This is the accepted manuscript made available via CHORUS. The article has been published as:

Large scale structure reconstruction with short-wavelength modes
Peikai Li, Scott Dodelson, and Rupert A. C. Croft
Phys. Rev. D 101, 083510 — Published 6 April 2020
DOI: 10.1103/PhysRevD.101.083510
Large Scale Structure Reconstruction with Short-Wavelength Modes

Peikai Li,1,2 Scott Dodelson,1,2 and Rupert A. C. Croft1,2

1 Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA
2 McWilliams Center for Cosmology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

(Dated: February 19, 2020)

Large scale density modes are difficult to measure because they are sensitive to systematic observational errors in galaxy surveys, but we can study them indirectly by observing their impact on small scale perturbations. Cosmological perturbation theory predicts that second-order density inhomogeneities are a convolution of a short- and a long-wavelength mode. This arises physically because small scale structures grow at different rates depending on the large scale environment in which they reside. This induces an off-diagonal term in the two-point statistics in Fourier space that we use as the basis for a quadratic estimator for the large scale field. We demonstrate that this quadratic estimator works well on an N-body simulation of size $(2.5h^{-1}\text{Gpc})^3$. In particular, the quadratic estimator successfully reconstructs the long-wavelength modes using only small-scale information. This opens up novel opportunities to study structure on the largest observable scales.

I. INTRODUCTION

Measuring the distribution of matter on large scales is one of the goals of cosmological surveys [1][2]. The information contained on large scales may provide information about issues ranging from the turnover in the power spectrum (and therefore the total matter density) to the accelerated universe to anomalies observed in the cosmic microwave background to primordial non-Gaussianity. In [3], a method of using delensing with intensity mapping has been proposed to directly measure large scale modes. But generally speaking, direct measurements are difficult because of observational systematic effects, so indirect approaches have been considered. As pointed out by [4], 21cm intensity mapping is one area where due to foregrounds, large-scale (line of sight) modes will be impossible to measure directly.

Small scale structure grows differently in the presence of an large-scale overdensity: it is as if the mean background density is larger than on average. This relation between long- and short-wavelength modes has been discussed in recent years [5][6][7][8][9][10][11][12][13][14][15][16]. The method of using the small scale position-dependent power spectrum to compute the squeezed-limit bispectrum also indicates that small scale perturbations can be used to infer large scale information [17][18]. In order to fully achieve this goal, we construct a quadratic estimator to measure long-wavelength modes indirectly.

Standard perturbation theory (SPT) [19][20][21] identifies the second-order contribution to a short-wavelength mode as a convolution of a short- and a long-wavelength mode. Abstractly, this is similar to cosmic microwave background (CMB) lensing [22][23], where the CMB temperature field has a second-order correction due to the gravitational field along the line of sight. Similarly in our case the short-wavelength mode’s nonlinear terms are related to its large scale environment. The construction of a CMB lensing quadratic estimator makes use of the fact that small scale two-point correlations of CMB temperature modes have off-diagonal terms due to large scale perturbations caused by gravitational lensing. The same statistical feature shows up in our case as well – the off-diagonal terms of the small scale correlations are no longer zero, due to the effect of large scale modes. Thus we can create a quadratic estimator for long-wavelength modes using exactly the same formalism.

We begin with a brief review of SPT up to second-order, build the quadratic estimator, and then assess its detectability. We then apply the estimator to data from a large N-body simulation and demonstrate that it successfully extracts the large scale modes. We use a flat ΛCDM model with Planck Collaboration XVI (2014) [24] cosmological parameters in this work (to match the parameters of the N-body simulation).

II. STANDARD PERTURBATION THEORY

Starting from a perfect pressureless fluid, the nonrelativistic cosmological fluid equations are the continuity, Euler and Poisson equations:

\[
\frac{\partial \bar{\delta}(\vec{x}, \tau)}{\partial \tau} + \bar{\nabla} \cdot [ (1 + \delta(\vec{x}, \tau)) \bar{v}(\vec{x}, \tau) ] = 0 \tag{1}
\]

\[
\left[ \frac{\partial}{\partial \tau} + \bar{v}(\vec{x}, \tau) \cdot \bar{\nabla} \right] \bar{v}(\vec{x}, \tau) = - \frac{da}{d\tau} \frac{\bar{v}(\vec{x}, \tau)}{a} - \bar{\nabla} \Phi \tag{2}
\]

\[
\bar{\nabla}^2 \Phi = 4\pi G a^2 \bar{\rho}_m \delta(\vec{x}, \tau). \tag{3}
\]

Here $a$ is the cosmological scale factor, $\Phi$ is the 3D gravitational potential and $\bar{\rho}_m(a)$ is the mean matter density. These equations fully determine the time evolution of the local density contrast $\delta$ and the peculiar velocity field $\bar{v} = d\vec{x}/d\tau$. We can solve these equations perturbatively...
in Fourier space [25]:

$$\delta(\vec{k}, \tau) = \sum_{n=1}^{\infty} \delta^{(n)}(\vec{k}, \tau) = \sum_{n=1}^{\infty} D_n^1(\tau) \delta_n(\vec{k})$$  \hspace{1cm} (4)$$

$$\theta(\vec{k}, \tau) = \sum_{n=1}^{\infty} \theta^{(n)}(\vec{k}, \tau)$$

$$= - \frac{d \ln D_1(\tau)}{d\tau} \sum_{n=1}^{\infty} D_n^0(\tau) \theta_n(\vec{k})$$  \hspace{1cm} (5)$$

where $D_1$ is the linear growth factor. The first order term $\delta^{(1)}$ corresponds to linear evolution. The linear power spectrum is given by this first order term via averaging over modes in Fourier space:

$$\langle \delta^{(1)}(\vec{k}, \tau) \delta^{(1)}(\vec{k}', \tau) \rangle = (2\pi)^3 \delta_D(\vec{k} + \vec{k}') P_{\text{lin}}(k, \tau)$$  \hspace{1cm} (6)$$

Here $\delta_D$ is the Dirac delta function. Substituting the perturbative series Eq. (4) and Eq. (5) into the Fourier transformed fluid equations Eq. (1)-Eq. (3) leads to an expression for the second-order density contrast:

$$\delta^{(2)}(\vec{k}, \tau) = \int \frac{d^3\vec{k}_1}{(2\pi)^3} F_2(\vec{k}_1, \vec{k} - \vec{k}_1) \delta^{(1)}(\vec{k}_1, \tau) \delta^{(1)}(\vec{k} - \vec{k}_1, \tau)$$  \hspace{1cm} (7)$$

with

$$F_2(\vec{k}_1, \vec{k}_2) = \frac{5}{7} + \frac{2}{7} \frac{(\vec{k}_1 \cdot \vec{k}_2)^2}{k_1^2 k_2^2} + \frac{\vec{k}_1 \cdot \vec{k}_2}{2k_1 k_2} \left[ k_1 + k_2 \right].$$  \hspace{1cm} (8)$$

Note that Eq. (4), Eq. (5) and Eq. (8) are completely accurate only in an Einstein-de Sitter universe and also assuming the case of a pressureless perfect fluid. Nonetheless for related calculations in an ΛCDM universe, the difference is found to be negligible [26], and thus we use the expressions from Eq. (4) and Eq. (8) throughout this work. Using this expression for $\delta^{(2)}$, we can calculate the two-point correlation of two short-wavelength modes $\vec{k}_s$ and $\vec{k}'_s$, in the squeezed limit $\vec{k}_s = \vec{k}_1 + \vec{k}'_1$ with $\vec{k}_s, \vec{k}'_s \gg \vec{k}_l$. Here $\vec{k}_l$ corresponds to a long-wavelength mode, and we suppress the time dependence. To second order,

$$\langle \delta(\vec{k}_s) \delta(\vec{k}'_s) \rangle|_{\vec{k}_s + \vec{k}'_s = \vec{k}_l} = \langle \delta^{(1)}(\vec{k}_s) \delta^{(2)}(\vec{k}'_s) \rangle + \langle \delta^{(2)}(\vec{k}_s) \delta^{(1)}(\vec{k}'_s) \rangle.$$  \hspace{1cm} (9)$$

Substituting Eq. (7) into the first bracket we get:

$$\langle \delta^{(1)}(\vec{k}_s) \delta^{(2)}(\vec{k}'_s) \rangle = \int \frac{d^3\vec{k}_1}{(2\pi)^3} F_2(\vec{k}_1, \vec{k}'_s - \vec{k})$$

$$\times \langle \delta^{(1)}(\vec{k}_s) \delta^{(1)}(\vec{k}'_s - \vec{k}) \delta^{(1)}(\vec{k}) \rangle.$$  \hspace{1cm} (10)$$

When one of the wavenumbers in the 3-point function in Eq. (10) is very small, that mode can be considered as a background mode. The small scale modes evolve in the presence of whatever long wavelength modes happen to be present. Therefore, we can take the long-wavelength mode out of the bracket:

$$\langle \delta^{(1)}(\vec{k}_s) \delta^{(1)}(\vec{k}'_s - \vec{k}) \delta^{(1)}(\vec{k}) \rangle = \langle \delta^{(1)}(\vec{k}_s) \delta^{(1)}(\vec{k}'_s - \vec{k}) \delta^{(1)}(\vec{k}) \rangle \delta^{(1)}(\vec{k}).$$  \hspace{1cm} (11)$$

Delete: The first term on the right occurs when $\vec{k}$ is

FIG. 1. Eq. (11) occurs when $\vec{k} \ll \vec{k}_s, \vec{k}'_s$ or $\vec{k}'_s - \vec{k} \ll \vec{k}_s, \vec{k}'_s$, which corresponds to the two terms shown in this figure.

small and the second when $\vec{k}'_s - \vec{k}$ is small. Using Eq. (6), Eq. (10) then becomes:

$$\int \frac{d^3\vec{k}}{(2\pi)^3} F_2(\vec{k}, \vec{k}'_s - \vec{k}) \langle \delta^{(1)}(\vec{k}_s) \delta^{(1)}(\vec{k}'_s - \vec{k}) \delta^{(1)}(\vec{k}) \rangle$$

$$= \int d^3\vec{k} F_2(\vec{k}, \vec{k}'_s - \vec{k}) \delta_D(\vec{k}_s + \vec{k}'_s - \vec{k}) P_{\text{lin}}(k)$$

$$+ \int d^3\vec{k} F_2(\vec{k}, \vec{k}'_s - \vec{k}) \delta_D(\vec{k}_s + \vec{k}'_s) P_{\text{lin}}(k)$$

$$= 2F_2(-\vec{k}_s, \vec{k}_s) P_{\text{lin}}(k)$$  \hspace{1cm} (12)$$

where we take advantage of the fact that $F_2$ is a symmetric function. Finally we have:

$$\langle \delta(\vec{k}_s) \delta(\vec{k}'_s) \rangle = f(\vec{k}_s, \vec{k}'_s) \delta^{(1)}(\vec{k}_l)$$  \hspace{1cm} (13)$$

with

$$f(\vec{k}_s, \vec{k}'_s) = 2F_2(-\vec{k}_s, \vec{k}_s + \vec{k}'_s) P_{\text{lin}}(k)$$

$$+ 2F_2(-\vec{k}_s, \vec{k}_s + \vec{k}'_s) P_{\text{lin}}(k).$$  \hspace{1cm} (14)$$

This suggests that we can estimate long-wavelength modes using short-wavelength modes. Since the left-hand side of Eq. (13) has only short modes while the right-hand side of it is sensitive to long modes.

III. QUADRATIC ESTIMATOR

We can now construct the quadratic estimator for long-wavelength modes starting from Eq. (13) and summing
over as many pairs as possible with weights that maximize the signal-to-noise. As with the case of CMB lensing, we can write the general form of the estimator by averaging over pairs of short-wavelength modes:

$$\hat{\delta}^{(1)}(\vec{k}_l) = A(\vec{k}_l) \int \frac{d^3 \vec{k}_s}{(2\pi)^3} g(\vec{k}_s, \vec{k}_l) \delta(\vec{k}_s) \delta(\vec{k}_l),$$  \hspace{1cm} (15)

with $g$ being a weighting function, $\vec{k}_s = \vec{k}_l - \vec{k}_s$ and $A$ is the normalization prefactor defined by requiring that $\langle \hat{\delta}^{(1)}(\vec{k}_l) \rangle = \delta^{(1)}(\vec{k}_l)$:

$$A(\vec{k}_l) = \left[ \int \frac{d^3 \vec{k}_s}{(2\pi)^3} g(\vec{k}_s, \vec{k}_l) f(\vec{k}_s, \vec{k}_s) \right]^{-1}$$  \hspace{1cm} (16)

In the absence of shot noise the Gaussian noise is given by:

$$\langle \hat{\delta}^{(1)}(\vec{k}_l) \hat{\delta}^{(1)*}(\vec{k}_l') \rangle = (2\pi)^3 \delta_D(\vec{k}_l - \vec{k}_l') |P_{\text{lin}}(k_l) + N(\vec{k}_l)|$$  \hspace{1cm} (17)

with

$$N(\vec{k}_s) = 2A^2(\vec{k}_l) \times \int \frac{d^3 \vec{k}_s}{(2\pi)^3} g^2(\vec{k}_s, \vec{k}_l - \vec{k}_s) P_{\text{nl}}(k_s) P_{\text{nl}}(|\vec{k}_l - \vec{k}_s|)$$  \hspace{1cm} (18)

where $P_{\text{nl}}$ is the nonlinear power spectrum. Minimizing the noise term we can fix the form of $g$ to be:

$$g(\vec{k}_s, \vec{k}_s') = \frac{f(\vec{k}_s, \vec{k}_s')}{2P_{\text{nl}}(k_s)P_{\text{nl}}(k_s')} = \frac{F_2(-\vec{k}_s, \vec{k}_s + \vec{k}_s') P_{\text{lin}}(k_s) + F_2(-\vec{k}_s', \vec{k}_s + \vec{k}_s') P_{\text{lin}}(k_s')}{P_{\text{nl}}(k_s)P_{\text{nl}}(k_s')}$$  \hspace{1cm} (19)

The noise term reduces simply to $N(\vec{k}_l) = A(\vec{k}_l)$. We find by testing that the value of $N$ is very insensitive to the choice of the lower limit of the integration Eq. (16), since most of the contribution comes from large $k_s$.

Assuming Gaussian noise, the projected detectability of a $P(k_l)$ measurement using the quadratic estimator can be expressed as:

$$\frac{1}{\sigma^2(\vec{k}_l)} = \frac{V k_l^2 \Delta k}{(2\pi)^2} \left[ \frac{P_{\text{lin}}(k_l)}{P_{\text{lin}}(k_l) + N(\vec{k}_l)} \right]^2$$  \hspace{1cm} (20)

where $V$ is the volume of a survey and we compute the detectability for a set of narrow $k_l$-bins each separated by width $\Delta k$. In Fig. 2, we show the projected errors on the long-wavelength power spectrum using this quadratic estimator in a large survey. The current largest scale published measurement of the three dimensional power spectrum is for scale $0.02 h^{-1}$ Mpc$^{-1}$, from [27]. And our Fig. 2 shows that it should be possible to make measurements using our method on scales of 0.002 h Mpc$^{-1}$, which are $\sim 10$ times larger. The upper limit of the $k_s$ integration in Eq. (16) is set to be 0.22 h Mpc$^{-1}$. We will see that this choice of the upper limit is reasonable for our current construction at $z = 0$. Also notice that $P_{\text{lin}}(k_l)$ dominates over $N(k_l)$ in Eq. (20) for this upper limit, thus the projected error bars are only slightly wider than the cosmic variance error bars ($N = 0$).

**FIG. 2.** Long-wavelength power spectrum and its error from Eq. (20) which can be expressed as $P(k_l) \sigma(k_l)$. We assume a toy survey of boxsize $L = 2.5 h^{-1}$ Gpc, thus volume $V = L^3$ and width $\Delta k = 2\pi/L$. Moreover, we set the integration range for $k_s$ from 0.03 h Mpc$^{-1}$ to 0.22 h Mpc$^{-1}$.

**IV. DEMONSTRATION WITH AN N-BODY SIMULATION**

We test the power of the quadratic estimator using data from a cosmological N-body simulation. We use the $z = 0$ snapshot from BigMPDL, one of the Multi-Dark cosmological simulations [28]. The cubical box side length of BigMDPL is $2.5 h^{-1}$ Gpc. We use the dark matter particle data to compute the matter density field, leaving the effect of using galaxies or halos to trace the field [29] to future work.

We use the code nbodykit [30] to measure the Fourier density modes, and Eq. (15) to estimate the long wavelength modes from the measured short wavelength modes. How well the estimator works can be seen from Fig. 3, where we show histograms of the ratio of the estimated mode amplitudes $\hat{\delta}(\vec{k}_l)$ to their true amplitudes $\delta(\vec{k}_l)$ for different values of $k_l$. The two panels show the differences between a short wavelength mode cutoff of $k_s = 0.22 h$ Mpc$^{-1}$ and $k_s = 0.37 h$ Mpc$^{-1}$. Notice that second-order SPT becomes less accurate as shorter wavelengths are used and will produce a bias of our quadratic estimator. The figure shows that individual mode amplitudes are unbiased when $k_{s,\text{max}} = 0.22 h$ Mpc$^{-1}$, while for $k_{s,\text{max}} = 0.37 h$ Mpc$^{-1}$, the results are biased (the center of the ratio is $\sim 20\%$ too high, and the histogram of the polar angle of $\hat{\delta}(\vec{k}_l)/\delta(\vec{k}_l)$ is less peaked at 0.).
FIG. 3. Comparison of measured and predicted Fourier mode amplitudes: histograms of number counts of $|\hat{\delta}(\vec{k}_l)/\delta(\vec{k}_l)|$ for two different $k_s$ integration ranges. $x$-axis shows the magnitude of $|\hat{\delta}(\vec{k}_l)/\delta(\vec{k}_l)|$.

Another way of examining the success of the quadratic estimator is to transform the estimated density field back to real space to form $\hat{\delta}(\vec{x})$ and then compare with the actual large scale density field $\delta(\vec{x})$ in the simulation. The 7 panels in the top two rows of Fig. 4 compare these two fields; each panel is a slice of the full simulations. The bottom panel shows the difference between the estimated and true density fields. It is apparent that the differences are much smaller than the overdensities; equivalently the estimator does an excellent job of extracting the large scale density field.

V. CONCLUSION

In this paper, we have proposed a new and potentially powerful method to measure long-wavelength modes without having to actually measure large scale structure directly. Similarly to this construction, kSZ velocities [31] might also be a good tracer of large-scale modes. We can take the advantage of it’s small scale information and potentially get a better constraint of large-scale modes. We will leave this part to future work.

The estimator works well on an N-Body simulation, so applying this estimator to survey data is the logical next step. Among the issues that must be faced when dealing with a spectroscopic galaxy survey are: galaxy bias, redshift space distortions [32], and light-cone effects. We do not expect any of these to be show-stoppers, so it is tempting to speculate about the possibilities that will open up with this estimate of the large scale density field.

First, we can hope to measure 3D clustering on scales larger than the scale entering the horizon at matter-radiation equality without worrying about large-scale systematic effects. General relativistic effects are strongest on large scales (e.g., [33]), and these could be detected. There is evidence of large scale anomalies, in the CMB, that could be confronted with maps of large scale structure obtained with this estimator. One physical mechanism that has been proposed as a possible explanation for the deficit in the large-angle CMB temperature correlations is a suppression of primordial power on $\sim$ Gpc scales [34]. It would be very useful to verify if this new physics is also present in other probes of large-scale structure. Primordial non-Gaussianity generated by inflation leaves an imprint on the largest scales. There is even the possibility of cross-correlating the large-scale matter field with the CMB itself to extract information about the longest wavelength modes in the universe. Since the current epoch of acceleration is a large-scale, late-time effect, there is the possibility of learning about the mechanism responsible for acceleration. Although, as mentioned above, challenges remain, there is also the possibility of using even smaller wavelength modes in our estimator by going to higher order in perturbation theory.

ACKNOWLEDGMENTS

We thank Adam Solomon, Duncan Campbell, Fabian Schmidt and Lam Hui for resourceful discussions. This work is supported by U.S. Dept. of Energy contract DE-SC0019248 and NSF AST-1009193. The BigMDPL simulation was performed at LRZ Munich within the PRACE project pr86bu. The CosmoSim database (www.cosmosim.org) providing the file access is a service by the Leibniz-Institute for Astrophysics Potsdam (AIP).
FIG. 4. Comparison of the true density field in the BigMPDL simulation ($\delta(\vec{x})$) computed using the directly measured large-scale modes, top row) and the density field from the quadratic estimator ($\hat{\delta}(\vec{x})$, middle row). The bottom row shows their difference. Each panel represents a slice through the simulation volume, $2.5\, h^{-1}\text{Gpc}$ wide, and one cell ($0.36\, h^{-1}\text{Gpc}$) thick. The upper limit of $\vec{k}_s$ is $0.22\, h\text{Mpc}^{-1}$. 
$P_{\text{lin}}(k) \left[ (h^{-1}\text{Mpc})^3 \right]$
