REWRITE TITLE??
New Probes of Large Scale Structure

Peikai Li,[1,2] Rupert A. C. Croft,[1,2] and Scott Dodelson[1,2]

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

(Dated: May 21, 2021)

This is the second paper in a series where we propose a method of indirectly measuring large scale structure using information from small scale perturbations. The idea is to build a quadratic estimator from small scale modes that provides a map of structure on large scales. We demonstrated in the first paper that the quadratic estimator works well on a dark-matter-only N-body simulation at a snapshot of $z = 0$. Here we generalize the theory to the case of a light cone halo catalog with a non-cubic region taken into consideration. We successfully apply the generalized version of the quadratic estimator to a light cone halo catalog of an N-body simulation of size $\sim 15.03\,(h^{-1}\,\mathrm{Gpc})^3$. The most distant point in the light cone is at a redshift of 1.42, which indicates that we might be able to apply our method to next generation galaxy surveys.

I. INTRODUCTION

Directly measuring the distribution of matter on large scales is extremely difficult right now as pointed out, e.g., by [1]. The attempt to use small scale perturbations to infer large scale information has been frequently discussed in recent years [2][3][4][5][6]. In our first work [7], we proposed a method of indirectly measuring large scale structure using the small scale density contrast. Physically, long- and short-wavelength modes are correlated because small scale modes will grow differently depending on the large scale structure they reside in. This phenomenon leaves a signature in Fourier space: the two-point statistics of short-wavelength matter density modes will have non-zero off-diagonal terms proportional to long-wavelength modes. This is our starting point for constructing the quadratic estimator for long-wavelength modes. We tested the power of the quadratic estimator using a dark-matter-only catalog from an N-body simulation in the first paper. In this work, we generalize Ref. [7] to account for two main effects that must be accounted for before applying the techniques to upcoming surveys [8][9][10]: (i) we observe galaxies, not the dark matter field; (ii) we observe in a non-cubic light cone not a snapshot. We should be able to apply our method to real surveys in near future.

First we need to account for galaxy bias [11][12]. Galaxy bias is a term relating the galaxy number density contrast to the matter density contrast [13][14]. We consider up to second order bias, as done in recent treatments of galaxy surveys. Meanwhile, analytically the generalization to even higher order biases is straightforward. We adopt the most commonly used second-order galaxy bias model and assume all the bias parameters to be a constant while we are considering a large volume across wide redshift range.

Observationally a galaxy catalog will be in a light cone [15] instead of a snapshot. The typical treatment is to cut a light cone into several thin redshift bins [16] and analyze the properties within each bin. Doing this, though, loses information about the long-wavelength modes along line of sight. Thus in this paper we propose a method of considering all the galaxies in a light cone together, using the famous Feldman-Kaiser-Peacock (FKP) estimator [17] to account for the evolution of the galaxy number density. Using an octant volume (which can be applied to a more generalized shape), we test the quadratic estimator for long-wavelength modes using information from non-zero off-diagonal terms as in [7]. Also notice that the FKP description corresponds to the monopole part of the estimator in redshift space (e.g. the Yamamoto estimator [18][19]). So our formalism is able to reconstruct the large scale monopole power spectrum which is the main goal in learning the large scale matter distributions of the 3D universe.

We begin with a brief review of the formalism developed in Ref. [7], then present our treatment of the galaxy number density contrast in a light cone and then build the quadratic estimator. Finally we apply the estimator to a light-cone halo N-body simulations and successfully extract large scale modes accounting for these effects.

II. REVIEW OF QUADRATIC ESTIMATOR

We first review the construction of a quadratic estimator of a dark-matter-only catalog [7] before moving to a halo catalog. Starting from the perturbative series of the matter density contrast in Fourier space up to second order [20][21]:

$$\begin{aligned}\delta_{\mathrm{m}}(\vec{k};z) &= \delta_{\mathrm{m}}^{(1)}(\vec{k};z)+\delta_{\mathrm{m}}^{(2)}(\vec{k};z)\\ &= \delta_{\mathrm{m}}^{(1)}(\vec{k};z)\\ &+\int \frac{d^3\vec{k}'}{(2\pi)^3}\delta_m^{(1)}(\vec{k}';z)\delta_m^{(1)}(\vec{k}-\vec{k}';z)F_2(\vec{k}',\vec{k}-\vec{k}') \quad (1)\end{aligned}$$

where "m" stands for matter, the superscript $i=1,2,\cdots$ corresponds to the i-th order term of the expansion, and $\delta_{\mathrm{m}}(\vec{k};z)$ is the full Fourier space matter density contrast in a snapshot when the redshift was equal to z. And the kernel F_2 is a function particularly insensitive to the choice of cosmological parameters in a dark-energy-dominated universe [22]:

$$F_2(\vec{k}_1,\vec{k}_2)=\frac{5}{7}+\frac{2}{7}\frac{(\vec{k}_1\cdot\vec{k}_2)^2}{k_1^2k_2^2}+\frac{\vec{k}_1\cdot\vec{k}_2}{2k_1k_2}\left[\frac{k_1}{k_2}+\frac{k_2}{k_1}\right]. \quad (2)$$

Thus, $\delta_{\mathrm{m}}^{(1)}$ is the linear density contrast, and the second order term $\delta_{\mathrm{m}}^{(2)}$ can be written as a convolution-like integral using the first order term.

When evaluating the two-point function of the full density contrast, cross-terms appear. For example, $\langle\delta_{\mathrm{m}}^{(1)}(\vec{k};z)\delta_{\mathrm{m}}^{(2)}(\vec{k}';z)\rangle$ is proportional to $\delta_{\mathrm{m}}^{(1)}(\vec{k}+\vec{k}';z)$ if both $\vec{k}$ and $\vec{k}'$ correspond to short wavelengths but their sum is small (long wavelength). Explicitly, keeping terms up to second order,

$$\langle\delta_{\mathrm{m}}(\vec{k}_s;z)\delta_{\mathrm{m}}(\vec{k}_s';z)\rangle = f(\vec{k}_s,\vec{k}_s';z)\delta_{\mathrm{m}}^{(1)}(\vec{k}_l;z). \quad (3)$$

Here $\vec{k}_s$ and $\vec{k}_s'$ are two short-wavelength modes and $\vec{k}_l$ is a long-wavelength mode ($\vec{k}_s,\vec{k}_s'\gg\vec{k}_l$). They satisfy the squeezed-limit condition $\vec{k}_s+\vec{k}_s'=\vec{k}_l$ and f is given by:

$$\begin{aligned}f(\vec{k}_s,\vec{k}_s';z) &= 2F_2(-\vec{k}_s,\vec{k}_s+\vec{k}_s')P_{\mathrm{m}}^{(1)}(k_s;z)\\ &+2F_2(-\vec{k}_s',\vec{k}_s+\vec{k}_s')P_{\mathrm{m}}^{(1)}(k_s';z). \quad (4)\end{aligned}$$

Here $P_{\mathrm{m}}^{(1)}$ is the linear matter power spectrum. Eq. (3) indicates that we can estimate the long-wavelength modes using small scale information with the following minimum variance quadratic estimator:

$$\hat{\delta}_{\mathrm{m}}^{(1)}(\vec{k}_l;z)=A(\vec{k}_l;z)\int\frac{d^3\vec{k}_s}{(2\pi)^3}g(\vec{k}_s,\vec{k}_s';z)\delta_{\mathrm{m}}(\vec{k}_s;z)\delta_{\mathrm{m}}(\vec{k}_s';z) \quad (5)$$

with $\vec{k}_s'=\vec{k}_l-\vec{k}_s$. The normalization factor A is defined by requiring that $\langle\hat{\delta}_{\mathrm{m}}^{(1)}(\vec{k}_l;z)\rangle=\delta_{\mathrm{m}}^{(1)}(\vec{k}_l;z)$, and the weighting function g is obtained by minimizing the noise. They can be expressed as:

$$\begin{aligned}A(\vec{k}_l;z) &= \left[\int\frac{d^3\vec{k}_s}{(2\pi)^3}g(\vec{k}_s,\vec{k}_s';z)f(\vec{k}_s,\vec{k}_s';z)\right]^{-1}\\ g(\vec{k}_s,\vec{k}_s';z) &= \frac{f(\vec{k}_s,\vec{k}_s';z)}{2P_{\mathrm{m}}(k_s;z)P_{\mathrm{m}}(k_s';z)} \quad (6)\end{aligned}$$

where P_{m} is the nonlinear matter power spectrum. With this choice of the weighting function g, the noise on the estimator $N(\vec{k}_l;z)=A(\vec{k}_l;z)$ if non-gaussian terms in the four-point function are neglected. Therefore, the projected detectability of a power spectrum measurement using this quadratic estimator can be written as:

$$\frac{1}{\sigma^2(k_l;z)}=\frac{Vk_l^2\Delta k}{(2\pi)^2}\left[\frac{P_{\mathrm{m}}^{(1)}(k_l;z)}{P_{\mathrm{m}}^{(1)}(k_l;z)+A(k_l;z)}\right]^2, \quad (7)$$

where V is the volume of a survey and Δk is the width of long-wavelength mode bins. We also take advantage of the fact that $A(\vec{k}_l;z)$ does not depend on the direction of the long mode $\vec{k}_l$.

III. GENERALIZATION: BIAS MODEL AND FKP ESTIMATOR

Similar to Eq. (1), we use the most commonly used Eulerian non-linear and non-local galaxy bias model np to second-order first proposed by [23]:

$$\begin{aligned}\delta_{\mathrm{g}}(\vec{k};z) &= b_1\delta_m^{(1)}(\vec{k};z)\\ &+\int\frac{d^3\vec{k}'}{(2\pi)^3}\delta_m^{(1)}(\vec{k}';z)\delta_m^{(1)}(\vec{k}-\vec{k}';z)\mathcal{F}_2(\vec{k}',\vec{k}-\vec{k}')\ . \quad (8)\end{aligned}$$

here "g" denotes galaxy, and b_1 is the linear bias parameter relating galaxy and the matter density contrasts. The kernel $\mathcal{F}_2$ is given by:

$$\mathcal{F}_2(\vec{k}_1,\vec{k}_2)=b_1F_2(\vec{k}_1,\vec{k}_2)+\frac{b_2}{2}+\frac{b_{s^2}}{2}S_2(\vec{k}_1,\vec{k}_2) \quad (9)$$

with S_2 given by:

$$S_2(\vec{k}_1,\vec{k}_2)=\frac{(\vec{k}_1\cdot\vec{k}_2)^2}{k_1^2k_2^2}-\frac{1}{3}\ . \quad (10)$$

Comparing the perturbative series of the galaxy density contrast Eq. (8) with that of the matter density contrast Eq. (1), we see that the difference with the first order term is an extra coefficient b_1. And the second order term is almost the same with a simple replacement of the kernel function. This implies that we can easily generalize to the case of a galaxy catalog in a snapshot. In the case of a galaxy catalog of a light cone, usuallyt we use the Feldman-Kaiser-Peacock (FKP) estimator to estimate the observed galaxy power spectrum [17]:

$$F(\vec{r})\equiv I^{-1/2}w_{\mathrm{FKP}}(\vec{r})[n_g(\vec{r})-\alpha n_s(\vec{r})] \quad (11)$$

with

$$I \equiv \int_V d^3\vec{r} w_{\rm FKP}^2(\vec{r})\langle n_g\rangle^2(\vec{r}) \ . \tag{12}$$

Here n_g is the observed galaxy number density and n_s is the corresponding synthetic catalog. The constant α is the ratio of the observed number density to the synthetic catalog's number density. The FKP weight $w_{\rm FKP}(\vec{r})$ is usually defined as:

$$w_{\rm FKP}(\vec{r}) = \frac{1}{1+\langle n_g\rangle(\vec{r})P_0} \tag{13}$$

Notice in real surveys we will have other types of weights [13][24], which can be easily included by the formalism in this section. The FKP estimator $F(\vec{r})$ is related to the observed galaxy power spectrum $P_{g,\rm obs}(\vec{k})$ by considering the following expectation value (diagonal elements) in Fourier space:

$$\begin{aligned}\langle|F(\vec{k})|^2\rangle &= \int\frac{d^3\vec{k}'}{(2\pi)^3}P_g(k';z_{\rm eff})|W(\vec{k}-\vec{k}')|^2 + P_{\rm shot\ noise}\\ &= \langle|\delta_{g,W}(\vec{k};z_{\rm eff})|^2\rangle + P_{\rm shot\ noise} \equiv P_{g,\rm obs}(\vec{k})\end{aligned} \tag{14}$$

here $z_{\rm eff}$ is the effective redshift of the whole light cone. Also the window function $W(\vec{k})$, the shot noise spectrum $P_{\rm shot\ noise}$ and the windowed galaxy density contrast $\delta_{g,W}(\vec{k};z)$ are given respectively by[1]:

$$W(\vec{k}) = I^{-1/2}\int_V d^3\vec{r}\langle n_g\rangle(\vec{r})w_{\rm FKP}(\vec{r})e^{-i\vec{k}\cdot\vec{r}} \tag{15}$$

$$P_{\rm shot\ noise} = (1+\alpha)I^{-1}\int_V d^3\vec{r}\langle n_g\rangle(\vec{r})w_{\rm FKP}^2(\vec{r}) \tag{16}$$

$$\delta_{g,W}(\vec{k}) \equiv \int\frac{d^3\vec{k}'}{(2\pi)^3}\delta_g(\vec{k}')W(\vec{k}-\vec{k}') \ . \tag{17}$$

We want to calculate the off-diagonal term of the FKP estimator $F(\vec{k})$ given the fact that $F(\vec{k})$ is the observable of a galaxy light cone survey instead of $\delta_{g,W}$. Notice that the two point functions of $n_g(\vec{r})-\alpha n_s(\vec{r})$ can be written as [17]:

$$\begin{aligned}&\langle[n_g(\vec{r})-\alpha n_s(\vec{r})][n_g(\vec{r}')-\alpha n_s(\vec{r}')]\rangle\\ &= \langle n_g\rangle(\vec{r})\langle n_g\rangle(\vec{r}')\xi_g(\vec{r}-\vec{r}') + (1+\alpha)\langle n_g\rangle(\vec{r})\delta_{\rm D}(\vec{r}-\vec{r}')\end{aligned} \tag{18}$$

Assuming the squeezed limit $\vec{k}_s+\vec{k}'_s = \vec{k}_l$ and use the expression above, we can write the off-diagonal term as:

$$\begin{aligned}&\langle F(\vec{k}_s)F(\vec{k}'_s)\rangle\\ &= \langle\delta_{g,W}(\vec{k}_s;z_{\rm eff})\delta_{g,W}(\vec{k}'_s;z_{\rm eff})\rangle + Q_{\rm shot\ noise}(\vec{k}_l)\end{aligned} \tag{19}$$

[1] One interesting thing to notice here is: in the original paper of the FKP estimator [17] and almost every work after, $W(\vec{k})$ is defined as the complex conjugate of this work. In their cases, $W(\vec{k})$ only shows up in the form of $|W(\vec{k})|^2$ so it won't make a difference. However in this work we demonstrate using simulations that it should be $e^{-i\vec{k}\cdot\vec{r}}$ instead of $e^{+i\vec{k}\cdot\vec{r}}$ in the integrand, same as the definition of the Fourier transform.

with the"off-diagonal shot noise" $Q(\vec{k}_l)$ given by:

$$Q(\vec{k}_l) = (1+\alpha)I^{-1}\int_V d^3\vec{r}\langle n_g\rangle(\vec{r})w_{\rm FKP}^2(\vec{r})e^{i\vec{k}_l\cdot\vec{r}} \ . \tag{20}$$

The two point function $\langle\delta_{g,W}(\vec{k}_s;z_{\rm eff})\delta_{g,W}(\vec{k}'_s;z_{\rm eff})\rangle$ up to second order can be simply expressed as:

$$\langle\delta_{g,W}\delta'_{g,W}\rangle = \langle\delta^{(1)}_{g,W}\delta'^{(1)}_{g,W}\rangle + \langle\delta^{(1)}_{g,W}\delta'^{(2)}_{g,W}\rangle + \langle\delta^{(2)}_{g,W}\delta'^{(1)}_{g,W}\rangle \tag{21}$$

by defining $\delta_{g,W} \equiv \delta_{g,W}(\vec{k}_s;z_{\rm eff})$ and $\delta'_{g,W} \equiv \delta_{g,W}(\vec{k}'_s;z_{\rm eff})$. Notice one major difference here is that, for a non-cubic region, the leading order term would also be non-zero unlike last section II:

$$\begin{aligned}&\langle\delta^{(1)}_{g,W}(\vec{k}_s;z_{\rm eff})\delta^{(1)}_{g,W}(\vec{k}'_s;z_{\rm eff})\rangle\\ &= \int\frac{d^3\vec{k}}{(2\pi)^3}\int\frac{d^3\vec{k}'}{(2\pi)^3}W(\vec{k}-\vec{k}_s)W(\vec{k}'-\vec{k}'_s)\\ &\times b_1^2(2\pi)^3\delta_{\rm D}(\vec{k}-\vec{k}')P^{(1)}_m(k;z_{\rm eff})\\ &= b_1^2\int\frac{d^3\vec{k}}{(2\pi)^3}W(\vec{k}-\vec{k}_s)W(-\vec{k}-\vec{k}'_s)P^{(1)}_m(k;z_{\rm eff})\end{aligned} \tag{22}$$

where $\delta_{\rm D}$ is the Dirac delta function. This term would vanish since in the case of a cube, $W(\vec{k})$ would be close to a Dirac delta function. While for a non-cubic region, this term is no longer zero. Notice that this leading order term can be fully determined numrically.

Using the expressions of Eq. (8) and Eq. (17), we can compute the second order two-point correlation of two short-wavelength modes $\delta_{g,W}(\vec{k}_s;z_{\rm eff})$ and $\delta_{g,W}(\vec{k}'_s;z_{\rm eff})$. Use $\langle\delta^{(1)}_{g,W}(\vec{k}_s;z_{\rm eff})\delta^{(2)}_{g,W}(\vec{k}'_s;z_{\rm eff})\rangle$ as an example:

$$\begin{aligned}&\langle\delta^{(1)}_{g,W}(\vec{k}_s;z_{\rm eff})\delta^{(2)}_{g,W}(\vec{k}'_s;z_{\rm eff})\rangle\\ &= b_1\int\frac{d^3\vec{k}}{(2\pi)^3}\int\frac{d^3\vec{k}'}{(2\pi)^3}W(\vec{k}_s-\vec{k})W(\vec{k}'_s-\vec{k}')\\ &\times\langle\delta^{(1)}_m(\vec{k};z_{\rm eff})\delta^{(2)}_g(\vec{k}';z_{\rm eff})\rangle\end{aligned} \tag{23}$$

Notice that we have compute the bracket $\langle\delta^{(1)}_m(\vec{k};z_{\rm eff})\delta^{(2)}_g(\vec{k}';z_{\rm eff})\rangle$ before in our last work [7], with F_2 replaced by $\mathscr{F}_2$, and the result gives:

$$\begin{aligned}&\langle\delta^{(1)}_m(\vec{k};z_{\rm eff})\delta^{(2)}_g(\vec{k}';z_{\rm eff})\rangle\\ &= 2\mathscr{F}_2(-\vec{k},\vec{k}+\vec{k}')P^{(1)}_m(k;z_{\rm eff})\delta^{(1)}_m(\vec{k}+\vec{k}';z_{\rm eff})\end{aligned} \tag{24}$$

Thus we can further express the bracket as:

$$\begin{aligned}&\langle\delta^{(1)}_{g,W}(\vec{k}_s;z_{\rm eff})\delta^{(2)}_{g,W}(\vec{k}'_s;z_{\rm eff})\rangle\\ &= 2b_1\int\frac{d^3\vec{k}}{(2\pi)^3}\int\frac{d^3\vec{k}'}{(2\pi)^3}W(\vec{k}_s-\vec{k})W(\vec{k}'_s-\vec{k}')\\ &\times\mathscr{F}_2(-\vec{k},\vec{k}+\vec{k}')P^{(1)}_m(k;z_{\rm eff})\delta^{(1)}_m(\vec{k}+\vec{k}';z_{\rm eff})\end{aligned} \tag{25}$$

In order to extract a term proportional to $\delta^{(1)}_{g,W}(\vec{k}_l;z_{\rm eff})$ which has the expression:

$$\delta^{(1)}_{g,W}(\vec{k}_l;z_{\rm eff}) = b_1\int\frac{d^3\vec{k}}{(2\pi)^3}\delta^{(1)}_m(\vec{k};z_{\rm eff})W(\vec{k}_l-\vec{k}). \tag{26}$$

We want to extract $W(\vec{k}'_s - \vec{k}')$ out of the integral using the fact that, given a large enough volume, $W(\vec{k})$ is peaked at $\vec{k} = 0$ and also[2]:

$$\int \frac{d^3\vec{k}}{(2\pi)^3} W(\vec{k}) = W(\vec{r} = 0) \equiv C \tag{27}$$

Thus we have the following approximations:

$$\begin{aligned}
&\langle \delta^{(1)}_{g,W}(\vec{k}_s; z_{\rm eff}) \delta^{(2)}_{g,W}(\vec{k}'_s; z_{\rm eff}) \rangle \\
&= 2b_1 \int \frac{d^3\vec{k}}{(2\pi)^3} \int \frac{d^3\vec{k}'}{(2\pi)^3} W(\vec{k}_s - \vec{k} + \vec{k}') W(\vec{k}'_s - \vec{k}') \\
&\times \mathcal{F}_2(-\vec{k} + \vec{k}', \vec{k}) P^{(1)}_m(|\vec{k} - \vec{k}'|; z_{\rm eff}) \delta^{(1)}_m(\vec{k}; z_{\rm eff}) \\
&\simeq 2C\mathcal{F}_2(-\vec{k}_s, \vec{k}_s + \vec{k}'_s) P^{(1)}_m(k_s; z_{\rm eff}) \\
&\times b_1 \int \frac{d^3\vec{k}}{(2\pi)^3} \delta^{(1)}_m(\vec{k}; z) W(\vec{k}_s + \vec{k}'_s - \vec{k}) \\
&= 2C\mathcal{F}_2(-\vec{k}_s, \vec{k}_s + \vec{k}'_s) P^{(1)}_m(k_s; z_{\rm eff}) \delta^{(1)}_{g,W}(\vec{k}_l; z_{\rm eff})
\end{aligned} \tag{28}$$

where in the first step, we use a redefinition of dummy variables. With this calculation above, we can recover the long-wavelength modes from the off-diagonal two-point functions of short-wavelength modes:

$$\begin{aligned}
&\langle F(\vec{k}_s) F(\vec{k}'_s) \rangle - Q_{\text{shot noise}}(\vec{k}_l) \\
&- b_1^2 \int \frac{d^3\vec{k}}{(2\pi)^3} W(\vec{k} - \vec{k}_s) W(-\vec{k} - \vec{k}'_s) P^{(1)}_m(k; z_{\rm eff}) \\
&= f(\vec{k}_s, \vec{k}'_s; z_{\rm eff}) \delta^{(1)}_{g,W}(\vec{k}_l; z_{\rm eff})
\end{aligned} \tag{29}$$

with

$$\begin{aligned}
f(\vec{k}_s, \vec{k}'_s; z_{\rm eff}) &= 2C\mathcal{F}_2(-\vec{k}_s, \vec{k}_s + \vec{k}'_s) P^{(1)}_m(k_s; z_{\rm eff}) \\
&+ 2C\mathcal{F}_2(-\vec{k}'_s, \vec{k}_s + \vec{k}'_s) P^{(1)}_m(k'_s; z_{\rm eff}).
\end{aligned} \tag{30}$$

Notice that the f is almost identical to the f function in section II, simply with a replacement of the F_2 function and an extra coefficient C. Then the quadratic estimator can be similarly formed as:

$$\begin{aligned}
&\hat{\delta}^{(1)}_{g,W}(\vec{k}_l; z_{\rm eff}) = \mathcal{A}(\vec{k}_l; z_{\rm eff}) \int \frac{d^3\vec{k}_s}{(2\pi)^3} g(\vec{k}_s, \vec{k}'_s; z_{\rm eff}) \\
&\times \Bigg[F(\vec{k}_s) F(\vec{k}'_s) - Q_{\text{shot noise}}(\vec{k}_l) \\
&- b_1^2 \int \frac{d^3\vec{k}}{(2\pi)^3} W(\vec{k} - \vec{k}_s) W(-\vec{k} - \vec{k}'_s) P^{(1)}_m(k; z_{\rm eff}) \Bigg]
\end{aligned} \tag{31}$$

with $\vec{k}'_s = \vec{k}_l - \vec{k}_s$ and g being the weighting function. Notice here the only difference is that we subtract off the non-zero leading order terms due to the non-cubic shape of the galaxy catalog, and these two terms can be calculated numerically. By requiring that $\langle \hat{\delta}^{(1)}_{g,W}(\vec{k}_l; z_{\rm eff}) \rangle = \hat{\delta}^{(1)}_{g,W}(\vec{k}_l; z_{\rm eff})$ we can similarly determine the normalization function $\mathcal{A}$:

$$\mathcal{A}(\vec{k}_l; z_{\rm eff}) = \left[\int \frac{d^3\vec{k}_s}{(2\pi)^3} g(\vec{k}_s, \vec{k}'_s; z_{\rm eff}) f(\vec{k}_s, \vec{k}'_s; z_{\rm eff}) \right]^{-1} \tag{32}$$

Similar to our last work, by minimizing the noise we get the expression for the weighting function g:

$$\begin{aligned}
g(\vec{k}_s, \vec{k}'_s; z_{\rm eff}) &= \frac{f(\vec{k}_s, \vec{k}'_s; z_{\rm eff})}{2P_{g,\rm obs}(\vec{k}_s) P_{g,\rm obs}(\vec{k}'_s)} \\
&= C\Bigg[\frac{\mathcal{F}_2(-\vec{k}_s, \vec{k}_s + \vec{k}'_s) P^{(1)}_m(k_s; z_{\rm eff})}{P_{g,\rm obs}(\vec{k}_s) P_{g,\rm obs}(\vec{k}'_s)} \\
&\quad + \frac{\mathcal{F}_2(-\vec{k}'_s, \vec{k}_s + \vec{k}'_s) P^{(1)}_m(k'_s; z_{\rm eff})}{P_{g,\rm obs}(\vec{k}_s) P_{g,\rm obs}(\vec{k}'_s)} \Bigg]
\end{aligned} \tag{33}$$

here $P_{g,\rm obs}$ is the full observed galaxy power spectrum. With this choice of g the noise term $\mathcal{N}$ is identical to the normalization factor $\mathcal{A}$. And the projected detectability is defined similarly as Eq. (7):

$$\frac{1}{\sigma(k_l; z_{\rm eff})^2} = \frac{V k_l^2 \Delta k}{(2\pi)^2} \left[\frac{P^{(1)}_m(k_l; z_{\rm eff})}{P^{(1)}_m(k_l; z_{\rm eff}) + \mathcal{A}(k_l; z_{\rm eff})} \right]^2 \tag{34}$$

Using the quadratic estimator Eq. (31) we can use small scale information of the non-cubic whole light cone to infer the large scale field of the windowed galaxy density contrast $\delta_{g,W}(\vec{r})$.

IV. DEMONSTRATION WITH AN N-BODY SIMULATION

We use the MICE Grand Challenge light cone N-body simulation (MICE-GC) [25][26][27] to demonstrate the power of the estimator in a light cone. The catalog contains one octant of the full sky up to $z = 1.42$ (comoving distance $3062\, h^{-1}$ Mpc) without simulation box repetition, as shown in Fig. 1. This simulation used a flat ΛCDM model with cosmological parameters $\Omega_{\rm m} = 0.25$, $\sigma_8 = 0.8$, $n_{\rm s} = 0.95$, $\Omega_{\rm b} = 0.044$, $\Omega_\Lambda = 0.75$, $h = 0.7$.

We consider the halo catalog in this light cone with halo masses between $2.2 \times 10^{12} h^{-1} M_\odot < M < 10^{14} h^{-1} M_\odot$, which is a quite wide mass bin. We obtained similar results in other mass bins as well. The effective redshift of this light cone is $z_{\rm eff} = 0.76$. We assume the bias parameter b_1 and b_2 to be free parameters of the model, and use FAST-PT [28] to determine the bias parameters to be:

$$\begin{aligned}
b_1 &= 1.88 \\
b_2 &= 3.13\,.
\end{aligned} \tag{35}$$

The remaining bias parameter b_{s^2} can be constrained by assuming the bias model is local in Lagrangian space [29]:

$$b_{s^2} = -\frac{4}{7}(b_1 - 1) = -0.50\,. \tag{36}$$

[2] The result would remain the same even if the origin $\vec{r} = 0$ is outside the region V.

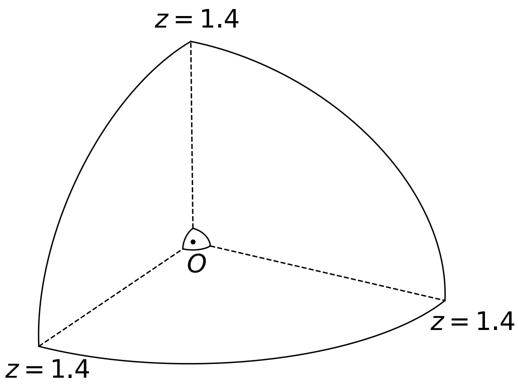

FIG. 1. The survey region of the MICE-GC simulation, which is an octant. Notice due to some technical reasons there are no galaxies near the origin O, so a small octant is removed from the survey region.

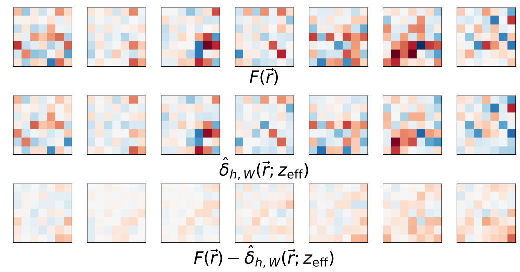

FIG. 2. Comparison of the true real space FKP estimator field in the MICE-GC simulation ($F(\vec{r})$ computed using the directly measured large scale modes, top row) and the windowed halo density field from the quadratic estimator ($\hat{\delta}_{h,W}(\vec{r})$, middle row). The bottom row shows their difference. Each cell is $(0.44\,h^{-1}\mathrm{Mpc})$ thick. And the upper limit of $\vec{k}_s$ is $0.48\,h\,\mathrm{Mpc}^{-1}$.

We use the quadratic estimator Eq. (31) to get the reconstructed windowed galaxy density field and transform them back into real space. Then compare this estimated result with the directly measured FKP estimator in real space in Fig. **??**. **We use the information of small scale modes up to $k_{s,\max} = 0.48\,h\,\mathrm{Mpc}^{-1}$. Since we observe that the result would get much better if we include more small scale modes.** We also plot in Fig. 3 the directly measured large scale halo power spectrum with the cosmic variance versus the estimated large scale power spectrum using Eq. (31) with detectability given by Eq. (34). We see that our quadratic estimator gives a good estimation of the linear matter power spectrum and the variance of the estimation is only slightly larger than the cosmic variance.

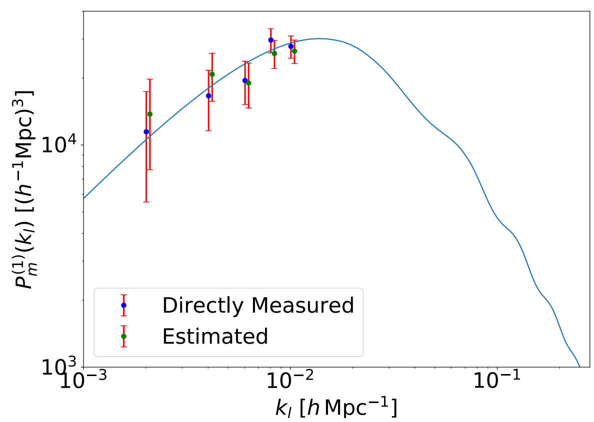

FIG. 3. Inferred linear matter power spectrum from direct measurement versus indirect estimation using our quadratic estimator, both from the MICE-GC light-cone halo catalog. The error for the direct measurement can be derived from the cosmic variance. And the error for the indirect estimation can be expressed as $P_m^{(1)}(k_l)\sigma(k_l)$, where $\sigma(k_l)$ is from Eq. (34) after scaling.

Notice in Fig. **??** we cannot have a direct measurement of the $\delta_{g,W}$ field, but both of them encode almost the same large scale information of the light cone catalog. Since the large scale we are observing is about $10^{-3}h\,\mathrm{Mpc}^{-1} < k_l < 10^{-2}h\,\mathrm{Mpc}^{-1}$ where the magnitude of the observed power spectrum Fig. 3 is much greater than the shot noise term where in this case $P_{\text{shot noise}} \simeq 1000\,(h^{-1}\mathrm{Mpc})^3$. From Fig. **??** we see that our quadratic estimator is still able to extract large scale information, especially large over- and under-density cells are well reconstructed. The difference seems slightly larger when we go to higher redshift (corresponding to the panels on the right) and is worst on the very right panel.

V. CONCLUSION

A. Summary

In prior work [7] we have shown that the amplitude and phase of large scale density fluctuations can be recovered by applying a quadratic estimator to measurements of small scale Fourier modes and their correlations. In this paper we extend that work (which was limited to a matter density field at a single instance in cosmic time) to a light cone galaxy catalog in order to make it applicable to observational data. All extensions are tested on appropriate mock survey datasets derived from N-body simulations.

B. Discussion

Our formalism includes the major effects that are relevant for an application to observational data. There are some minor aspects however which will need to be dealt with when this occurs. One is the fact that we have tested on homogeneous mock surveys, when real observations will include masked data (to account for bright stars for example), and a potentially more complex window function. In a spectroscopic survey, the observed distribution of galaxies is distorted and squashed when we use their redshift as an indicator of their radial distance due to galaxies' peculiar velocity. This effect is known as redshift space distortion [30] and the FKP formalism corresponds to the monopole moment in redshift space. And we left the generalization to include higher order power spectrum multipoles (quadrapole, hexadecapole) to future work.

At present the large scale limitations on direct measurement of clustering are observational systematics (e.g., [31]). These include angular variations in obscuration, seeing, sky brightness, colors, extinction and magnitude errors. Because these result in relatively small modulations of the measured galaxy density, they will affect large scale modes most importantly, hence the utility of our indirect measurements of clustering on these scales. Quantification of these effects on the scales for which we do measure clustering will still be needed though. It will be also be instructive to apply large scale low amplitude modulations to our mock surveys in order to test how well the quadratic estimator works with imperfect data. Even small scale issues with clustering, such as fiber collisions [32] could affect our reconstruction, depending on how their effects propagate through the quadratic estimator.

Observational datasets exist at present which could be used to carry out measurements using our methods. These include the SDSS surveys BOSS [33] and eBOSS [34] (both luminous red galaxies and emission line galaxies). Substantial extent in both angular coordinates and redshift are necessary, so that deep but narrow surveys such as VIMOS [35] or DEEP2 [36] would not be suitable. In the near future, the available useful data will increase rapidly with the advent of WEAVE [37] and DESI [10]. Space based redshift surveys with EUCLID [38] and WFIRST [9] will expand the redshift range, and SPHEREx [39], due for launch even earlier will offer maximum sky coverage, and likely the largest volume of all.

In order to model what is expected from all these datasets, the effective range of wavelengths used in the reconstruction of large scale modes should be considered. Surveys covering large volumes but with low galaxy number density will have large shot noise contributions to density fluctuations, and this will limit the range of scales that can be used. For example, in our present work we have successfully tested number densities of $\sim 3 \times 10^{-3}$ galaxies per $(\mathrm{Mpc/h})^3$. Surveys such as the eBOSS quasar redshift survey [40] with a number density ~ 100 times lower will not be useful, for example.

Once an indirect measurement of large scale modes has been made from an observational dataset, there are many different potential applications. We can break these up into two groups, involving the power spectrum itself, and the map (and statistics beyond $P_{\mathrm{m}}(k)$) which can be derived from it.

First, because of the effect of observational systematics mentioned above, and the fact the our indirect estimate of clustering is sensitive to fluctuations beyond the survey boundaries itself, then it is likely that the measurement we propose would correspond to the largest scale estimate of three dimensional matter clustering yet made. This would in itself be an exciting test of theories, for example probing the power spectrum beyond the matter-radiation equality turnover, and allowing access to the Harrison-Zeldovich portion. There has been much work analyzing large scale anomalies in the clustering measured from the CMB [41][42][43], and it would be extremely useful to see if anything comparable is seen from galaxy large scale structure data. On smaller scales, one could use the matter-radiation equality turnover as a cosmic ruler [44], and this would allow comparison to measurements based on BAO [45].

Second, there will be much information in the reconstructed maps of the large scale densities (such as Fig. 2). One could look at statistics beyond the power spectrum, such as counts-in-cells [46], or the bispectrum, and see how consistent they are with model expectations. One can also compare to the directly measured density field and obtain information on the large scale systematic effects which are modulating the latter. Cross-correlation of the maps with those of different tracers can also be carried out. For example the large scale potential field inferred can be used in conjunction with CMB observations to constrain the Integrated Sachs Wolfe effect[47].

In general, as we will be looking at large scale fluctuations beyond current limits by perhaps an order of magnitude in scale or more, one may expect to find interesting constraints on new physics. For example evidence for the ΛCDM model was seen in the first reliable measurements of large scale galaxy clustering on scales greater than $10\,h^{-1}\mathrm{Mpc}$ (e.g., [48]). Moving to wavelengths beyond $2\pi/(k = 0.02) \sim 300\,\mathrm{Mpc}$ may yet lead to more surprises.

ACKNOWLEDGMENTS

We thank Duncan Campbell and Rachel Mandelbaum for resourceful discussions. We also thank Enrique Gaztanaga for providing us with the 1 in 700 matter particles' positions of Mice-GC simulation for a test. This work is supported by U.S. Dept. of Energy contract DE-

SC0019248 and NSF AST-1909193. This work has made use of CosmoHub. CosmoHub has been developed by the Port d'Informació Científica (PIC), maintained through a collaboration of the Institut de Física d'Altes Energies (IFAE) and the Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), and was partially funded by the "Plan Estatal de Investigación Científica y Técnica y de Innovación" program of the Spanish government.

[1] C. Modi, M. White, A. Slosar, and E. Castorina, J. Cosmol. Astropart. P. **11**, 023 (2019), arXiv:1907.02330 [astro-ph.CO].
[2] T. Baldauf, U. Seljak, L. Senatore, and M. Zaldarriaga, J. Cosmol. Astropart. P. **10**, 031 (2011), arXiv:1106.5507 [astro-ph.CO].
[3] D. Jeong and M. Kamionkowski, Phys. Rev. Lett. **108**, 251301 (2012), arXiv:1203.0302 [astro-ph.CO].
[4] Y. Li, W. Hu, and M. Takada, Phys. Rev. D **90**, 103530 (2014), arXiv:1408.1081 [astro-ph.CO].
[5] H.-M. Zhu, U.-L. Pen, Y. Yu, X. Er, and X. Chen, Phys. Rev. D **93**, 103504 (2016), arXiv:1511.04680 [astro-ph.CO].
[6] A. Barreira and F. Schimidt, J. Cosmol. Astropart. P. **06**, 03 (2017), arXiv:1703.09212 [astro-ph.CO].
[7] P. Li, S. Dodelson, and R. A. C. Croft, Phys. Rev. D **101**, 083510 (2020), arXiv:2001.02780 [astro-ph.CO].
[8] LSST Daek Energy Science Collaboration, (2012), arXiv:1211.0310 [astro-ph.CO].
[9] WFIRST Science Definition Team, (2012), arXiv:1208.4012 [astro-ph.IM].
[10] DESI Collaboration, (2019), arXiv:1907.10688 [astro-ph.IM].
[11] A. V. Kravtsov and A. Klypin, Anatoly, Astrophys. J. **520**, 437 (1999), arXiv:astro-ph/9812311 [astro-ph].
[12] V. Desjacques, D. Jeong, and F. Schmidt, Phys. Rept. **733**, 1 (2018), arXiv:1611.09787 [astro-ph.CO].
[13] H. Gil-Marín, J. Noreña, L. Verde, W. J. Percival, C. Wagner, M. Manera, and D. P. Schneider, Mon. Not. Roy. Astron. Soc. **451**, 539 (2015), arXiv:1407.5668 [astro-ph.CO].
[14] H. Gil-Marín, W. J. Percival, L. Verde, J. R. Brownstein, C.-H. Chuang, F.-S. Kitaura, S. A. Rodríguez-Torres, and M. D. Olmstead, Mon. Not. Roy. Astron. Soc. **465**, 1757 (2017), arXiv:1606.00439 [astro-ph.CO].
[15] S. M. Carroll, (1997), arXiv:gr-qc/9712019 [gr-qc].
[16] C.-H. Chuang *et al.* (BOSS), Mon. Not. Roy. Astron. Soc. **471**, 2370 (2017), arXiv:1607.03151 [astro-ph.CO].
[17] H. A. Feldman, N. Kaiser, and J. A. Peacock, Astrophys. J. **426**, 23 (1994), arXiv:astro-ph/9304022.
[18] K. Yamamoto, M. Nakamichi, A. Kamino, B. A. Bassett, and H. Nishioka, Publ. Astron. Soc. Jap. **58**, 93 (2006), arXiv:astro-ph/0505115.
[19] D. Bianchi, H. Gil-Marín, R. Ruggeri, and W. J. Percival, Mon. Not. Roy. Astron. Soc. **453**, L11 (2015), arXiv:1505.05341 [astro-ph.CO].
[20] B. Jain and E. Bertschinger, Astrophys. J. **431**, 495 (1994), arXiv:astro-ph/9311070 [astro-ph].
[21] F. Bernardeau, S. Colombi, E. Gaztañaga, and Scoccimarro, Phys. Rept. **367**, 1 (2012), arXiv:astro-ph/0112551 [astro-ph].
[22] R. Takahashi, Prog. Theor. Phys. **120**, 549–559 (2008), arXiv:0806.1437 [astro-ph.CO].
[23] P. McDonald and A. Roy, JCAP **08**, 020 (2009), arXiv:0902.0991 [astro-ph.CO].
[24] H. Gil-Marín *et al.*, Mon. Not. Roy. Astron. Soc. **477**, 1604 (2018), arXiv:1801.02689 [astro-ph.CO].
[25] P. Fosalba, M. Crocce, E. Gaztañaga, and F. J. Castander, Mon. Not. Roy. Astron. Soc. **448**, 2987 (2015), arXiv:1312.1707 [astro-ph.CO].
[26] M. Crocce, F. J. Castander, E. Gaztañaga, P. Fosalba, and J. Carretero, Mon. Not. Roy. Astron. Soc. **453**, 1513 (2015), arXiv:1312.2013 [astro-ph.CO].
[27] P. Fosalba, E. Gaztañaga, F. J. Castander, and M. Crocce, Mon. Not. Roy. Astron. Soc. **447**, 1319 (2015), arXiv:1312.2947 [astro-ph.CO].
[28] J. E. McEwen, X. Fang, C. M. Hirata, and J. A. Blazek, JCAP **09**, 015 (2016), arXiv:1603.04826 [astro-ph.CO].
[29] T. Baldauf, U. Seljak, V. Desjacques, and P. McDonald, Phys. Rev. D **86**, 083540 (2012), arXiv:1201.4827 [astro-ph.CO].
[30] N. Kaiser, Mon. Not. Roy. Astron. Soc. **227**, 1 (1987).
[31] S. Ho *et al.*, Astrophys. J. **761**, 24 (2012), arXiv:1201.2137 [astro-ph.CO].
[32] C. Hahn, R. Scoccimarro, M. R. Blanton, J. L. Tinker, and S. A. Rodríguez-Torres, Mon. Not. Roy. Astron. Soc. **467**, 1940 (2017), arXiv:1609.01714 [astro-ph.CO].
[33] K. S. Dawson *et al.*, Astron. J. **145**, 10 (2013), arXiv:1208.0022 [astro-ph.CO].
[34] K. S. Dawson *et al.*, Astron. J. **151**, 44 (2016), arXiv:1508.04473 [astro-ph.CO].
[35] O. Le Fèvre *et al.*, Astron. Astrophys. **576**, A79 (2015), arXiv:1403.3938 [astro-ph.CO].
[36] A. L. Coil, J. A. Newman, M. C. Cooper, M. Davis, S. Faber, D. C. Koo, and C. N. Willmer, Astrophys. J. **644**, 671 (2006), arXiv:astro-ph/0512233.
[37] G. Dalton *et al.*, Proceedings of the SPIE **9147** (2014), 10.1117/12.2055132, arXiv:1412.0843 [astro-ph.CO].
[38] J. Amiaux *et al.*, Proceedings of the SPIE **8842** (2021), 10.1117/12.926513, arXiv:1209.2228 [astro-ph.CO].
[39] O. Doré *et al.*, (2014), arXiv:1412.4872 [astro-ph.CO].
[40] M. Ata *et al.*, Mon. Not. Roy. Astron. Soc. **473**, 4773 (2018), arXiv:1705.06373 [astro-ph.CO].
[41] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, Adv. Astron. **2010**, 847541 (2010), arXiv:1004.5602 [astro-ph.CO].
[42] A. Rassat, J.-L. Starck, P. Paykari, F. Sureau, and J. Bobin, JCAP **08**, 006 (2014), arXiv:1405.1844 [astro-ph.CO].
[43] D. J. Schwarz, C. J. Copi, D. Huterer, and G. D. Starkman, Class. Quant. Grav. **33**, 184001 (2016), arXiv:1510.07929 [astro-ph.CO].
[44] J. Hasenkamp and J. Kersten, JCAP **08**, 024 (2013), arXiv:1212.4160 [hep-ph].
[45] R. Lazkoz, S. Nesseris, and L. Perivolaropoulos, JCAP **07**, 012 (2008), arXiv:0712.1232 [astro-ph].
[46] A. Yang and W. C. Saslaw, Astrophys. J. **729**, 123 (2011), arXiv:1009.0013 [astro-ph.CO].
[47] A. J. Nishizawa, PTEP **2014**, 06B110 (2014), arXiv:1404.5102 [astro-ph.CO].
[48] G. Efstathiou, W. J. Sutherland, and S. J. Maddox, Nature **348**, 705 (1990).