)[T_{l'}^{L',m'}Y_{l'm'}(\theta,\phi)]^* \propto e^{i(m-m')\phi}. \tag{3.3}$$

Upon averaging over the azimuthal angle ϕ of Fourier modes all the terms with $m \ne m'$ vanish. Another way to state this is that upon rotation by angle Ψ the correlators pick up a term $e^{i(m-m')\Psi}$, and in order for the power spectrum to be rotationally invariant we require $m = m'$. Putting it all together we find

$$P_{LL'}(\mathbf{k}) = \sum_{(l=L,L-2,..)}\sum_{(l'=L',L'-2,..;\ l'\ge l)}\sum_{m=0}^{l} P_{l,l'}^{L,L',m}(k)P_{l}^{m}(\mu)P_{l'}^{m}(\mu), \tag{3.4}$$

where $P_{l}^{m}(\mu=\cos\theta)$ are the associated Legendre polynomials, which determine the θ angular dependence of the spherical harmonics, $Y_{lm}(\theta,\phi) = \sqrt{(2l+1)(l-m)!/4\pi(l+m)!}P_{l}^{m}(\cos\theta)e^{im\phi}$. We absorbed all of the terms that depend on l and m and various constants into the definition of power spectra $P_{l,l'}^{L,L',m}(k)$, replaced the two helicity states $\pm m$ by a single one with $m > 0$, since their θ angular dependencies are the same, and we absorbed the factor of 2 into the definition of $P_{l,l'}^{L,L',m}(k)$. We also require $l' \ge l$ and absorb the factor of 2 into the definition of $P_{l,l'}^{L,L',m}(k)$ since the two terms have the same angular structure. Note that due to statistical isotropy the spectra $P_{l,l'}^{L,L',m}(k)$ depend only on amplitude of k, i.e. we have

$$P_{l,l'}^{L,L',m}(k) \propto \langle T_{l}^{L,m}(\mathbf{k})(T_{l'}^{L',m}(\mathbf{k}))^*\rangle. \tag{3.5}$$

All the angular structure is thus in associate Legendre polynomials $P_{l}^{m}(\mu)$.

Equations 3.2 and 3.4 are the main result of this paper. They show that there exists a well defined expansion in terms of cross and auto-power spectra of velocity moments. The expansion parameter is roughly defined as $k\mu u/\mathcal{H}$, where u is related to a typical gravitational velocity of the system (which should be of the order of hundreds of km/s, but note that we take higher and higher powers of these velocities in the series). The expansion is convergent if the expansion parameter is less than unity.

In terms of perturbation theory there is a close, but not one to one, relation between the lowest order of perturbation theory and the order of the moment expansion. Assuming δ and $ku/\mathcal{H}$ are of the same order, the lowest order of the contribution in terms of powers of power spectrum (i.e., quadratic in δ) is $(L+L')/2$ if $L+L'$ is even and $L > 0$, and $(L+L'+1)/2$ if odd and $L > 0$, while for $L = 0$ it is $L'/2 + 1$ if L' even or $(L'+1)/2$ if L' odd, but of course all higher order terms also enter.

These equations also show that there is a close relation between the order of the moments and their angular dependence. To understand the angular dependence we first note that associated Legendre polynomial $P_{l}^{m}(\mu)$ contains powers from 1 to μ^{l-m} for even l, and from μ to μ^{l-m}, for odd l, and is always multiplied with a power of $(1-\mu^2)^{m/2}$. Thus $P_{l,l'}^{L,L',m}(k)$ gets multiplied with powers of $(1-\mu^2)^m$ or $\mu(1-\mu^2)^m$ to $\mu^{l+l'-2m}(1-\mu^2)^m$, so the highest order is $\mu^{l+l'}$. In addition we have $\mu^{L+L'}$ dependence in equation 3.2, so the lowest contribution in powers of μ to $P^{ss}(k)$ is $\mu^{L+L'}$ if $L+L'$ is even or $\mu^{L+L'+1}$ if $L+L'$ is odd, and the highest is $\mu^{2(L+L')}$. Thus for $P_{00}(\mathbf{k})$ the only angular term is isotropic, for $P_{01}(\mathbf{k})$ the only angular term is μ^2, $P_{11}(\mathbf{k})$ and $P_{02}(\mathbf{k})$ contain both μ^2 and μ^4 etc. Note that only even powers of μ enter in the final expression, as required by the symmetry. We now proceed to look in more detail at the lowest order terms.

3.1 $P_{00}(\mathbf{k})$: the isotropic term

At the lowest order in the expansion we have correlation of real space density $T_{||}^0 = \delta(\mathbf{k})$ with itself. Density is a scalar of rank 0, $P_0(\mu) = 1$. The power spectrum is isotropic $P_{00}(\mathbf{k}) = P_{00}^{000}(k)$. This term is just the real space power spectrum and of course does not have any μ dependence since it is independent of redshift space distortions. For small values of μ this term always dominates, and in the limit $\mu = 0$ the transverse power spectrum becomes the real power spectrum $P_{00}(k)$. The real space power spectrum agrees with the linear one on large scales, $P_{00}(k) = P_{\rm lin}(k)$, slightly dips below the linear one around $k \sim 0.1 h/Mpc$, while on even smaller scales the nonlinear corrections cause it to increase over the linear one.

3.2 $P_{01}(\mathbf{k})$

At the next order in our expansion (not in perturbation theory PT, see below) we have correlations between the density $T_{||}^0(\mathbf{k}) = \delta(\mathbf{k})$ and radial component of momentum density $T_{||}^1(\mathbf{k}) = [(1+\delta)u_{||}](\mathbf{k})$. Momentum density can be decomposed into a scalar ($m = 0$) $T_1^{1,0}$ and two vector ($m = \pm 1$) components $T_1^{1,\pm 1}$, but only the scalar part correlates with the density $T_0^{0,0}$, which is a scalar. Thus the only contribution comes from $P_{0,1}^{0,1,0}(k) \propto \langle T_0^{0,0}(\mathbf{k})(T_1^{1,0}(\mathbf{k}))^* \rangle$,

$$P_{01}(\mathbf{k}) = P_{0,1}^{0,1,0}(k)\mu, \tag{3.6}$$

where we used $P_1^0(\mu) = \mu$.

The scalar mode of momentum can be obtained from the divergence of momentum and related to $\dot{\delta}$ using the continuity equation, which in terms of our quantities is

$$\dot{T}_0^{0,0} + ik T_1^{1,0} = 0. \tag{3.7}$$

This is an exact relation (for conserved quantities), in the sense that the vector part of momentum does not contribute to it, since it vanishes upon taking the divergence (i.e., vector components are orthogonal to $\mathbf{k}$ and the dot product is zero).

From this we get that

$$P_{01}(\mathbf{k}) = -ik^{-1}\mu P_{\delta,\dot{\delta}}(k) = -\frac{i\mu}{2k}\frac{dP_{00}(k)}{dt}. \tag{3.8}$$

The total contribution from this term to $P^{ss}(k)$ is

$$P_{01}^{ss}(\mathbf{k}) = \mu^2 \mathcal{H}^{-1}\frac{dP_{00}(k)}{d\tau} = \mu^2 \frac{dP_{00}(k)}{d\ln a}. \tag{3.9}$$

This is an exact relation for dark matter, valid also in the nonlinear regime. It shows that this term can be obtained directly from the redshift evolution of the dark matter power spectrum $P_{00}(k)$. On large scales it agrees with the linear theory predictions. If we write $P_{00}(k) = D(a)^2 P_{\rm lin}(k)$, with $D(a)$ the linear growth rate and $f = d\ln D/d\ln a$, then we find $P_{01}^{ss}(\mathbf{k}) = 2f\mu^2 P_{\rm lin}(k)$. We thus see that this order is of the same order in PT as the first order $P_{00}(k)$, the well known Kaiser result [2]. On smaller scales we expect the term to deviate from the linear one, just as for $P_{00}(k)$.

3.3 $P_{11}(\mathbf{k})$

The next term is the correlation of the momentum density $T_{||}^1(\mathbf{k})$ with itself. In this case the scalar ($m = 0$) $T_1^{1,0}(k)$ correlates with itself and the vector ($m = \pm 1$) components $T_1^{1,\pm 1}(k)$ also correlate with itself, so both components of momentum contribute,

$$P_{11}(\mathbf{k}) = P_{1,1}^{1,1,0}(k)[P_1^0(\mu)]^2 + P_{1,1}^{1,1,1}(k)[P_1^1(\mu)]^2. \tag{3.10}$$

In terms of the contribution to the redshift space power spectrum this gives

$$P_{11}^{ss}(\mathbf{k}) = \mathcal{H}^{-2}k^2\mu^2[P_{1,1}^{1,1,0}(k)\mu^2 + P_{1,1}^{1,1,1}(k)(1-\mu^2)]. \tag{3.11}$$

The scalar part of the momentum is the one that contributes to the continuity equation 3.7. In linear perturbation theory only the scalar contribution is non-zero and $P_{1,1}^{1,1,0}(k) = f^2 P_{\rm lin}(k)$. This term is also of linear order and collecting all terms at this order we obtain the usual expression [2]

$$P_{\rm lin}^{ss}(\mathbf{k}) = (1 + f\mu^2)^2 P_{\rm lin}(k). \tag{3.12}$$

However, we see from the above that there will be another contribution to both μ^2 and μ^4 terms from the vector part of momentum correlator $P_{1,1}^{1,1,1}(k) \propto \langle |T_1^{11}(k)|^2 \rangle$, which comes in at the second order in power spectrum. This vector part is often called the vorticity part of the momentum. In general this term is non-zero because vorticity of momentum does not vanish, even if vorticity of velocity vanishes for a single streamed fluid [16]. As seen from equation 3.11 this term always *adds* power to μ^2 term and subtracts it in μ^4 term (but is combined with a positive contribution from the scalar part in μ^4 term).

3.4 $P_{02}(\mathbf{k})$

At orders higher than $P_{11}(\mathbf{k})$ we no longer have any linear contributions, hence these terms are usually not of interest for extracting the cosmological information. However, these terms, including what is sometimes called the Fingers-of-God (FoG) effect, are known to be important on fairly large scales. Here we will limit the discussion to some general statements of their k and μ dependence, leaving their more precise calculations to future work.

There are two different terms that contribute to this term,

$$P_{02}(\mathbf{k}) = P_{0,0}^{0,2,0}(k)[P_0^0(\mu)]^2 + P_{0,2}^{0,2,0}(k)P_0^0(\mu)P_2^0(\mu). \tag{3.13}$$

In terms of the contribution to the redshift space power spectrum this gives

$$P_{02}^{ss}(\mathbf{k}) = -\left(\frac{k\mu}{\mathcal{H}}\right)^2 \left[P_{0,0}^{0,2,0}(k) + \frac{1}{2}P_{0,2}^{0,2,0}(k)(3\mu^2 - 1) \right]. \tag{3.14}$$

The first term is the correlation between the isotropic part of the mass weighted square of velocity, i.e. the energy density, $T_0^{2,0} = (1 + \delta)u^2$ and the density field $T_0^{0,0} = \delta$. The second term comes from the scalar part of the anisotropic stress $T_2^{2,0}$ correlated with the density $T_0^{0,0} = \delta$.

On physical grounds we expect the first term to be large in systems with a large rms velocity resulting in a term scaling as $P_{0,0}^{0,2,0}(k) \sim P_{00}(k)\sigma^2$, where σ^2 has units of velocity squared, but is not simply the volume averaged velocity squared (see below). The contribution of this term to P^{ss} goes as $-(k\mu/\mathcal{H})^2\sigma^2 P_{00}(k)$, i.e. it is a damping term suppressing the linear power spectrum, with the effect increasing towards higher k (smaller scales). This is the lowest order FoG term, which we see contributes as $(k\mu)^2$ dependence and so affects the μ^2 term. It is a damping term that is always negative, while the corresponding μ^2 term from P_{11} always adds power. The scalar anisotropic stress-density correlator $P_{0,2}^{0,2,0}(k)$ also contributes to μ^2 angular term, as well as to μ^4 angular term, and is formally of the same order in perturbation theory as $P_{0,0}^{0,2,0}(k)$, but is likely to be smaller on physical grounds that velocity dispersion in virialized objects is isotropic and hence has a small anisotropic stress.

3.5 $P_{03}(\mathbf{k})$, $P_{12}(\mathbf{k})$, $P_{04}(\mathbf{k})$, $P_{13}(\mathbf{k})$ and $P_{22}(\mathbf{k})$

Since the lowest order in μ is $(L + L')$ or $L + L' + 1$, the terms of order higher than $P_{02}(\mathbf{k})$ do not contribute to μ^2 term. At the next order in μ we have terms of order μ^4. At this order there are 7 terms that contribute. $P_{11}(\mathbf{k})$ and $P_{02}(\mathbf{k})$ we already discussed: while $P_{11}(\mathbf{k})$ has a linear order term and is expected to dominate on large scales, $P_{02}(\mathbf{k})$ is second order in power spectrum. They both come with a prefactor of k^2. At the next order we have $P_{03}(\mathbf{k})$ and $P_{12}(\mathbf{k})$, both second order in power spectrum and each multiplied by k^3, followed by $P_{13}(\mathbf{k})$ and $P_{22}(\mathbf{k})$, also second order in power spectrum, but each multiplied by k^4, and by $P_{04}(\mathbf{k})$, third order in power spectrum. All of these terms also contribute to terms of higher order in μ^{2j}, up to μ^6 or μ^8 for these terms.

One can see from this discussion that the angular structure of higher order terms is considerably more complex than that of lower order terms and that all the terms even in powers of μ are being generated by RSD. However, there is a connection between the angular order, powers of k and lowest order in perturbation theory, such that only the low powers of k and low lowest order of PT contribute to the lowest orders in μ. Thus, at low values of $k\mu$, the series is convergent. To make these statements more quantitative a numerical or perturbative analysis is required, which will be presented elsewhere [17].

3.6 Shot noise and connections between the correlators

The correlators at the same order in powers of velocity, i.e. equal $L + L'$, contain nontrivial cancellations among them. To see this assume velocity is constant over a region of space $r \sim k_0^{-1}$. For example, large scale bulk flows lead to correlated velocities on small scales, giving rise to nearly equal velocities between nearby particles. On scales smaller than this velocity coherence scale, $k > k_0$, we can pull out these constant velocity terms from the correlators, to obtain

$$P_{LL'}(k)\delta_D(\mathbf{k} - \mathbf{k}') = \langle[(1 + \delta)u_\parallel^L](\mathbf{k})[(1 + \delta)u_\parallel^L]^*(\mathbf{k}')\rangle \sim P_{00}(k)\langle u_\parallel^{L+L'}\rangle\delta_D(\mathbf{k} - \mathbf{k}'), \tag{3.15}$$

where $\langle u_\parallel^{L+L'}\rangle$ is just a number corresponding to the spatial average of this term. So these terms are all equal as long as $L + L'$ is the same.

These terms enter into the sum in equation 3.1 with different prefactors and opposite signs, leading to cancellations between them. The lowest order example is that of P_{11} and P_{02}, which enter with equal prefactors but opposite signs, canceling any such contributions from each other. This is not surprising: bulk flows lead to "rigid body" displacements of particles but do not contribute to FoG effects, so their contribution to P_{02} must be canceled. As a result, only velocity dispersion type contributions lead to FoG effects.

In the extreme case this argument can be applied to the shot noise for these correlators, which is the contribution to the power spectrum caused by discreteness of tracers. It is well known that the shot noise of a density field sampled by tracers of number density $\bar{n}$ is given by $P_{00}(k) = \bar{n}^{-1}$. Analogous calculation for the moments gives

$$P_{LL'}(k) = \bar{n}^{-1}\langle u_\parallel^{L+L'}\rangle. \tag{3.16}$$

This expression is exact, since by definition a discrete tracer population only has a single value of velocity at any given position, so $\langle u_\parallel^{L+L'}\rangle$ will be the same for any pair of L, L' such that $L + L'$ is the same. These shot noise terms can be large if the tracer is sparse, i.e. if $\bar{n}$ is small. However, the argument above shows that these terms enter with opposiste signs in the final result and so these shot noise contributions cancel in the total sum of 3.1. This is expected: the only shot noise contribution to the total RSD power spectrum $P^{ss}(k)$ should be $\bar{n}^{-1}$. These examples show that these velocity moments are connected, and it is more natural to consider them together, such as $P_{11}(k) - P_{02}(k)$, where the shot noise and the bulk flow terms cancel out.

3.7 Relation to Legendre moments

In RSD analyses it is customary to integrate $P^{ss}(\mathbf{k})$ over the lowest order Legendre polynomials to obtain moments $P_l^{ss}(k)$,

$$P_l^{ss}(k) = (2l + 1)\int_0^1 P^{ss}(\mathbf{k})P_l(\mu)d\mu, \tag{3.17}$$

where $P_l(\mu)$ are the ordinary Legendre polynomials, $P_0(\mu) = 1$, $P_2(\mu) = (3\mu^2 - 1)/2$ and $P_4(\mu) = (35\mu^4 - 30\mu^2 + 3)/8$. Only the lowest 3 orders contain contributions from linear terms, so the analysis is usually limited to $l = 0, 2, 4$. The 0 moment is just the spherical average of the power spectrum in redshift space. The advantage of this expansion is that in a typical survey the moments are uncorrelated on scales small compared to the size of the survey.

Moments in even l can be viewed as an alternative way to expand in terms of even powers of μ. However, the expansion given in equation 3.4 is *not* an expansion in Legendre polynomials, since

it contains products of associated Legendre polynomials (including squares of ordinary Legendre polynomials of both even and odd orders). Hence there is no orthogonality between the moments of distribution function and Legendre moments $P_l^{ss}(k)$. So if we expand the angular dependence of any given term $P_{LL'}^{ss}(k)$ into Legendre polynomials, we will generate all orders up to $l = 2(L + L')$. This means for example that all terms will contribute to the monopole $l = 0$, all except $P_{00}^{ss}(k)$ to quadrupole $l = 2$ and all but $P_{00}^{ss}(k)$ and $P_{10}^{ss}(k)$ to hexadecupole $l = 4$. As a result we always have an infinite number of terms $P_{LL'}(\mathbf{k})$ contributing to any given Legendre series term $P_l^{ss}(k)$, with higher and higher powers of k. This expansion is thus considerably more complex than the expansion in powers of μ^{2j}, which has a finite number of terms for any given value of j.

The discussion above suggests it may be more beneficial to fit for powers of μ^{2j} rather than for Legendre moments, for example by fitting for μ^0, μ^2 and μ^4 terms, which contain linear order contributions, together with higher order terms μ^6, μ^8 etc., which we do not care for and can marginalize over in the end. However, Legedre moments are uncorrelated while powers of μ^{2j} are strongly correlated, so a marginalization over higher order terms will lead to a large increase in errors for higher k, given that these terms become very large at high k compared to lowe order terms. So this can only work if sufficiently strong priors are adopted for higher order terms μ^6, μ^8 etc. Such priors could come from simulations extracting individual higher order terms or from a parametrized model. This is pursued further in [17].

3.8   Applications to galaxies and issues of bias

The relation to other tracers such as galaxies is a rich subject worth exploring further with this method. In this paper we focus primarily on the dark matter, but all the derivations remain unchanged if the dark matter particles are replaced with some other tracers, such as galaxies or halos. In large scale structure we usually define bias as the ratio of galaxy power spectrum (shot noise subtracted) to matter power spectrum, $b^2(k) = P_{00}^{gg}(k)/P_{00}^{mm}(k)$. We can generalize the concept of bias to

$$b_{LL'}(\mathbf{k}) = \frac{P_{LL'}^{gg}(\mathbf{k})}{P_{LL'}^{mm}(\mathbf{k})}, \tag{3.18}$$

where $P_{LL'}^{gg}(\mathbf{k})$ is galaxy correlator and $P_{LL'}^{mm}(\mathbf{k})$ is the corresponding dark matter term. In linear theory we have $b_{00} = b_1^2$, $b_{01} = b_1$ and $b_{11} = 1$, independent of scale or angle, where linear bias is defined as $\delta_g = b_1 \delta_m$. Two ways to extract cosmological information from RSD are either by combining P_{00} and P_{01} to eliminate b_1, or to use P_{11} directly.

Before discussing RSD further it is useful to draw a comparison to weak lensing. In case of weak lensing we can measure both projected dark matter density or galaxy density, so we can perform a joint correlation analysis of galaxy clustering and weak lensing, where the galaxy auto-correlation is proportional to b^2 times matter correlations, cross-correlation between galaxies and weak lensing signal around them induced by the dark matter is proportional to b (the so called galaxy-galaxy lensing), while the weak lensing auto-correlation is independent of bias. Two ways to extract the signal are either using just shear-shear correlations tracing matter-matter correlations, or combining galaxy auto-correlation with galaxy-galaxy lensing to eliminate bias. This latter has higher signal to noise but is complicated due to the fact that bias is scale independent and the scale dependence depends on the galaxy properties [18]. To understand when this happens it is useful to expand galaxy density perturbation to second order in matter density, $\delta_g = b_1 \delta_m + b_2 \delta_m^2$. The second order terms will become important when they cannot be neglected against the first order terms, so the expansion parameter is $(b_2/b_1)\delta_m$. Since $\delta_m^{\rm rms}$ increases on small scales this scale dependent bias increases towards small scales. Typically we have $|b_2/b_1| < 0.4$ [18] and the corrections become important at $k \sim 0.1 h/Mpc$, where $\delta_m^{\rm rms} \sim 0.5$.

Returning back to RSD, our formalism is directly applicable to galaxies, except that all the velocity moments are mass weighted for the dark matter, $T_\parallel^{L,m} = (1 + \delta_m)u_\parallel^L$, and number density weighted for the galaxies, $T_\parallel^{L,g} = (1 + \delta_g)u_\parallel^L$. What this shows is that if the density distribution of galaxies differs from that of the dark matter then all the correlators of velocity moments will differ from each other, even those that appear independent of bias, such as P_{11}. In reality thus the

predictions of linear bias model will be modified, because even if galaxies are faithful tracers of the dark matter velocities at a given position, the weighting of the velocity moments differs: in one case they are weighted by the dark matter mass, in the other by the number of galaxies, and the two differ in their spatial distribution. This will result in scale dependence of the higher order bias terms $b_{LL'}$, just like it does for the b_{00} itself [19].

To quantify this further, for the lowest order momentum density term and for linear bias we must compare correlation of $(1 + b_1\delta_m)u_\parallel$ with itself to give $P_{11}^{gg}(\mathbf{k})$ or with $b_1\delta_m$ to give $P_{01}^{gg}(\mathbf{k})$. The auto-correlation will give the result that agrees with the dark matter only for $b_1 = 1$, or if $b_1\delta_m \ll 1$. In the same limit the cross-correlation will give linear bias b_1. So momentum density becomes velocity in the limit $b_1\delta_m \ll 1$, while requiring it to be scale independent relative to dark matter requires something like $(b_1 - 1)\delta_m \ll 1$, which for typical LRG galaxies ($b_1 \sim 2$) is in fact a more stringent requirement than that of a scale independent bias condition discussed above (b_2/b_1 versus b_1). This suggests that the scale dependence of the momentum density bias terms b_{01} and b_{11} defined in equation 3.18 extends to larger scales than scale dependent bias of density b_{00}.

The conclusion from this discussion is that the scale dependence of bias terms involving momentum density is a real concern in RSD and likely extends to relatively large scales ($k < 0.1h/Mpc$). In terms of the angular decomposition in powers of μ, the discussion of the scale dependent bias of RSD can be divided into a μ^2 term, which depends entirely on b_{01} in P_{01}, and the μ^4 term, for which the scale dependence of b_{11} term above is applicable, since that is the term that does not vanish on large scales in linear theory. In this sense RSD analysis is not the equivalent of a joint galaxy-weak lensing analysis, since weak lensing auto-correlation truly traces the dark matter directly, while in RSD this limit is achieved only on relatively large scales where $\delta_g^{rms} \ll 1$.

The discussion so far completely ignored FoG effects: for μ^2 term these are encoded in P_{02} and in vector part of P_{11}, which unlike P_{02} adds power rather than removes it, and these terms have their own physical interpretation and scale dependence unrelated to the scale dependent bias discussion above. While they can partially cancel the effects discussed above they are unlikely to achieve this exactly. In most of the literature so far only the scale dependence induced by FoG effects was discussed (although see [7, 12]). The simple linear bias model predicts that FoG effects scale with bias squared: the leading order term scales as b_1^2 both in $P_{02}^{gg} \propto \langle[b_1\delta v_\parallel^2](\mathbf{k})b_1\delta(-\mathbf{k})\rangle$ and in $P_{11}^{gg} \propto [b_1\delta v_\parallel]^2$. If we write $P_{02}^{gg}(k) - P_{11}^{gg}(k) = P_{00}^{gg}(k)\sigma^2$, then σ is independent of bias, since $P_{00}^{gg}(k) \propto b_1^2 P_{00}^{mm}(k)$.

4   Discussion

In this paper we present a distribution function approach to redshift space distortions. We show that the redshift space density can be expressed in terms of a sum over velocity moments and the redshift space power spectrum can be expressed in terms of correlators between the Fourier components of these moments. These moments are simple objects to calculate in any system: they are calculated by simply taking appropriate powers of radial velocity and summing over all particles. The lowest order moments are density, momentum density, stress energy density etc.

We have decomposed the moments into helicity eigenstates based on their transformation properties under rotation around the direction of the Fourier mode, a generalization of SVT decomposition in cosmological perturbation theory. We use rotational invariance to derive all of the allowed correlator terms, showing that only terms with the same helicity can contribute to the correlators. The moments of distribution function are complicated objects with many terms allowed by symmetries, specially at higher order, leading to a complicated angular and scale dependence, suggesting that treatments of RSD cannot be fully successful with simple ansatzes, such as the popular FoG velocity dispersion model with one free parameter [10, 11].

Despite the complexity of the general RSD description some general statements can be made. The lower order terms generally only contribute to low orders of expansion in μ^2, where μ is the angle between the Fourier mode and the line of sight. As an example, we have shown that only the scalar part of the momentum density correlates with the density, and this term can be written in terms of a time derivative of the power spectrum. This term only contributes to μ^2 angular dependence and contains a linear order term. But there is also the vector part of the momentum density-momentum

density correlation, the (scalar) energy density-density correlation, and the scalar part of anisotropic stress density-density correlation, all of which also contribute to the μ^2 term. They are all nonlinear and cannot dominate on very large scales, but likely dominate on small scales. The energy density-density correlation term is the term most closely related to the FoG velocity dispersion effect and is always negative, suppressing the power, but the other terms are formally of the same order in perturbation theory. We have shown that the vorticity part of momentum always adds to the RSD power of μ^2 term, and hence acts in the opposite direction to the FoG term. Our analysis cannot address which term has a larger amplitude, but it would be interesting to see if there are any systems where the terms that add power dominate over those that suppress it. The next angular term has μ^4 dependence and we identified 7 terms that contribute to it, of which one, scalar part of P_{11}, contains a linear contribution that does not vanish on large scales.

The fact that there are a finite number of velocity moment terms at each order of μ^{2j} expansion should be contrasted to the popular Legendre multipoles expansion (monopole, quadrupole and hexadecupole contain cosmological information), which receive contributions from all orders in moments of distribution function. This suggests that a better behaved analysis may be possible if instead of a multipole analysis the analysis is performed in terms of a μ^{2j} expansion, with the lowest 3 orders containing cosmological information and the rest treated as nuisance parameters.

It is important to emphasize that these moments are mass weighted quantities, and no volume averaged quantities ever enter into our expressions. This relates to one of the long standing issues in the treatment of RSD: many of the past treatments [10, 11, 20] have assumed that RSD trace correlations between velocities and dark matter and that the FoG effects multiply these density-velocity and velocity-velocity correlations, where FoG quantities are also defined as volume weighted quantities such as velocity dispersion $\sigma^2 = \langle u^2 \rangle$. But these volume weighted quantities are not well defined, specially for sparse biased systems such as galaxies or clusters. For a biased tracer with $b > 1$ one finds that voids with no tracers in them are enlarged, since, for $\delta < 0$, $1 + b\delta$ is closer to 0 than $1 + \delta$. This has forced some workers to use the dark matter velocity field instead, with unpredictable results [11]. Our expansion shows that it is more natural to define RSD in terms of mass or number weighted quantities, such as momentum density or energy density, the former replacing velocity and the latter replacing velocity dispersion. Mass and number weighted moments such as momentum or energy density are well defined even in voids (where they are simply zero). In this paper we show that there is a consistent expansion using mass weighted moments, and that the expansion is convergent on large scales.

The fact that all RSD quantities are density weighted also suggests that RSD effects will differ if the galaxy number density distribution differs from mass density distribution. We have shown that even a linear bias model induces scale dependent bias of the momentum density correlators, and that this scale dependence is likely to show up on relatively large scales, $k < 0.1\mathrm{h/Mpc}$. The success of RSD in extracting cosmological information depends entirely on our ability to model these various bias terms and relate them to each other. Similarly, the success of the approach presented here in modeling RSD depends on our ability to extract these moments from simulations and data and on our ability to model them with analytic models, such as perturbation theory. Providing physical interpretation of the terms, as done here, could enable one to develop more effective modeling, or provide a better physical understanding of limitations of RSD in extracting cosmological information. For example, it is relatively straight-forward to include the bias induced scale dependence effect at the lowest order of PT and we will present the results elsewhere [21]. In this paper we have focused on theory, conceptual issues and general symmetries, while applications to simulations and perturbation theory will be presented in upcoming work [17, 21].

Acknowledgments

References

[1] A. J. S. Hamilton, in D. Hamilton, ed., *The Evolving Universe* (1998), vol. 231 of *Astrophysics and Space Science Library*, pp. 185–+, arXiv:astro-ph/9708102.

[2] N. Kaiser, *Mon. Not. Roy. Astron. Soc.* **227**, 1 (Jul. 1987).

[3] M. White, Y. Song, and W. J. Percival, *Mon. Not. Roy. Astron. Soc.* **397**, 1348 (Aug. 2009), 0810.1518.

[4] P. McDonald and U. Seljak, Journal of Cosmology and Astro-Particle Physics **10**, 7 (Oct. 2009), 0810.0323.

[5] G. M. Bernstein and Y.-C. Cai, ArXiv e-prints (Apr. 2011), 1104.3862.

[6] A. F. Heavens, S. Matarrese, and L. Verde, *Mon. Not. Roy. Astron. Soc.* **301**, 797 (Dec. 1998), arXiv:astro-ph/9808016.

[7] R. Scoccimarro, *Phys. Rev.* **D 70**(8), 083007 (Oct. 2004).

[8] J. L. Tinker, *Mon. Not. Roy. Astron. Soc.* **374**, 477 (Jan. 2007), arXiv:astro-ph/0604217.

[9] A. Taruya, T. Nishimichi, and S. Saito, *Phys. Rev.* **D 82**(6), 063522 (Sep. 2010), 1006.0699.

[10] E. Jennings, C. M. Baugh, and S. Pascoli, *Mon. Not. Roy. Astron. Soc.* **410**, 2081 (Jan. 2011), 1003.4282.

[11] J. Tang, I. Kayo, and M. Takada, ArXiv e-prints (Mar. 2011), 1103.3614.

[12] B. A. Reid and M. White, ArXiv e-prints (May 2011), 1105.4165.

[13] M. Sato and T. Matsubara, ArXiv e-prints (May 2011), 1105.5007.

[14] P. J. E. Peebles, *The large-scale structure of the universe* (Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p., 1980).

[15] P. McDonald, *JCAP* **4**, 32 (Apr. 2011), 0910.1002.

[16] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, *Phys. Rept.* **367**, 1 (2002).

[17] T. Okumura et al., , in preparation (2011).

[18] T. Baldauf, R. E. Smith, U. Seljak, and R. Mandelbaum, *Phys. Rev.* **D 81**(6), 063531 (Mar. 2010), 0911.4973.

[19] N. Hamaus, U. Seljak, V. Desjacques, R. E. Smith, and T. Baldauf, *Phys. Rev.* **D 82**(4), 043515 (Aug. 2010), 1004.5377.

[20] M. Tegmark, M. R. Blanton, M. A. Strauss, F. Hoyle, D. Schlegel, R. Scoccimarro, M. S. Vogeley, D. H. Weinberg, I. Zehavi, A. Berlind, *et al.*, *Astrophys. J.* **606**, 702 (May 2004).

[21] Z. Vlah et al., , in preparation (2011).