Displacement field analysis via optimal transport: Multitracer approach to cosmological reconstruction

Farnik Nikakhtar[1,*] Ravi K. Sheth[2,3] Nikhil Padmanabhan[1,4] Bruno Lévy[5] and Roya Mohayaee[6,7]

[1]*Department of Physics, Yale University, New Haven, Connecticut 06511, USA*
[2]*Center for Particle Cosmology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*
[3]*The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, Trieste 34151, Italy*
[4]*Department of Astronomy, Yale University, New Haven, Connecticut 06511, USA*
[5]*Université Paris Saclay, Laboratoire de Mathématiques d'Orsay, Centre Inria de Saclay,*
1 Rue Honoré d'Estienne d'Orves, 91120, Palaiseau, France
[6]*Sorbonne Université, CNRS, Institut d'Astrophysique de Paris, 98bis Bld Arago, 75014 Paris, France*
[7]*Rudolf Peierls Centre for Theoretical Physics, University of Oxford,*
Parks Road, Oxford OX1 3PU, United Kingdom

We demonstrate the effectiveness of one of the many multitracer analyses enabled by optimal transport (OT) reconstruction. Leveraging a semidiscrete OT algorithm, we determine the displacements between initial and observed positions of biased tracers and the remaining matter field. With only redshift-space distorted final positions of biased tracers and a simple premise for the remaining mass distribution as input, OT solves the displacement field. This extracted field, assuming asymptotically uniform density and a gradient flow displacement, enables reconstruction of the initial overdensity fluctuation field. We show that the divergence of the OT displacement field is a good proxy of the linear density field, even though the method never assumes the linear theory growth. Additionally, this divergence field can be combined with the reconstructed protohalos to provide a higher signal-to-noise measurement of the baryon acoustic oscillations standard ruler than was possible with either measurement individually.

I. INTRODUCTION

Essentially all currently viable models for cosmological structure formation suggest that the initial conditions were almost exquisitely uniform. In fact, in simulations of cosmological structure formation, the initial density field is homogeneous (particles start from a grid), and subsequent evolution is driven by inhomogeneous perturbations in the initial velocity field. In this case, the initial velocity field is the gradient of a potential field, and the overdensity field is explicitly related to the divergence of the initial velocity field:

$$\delta_{\text{init}} \propto \nabla \cdot \boldsymbol{v}_{\text{init}} \propto \nabla^2 \Phi_{\text{init}}, \qquad (1)$$

where Φ_{init} is the initial gravitational potential [1]. Although the constant of proportionality depends on the cosmological model, the relation is otherwise general.

It is common to use $\boldsymbol{q}$ to denote the initial position of a particle, and $\boldsymbol{x}$ its position at some later time t. Then the *displacement* $\boldsymbol{S}$ is defined by

$$\boldsymbol{x}(t) = \boldsymbol{q} + \boldsymbol{S}(\boldsymbol{q}, t). \qquad (2)$$

Optimal transport (hereafter OT) theory is a fast-growing gem in mathematics that, broadly, computes distances and relations between different probability distributions [2]. In cosmology OT is a natural framework to use since it can *exactly* reconstruct the displacements $\boldsymbol{S}(\boldsymbol{q}, t)$ in (2) if they are gradients of a potential (see Refs. [3–5] and references therein). In cosmological structure formation, this is explicitly true in Lagrangian perturbation theory up to and including the second order (see e.g. [6]). This, and the fact that the initial conditions were "uniform" has motivated the use of OT methods for reconstructing the displacements, from knowledge of only the particle positions at some late time [3,7]. i.e. if n labels a particle, then OT determines estimates $\boldsymbol{S}_{\text{OT}n}$ of each of the displacements $\boldsymbol{S}_n$, knowing only the set of final positions $\boldsymbol{x}_n$. From this, $\boldsymbol{q}_{\text{OT}n} \equiv \boldsymbol{x}_n - \boldsymbol{S}_{\text{OT}n}$. Recent development of semidiscrete optimal transport methods [8,9] has transformed the field by vastly increasing the speed with which the displacements, and hence the initial conditions, can be reconstructed [10–12]. Note, the displacement $\boldsymbol{S}$ can be equivalently thought of as a function of $\boldsymbol{q}$ or $\boldsymbol{x}$. In computing derivatives,

---

*farnik.nikakhtar@yale.edu

123512-1

spatial derivatives with respect to q will be denoted as ∇_q, while those with respect to x will be denoted as ∇_x.

Such reconstructions are particularly attractive for studies which seek to constrain cosmological parameters from the statistical properties of the initial field [13]. The most used summary statistic of the field is the power spectrum $P(k)$, or its Fourier transform, the two-point correlation function $\xi(r)$. Typically, these are estimated directly from the particle distribution. However, the distribution of q_{OTn} is uniform (by design!), so $\xi_{OT}(r) = 0$ for all r. For this reason, Ref. [10] estimated two-point statistics of the slightly evolved positions $x_n(\alpha) = q_{OTn} + \alpha S_{OTn}$, for $\alpha \sim 0.1$. To compare this with the "initial" conditions in the simulations, they used the evolved positions $q_n + v_{\text{init}-n}/(afH)$ (the displacement is just initial speed times a suitably chosen time), but because $afH = a(d \ln D/d \ln a)(d \ln a/dt)$ depends on a cosmological model [$D(a)$ is the linear theory growth factor], matching α to $(afH)^{-1}$ gave the impression that the comparison was model dependent.

One of the goals of this work is to show that Eq. (1) offers a more direct, cosmology-independent estimate of the statistics of the initial fluctuation field. (Strictly speaking, to a good approximation, it is the initial fluctuation field evolved using linear theory to the time of observation.) We do this in two steps. In Sec. II, we consider the ideal case in which one starts from perfect information about the evolved dark matter field (all the x_n are known), as in Ref. [10]. In Sec. III, we show how to treat the case in which evolved positions of only a biased subset of the full field are known, as in Ref. [12]. This second step requires a guess about the distribution of the mass that is not directly associated with the biased tracers, which we refer to as the "dust." While previous work in the OT context primarily focused on reconstructing the biased field [14], here we show the dual reconstruction capability inherent in OT—namely, the reconstruction of both the protohalo field and the linear dark matter field. This leads to a discussion on the potential synergy between these reconstructions, providing a good context for contrasting our reconstruction scheme with other approaches. In a broad sense, the OT displacement field solves a nonlinear continuity equation, in contrast to standard reconstruction algorithms e.g. [13,15], which solve the linearized continuity equation. Compared to action minimization-based algorithms e.g. [16,17], which attempt to trace objects backwards along estimated trajectories, OT searches for an effective potential, the gradient of which maps each object from its observed position to its initial Lagrangian position—its solution is unique and the intermediate positions do not enter the analysis explicitly. Section IV discusses our results in the context of multitracer studies of large-scale structure, and highlights the fact that OT-like methods allow displacements and quantities built from the displacement field to be used as tracers of the cosmological fluctuation field. We summarize our results in Sec. V.

We demonstrate our methods using the dark matter and halo distributions in the same HADES cosmological simulations [18] that were used in Ref. [14]. The background cosmological model is flat Λ cold dark matter (ΛCDM) with $(\Omega_m, \Omega_b, h) = 0.3175, 0.049, 0.6711)$. The initial fluctuations were Gaussian with $P_{\text{Lin}}(k)$ having shape and amplitude parameters $(n_s, \sigma_8) = 0.9624, 0.833)$, and each simulation evolves 512^3 particles in a periodic box that is $L = 1h^{-1}$ Gpc on a side, so the particle mass is $m_p = 6.5 \times 10^{11} h^{-1} M_\odot$. We focus on the $z = 0$ outputs, where we consider the dark matter field, as well as halos more massive than $20m_p$; see Ref. [14] for further details about how the halos were identified. All our results are averaged over 20 simulation boxes.

II. PERFECT KNOWLEDGE OF THE FULL NONLINEAR FIELD

A. Divergence of displacement

Before we begin, it is important to distinguish between a number of measures of the initial power spectrum (PS). First is the theoretical PS which is used to generate the Fourier modes of the initial fluctuation field δ_{init}. We refer to the related power spectrum, scaled to $z = 0$ using linear theory, as $P_{\text{Lin}}(k)$. We use $P_{\text{IC}}(k)$ to denote the power spectrum of the scalar δ_{IC}, which is the initial fluctuation field δ_{init}, scaled to $z = 0$ using linear theory. This power spectrum should differ from $P_{\text{Lin}}(k)$ only because it is a single stochastic realization (of which P_{Lin} is the mean).

The Fourier modes of the initial velocities and densities are related: $v_{\text{init}} = ik\delta_{\text{init}}/k^2$, so that $\nabla \cdot v_{\text{init}} = -(afH)_{\text{init}}\delta_{\text{init}}$. The vector $v_{\text{init}}/(afH)_{\text{init}}$ has units of distance, and $(D_0/D_{\text{init}})v_{\text{init}}/(afH)_{\text{init}} \equiv S_{\text{Zel}}$ corresponds to the displacement vector, from the initial to the present time, in the Zeldovich approximation. We will define the scalar $\Theta_{\text{Zel}} \equiv \nabla_q \cdot S_{\text{Zel}}$, and use $P_{\Theta_Z\Theta_Z}(k)$ to denote the power spectrum of Θ_{Zel}. Since $\Theta_{\text{Zel}} = -(D_0/D_{\text{init}})\delta_{\text{init}} = -\delta_{\text{IC}}$, $P_{\Theta_Z\Theta_Z}(k)$ should be the same as $P_{\text{IC}}(k)$, so a comparison of the two measures the accuracy of our scheme for estimating velocity divergences in the simulations.

Note that $P_{\Theta_Z\Theta_Z}(k)$ will differ from the power spectrum of the Zeldovich-evolved density field, where particles move along straight lines following the gradient flow until $z = 0$, as detailed in [19]. In addition, since the true displacement of a particle, S_{true}, is different from the approximate S_{Zel} (the Zeldovich displacement is just the first-order term in Lagrangian perturbation theory), $P_{\Theta_Z\Theta_Z}$ differs from the power spectrum of the divergence of the actual displacements, $P_{\Theta\Theta}(k)$, where

$$\Theta \equiv \nabla_q \cdot S_{\text{true}}, \tag{3}$$

and we shall discuss $\nabla_x \cdot S$, the divergence at x rather than q, in Appendix C.

Finally, we define Θ_{OT} as the divergence of the displacements returned by our OT reconstructions. Recall that whereas the set of $S_{\text{true}} = x - q$ displacements uses a full cosmological simulation to obtain the evolved positions x from an initial grid of q positions, OT starts from this list of x positions, and returns a list of corresponding qs (Laguerre cell barycenters that are not necessarily on a grid).

B. Cross correlations

Before we look at the power spectra themselves, it is helpful to look at how closely these different quantities correlate with one another. To quantify the degree of correlation, we use the normalized coefficient

$$r_{AB}(k) \equiv \frac{\langle P_{AB}(k) \rangle}{\sqrt{P_{AA}(k) P_{BB}(k)}}. \tag{4}$$

(Note that we do not subtract shot noise from any measurements.) To help calibrate the figures throughout the paper, the thin gray curve in Fig. 1 shows that the nonlinear field is well correlated with the initial one only on very large scales ($r \geq 0.9$ only if $k \leq 0.1h$ Mpc^{-1}). We see that δ_{init} and Θ_Z are extremely tightly correlated (pink curve shows $r \sim 1$ for essentially all k), as expected. More interesting is that Θ and Θ_Z, and Θ and δ_{init}, are also quite well correlated on baryon acoustic oscillations (BAO) scales (red and green curves, which overlap almost exactly, have $r \geq 0.9$ for $k \leq 0.45h$ Mpc^{-1}). This matters, because the reconstructed Θ_{OT} is extremely well correlated with the true Θ (blue curve)—this, after all, is why OT reconstruction is interesting in the first place—so Θ_{OT} is well correlated with both Θ_Z and δ_{init} (yellow and purple curves, which overlap almost exactly, show $r \geq 0.9$ for $k \leq 0.35h$ Mpc^{-1}). Hence, compared to the nonlinear density field, the reconstructed Θ_{OT} field provides about $(0.35/0.1)^3 \sim 40\times$ more k modes with $r \sim 1$ to the initial fluctuation field.

This is nontrivial, because the explicit goal of OT is to return $S_{OT} = S_{\text{true}}$, and the blue curve in Fig. 1 suggests that it does this rather well. However, other than doing so while also returning a uniform density field, it does not explicitly try to reconstruct δ_{init} or Θ_Z. Since $S_{\text{true}} \neq S_{\text{Zel}}$, even if OT returned S_{true} perfectly, it would not perform better–in the sense of correlating with δ_{init}—than the green curve shown in Fig. 1.

Stated differently: if we use $\Phi_{OT}(q)$ to denote the potential which determines the OT-reconstructed displacements $\nabla\Phi_{OT}$, then, because $S_{OT} \approx S_{\text{true}} \neq S_{\text{Zel}}$, we know that $\Phi_{OT} \neq \Phi_{\text{init}}$. This is not surprising, since by definition, the gradient of the OT potential moves particles to their exact nonlinear positions, in contrast to $\nabla\Phi_{\text{init}}$, which would result in the Zeldovich-evolved field. Nevertheless, the correlation between $\Theta_{OT} \equiv k^2\Phi_{OT}$ and $\delta_{\text{init}} \equiv k^2\Phi_{\text{init}}$ (purple curve) is sufficiently good on BAO scales that we should be able to use Θ_{OT} as a proxy for δ_{init}. We test this explicitly in the next sections.
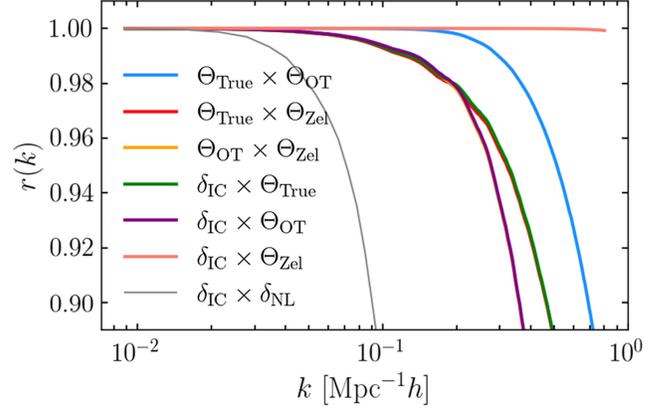
FIG. 1. Normalized correlation coefficients between the various divergences in real space, as defined in the text. On the $k \leq 0.2h$/Mpc scales that are most relevant to BAO analyses, the OT displacement divergences are remarkably well correlated with the true ones (blue), even though the OT algorithm includes no cosmological information. They are less well correlated with the initial density field (purple that overlaps with yellow), but are still much better than the nonlinear density (gray). The green curve overlaps with the red curve as well.

C. Power spectra and correlation functions

Figure 2 shows the power spectra associated with these different divergences. To guide the eye, the dashed curve in the top panel shows $P_{\text{Lin}}(k)$; it is very similar to P_{IC} and to $P_{\Theta_Z\Theta_Z}(k)$. The power spectrum of true divergences is also remarkably similar to $P_{\text{Lin}}(k)$, despite the fact that Θ is not as well correlated with δ_{init}.

We are most interested in Θ_{OT}. Although the curve for Θ_{OT} has noticeably less power than the others at large k, it is in excellent agreement on BAO scales. Note in particular that, whereas amplitudes are scaled out in $r(k)$, they matter here. Nevertheless, on BAO scales, the power spectrum of Θ_{OT} is an excellent tracer of $P_{\text{Lin}}(k)$. To emphasize the agreement, the middle panel shows the (square root of the) ratio of these $P(k)$ to $P_{\text{Lin}}(k)$. This shows that Θ_{OT} is within a few percent of linear theory on BAO scales; in fact, it traces the jaggedness at small k that is due to cosmic variance. The bottom panel shows the result of taking the ratio with respect to $P_{\Theta_Z\Theta_Z}(k)$ instead (we take the average of the ratio, rather than the ratio of the averages). This removes the cosmic variance, and shows that $\Theta_{OT}(k)$ has the linear theory amplitude to within a few percent for $k \leq 0.1h$ Mpc^{-1}. (While the nonlinear field also has the right amplitude at low k, the previous figure shows that it has $r(k) \sim 1$ over a much smaller range of k.)

Figure 3 shows the result of Fourier transforming the power spectra shown in the previous figure. In practice, this transformation is performed using a fast Fourier transform (FFT) on the 3D power spectrum calculated in the box. On BAO scales, all the curves are in excellent agreement with linear theory (shaded region shows the rms scatter between
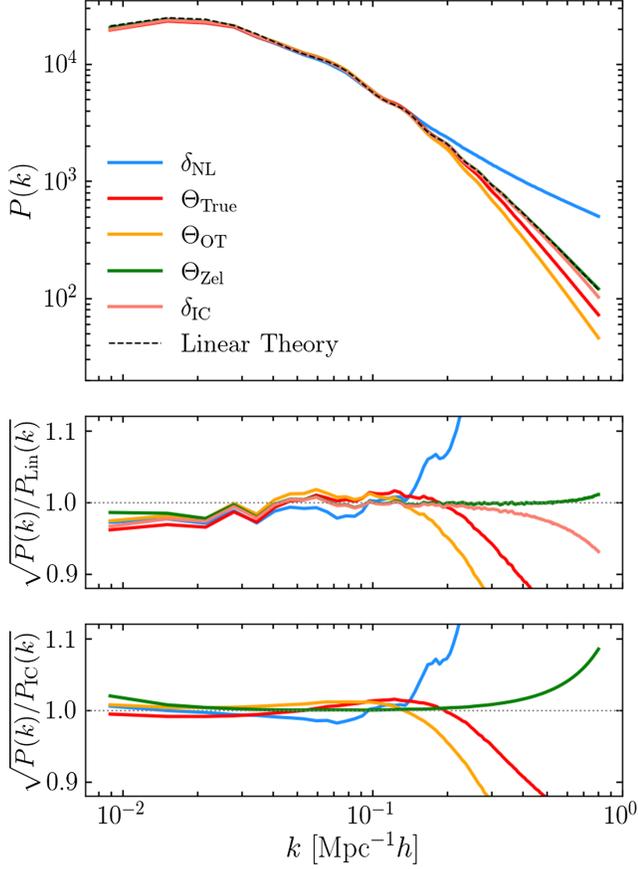
FIG. 2. Comparison of various power spectra in real space (as labeled). Dashed line shows the expected $z = 0$ linear theory curve, $P_{\mathrm{Lin}}(k)$, and uppermost blue line shows the actual, evolved, nonlinear curve $P_{\mathrm{NL}}(k)$. Yellow curve, for Θ_{OT}, represents our reconstruction of the shape and amplitude of $P_{\mathrm{Lin}}(k)$ from $P_{\mathrm{NL}}(k)$, despite having no cosmological information (e.g. linear theory growth factor). Middle: The result of dividing each of these curves by $P_{\mathrm{Lin}}(k)$ and taking the square root. Bottom: They were each divided by $P_{\mathrm{IC}}(k)$; this removes cosmic variance.

20 simulations), except for the one labeled nonlinear, for which the BAO feature is obviously smeared out. In particular, ξ of Θ_{OT} reproduces the shape and amplitude of ξ_{Lin} (on BAO scales) rather well. This demonstrates how one can use the OT displacements to reconstruct the shape and amplitude of ξ_{Lin} (on BAO scales) without using any cosmological information. In particular, it was not necessary to specify a growth factor when defining the reconstructed displacements, and the extra scaling used by Ref. [10] is unnecessary.

Before we end this section, we recall from the Introduction that one can think of OT as *approximately* returning the initial fluctuation field evolved using linear theory to the time of observation. This is because, although our results show that this approximation is very good, the OT reconstruction does differ slightly from ξ_{Lin}. We know that Φ_{OT} is not the linear Φ, but has higher-order corrections to it. It may be possible to construct a better estimate
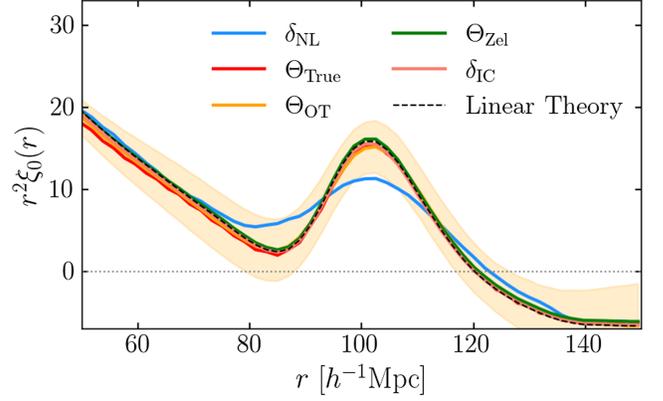
FIG. 3. Correlation functions corresponding to the power spectra shown in the previous figure. Note in particular that it was not necessary to specify a growth factor for the Θ_{OT} curve, but it is, nevertheless, indistinguishable from the other linear field curves.

of the linear Φ from the Φ_{OT}. Exploring this is currently work in progress.

III. BIASED, REDSHIFT-SPACE DISTORTED TRACERS

In the previous section we have outlined what could be achieved if one had a perfect knowledge of positions (but not velocities) in the nonlinear field. In practice, one usually only has information about a biased subset of the nonlinear field whose positions have been distorted by peculiar velocities, so we now discuss how to extend our method to this more realistic case.

A. Reconstructing both biased tracers and the total field

The OT approach assumes that a reasonably accurate mass estimate is available for each tracer, and all statistics are "mass weighted." However, it also requires knowledge of the amount and spatial distribution of the mass that is not observed, what we call the dust. Reference [14] shows that in order to model the missing dust one must assume values for two cosmological parameters: Ω_m and σ_8.

The curve that is second from top in Fig. 4 shows the cross correlation between the full initial and reconstructed fields when the OT algorithm is started with mass-weighted halos more massive than $10^{13} h^{-1} M_\odot$ plus the true dust (i.e. particles that were not in these halos) in real space (dashed). For comparison, the top (solid) curve shows the result of starting from the full nonlinear field (i.e. same as the yellow or purple curves in Fig. 1). Some of the degradation in $r(k)$ is because, in our reconstruction, we treat all the mass of a halo as being concentrated at its center, and the mass of the reconstructed protohalo as being located at the barycenter of its Laguerre cell, so that all of a halo's mass has the same displacement vector: $\boldsymbol{x}_{\mathrm{com}} - \boldsymbol{q}_{\mathrm{com}}$. A better model would
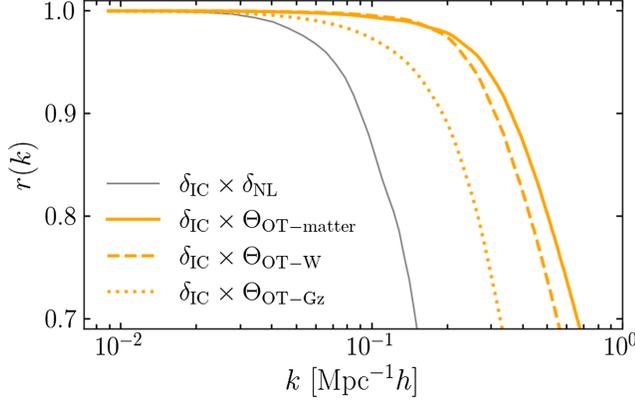
FIG. 4. Dependence of normalized cross correlation between the full initial and reconstructed fields, when the OT algorithm is started from the full nonlinear field (solid; OT-matter), from mass-weighted halos plus the true dust in real space (dashed; OT-W), and from mass-weighted halos plus the Wiener filter model for the dust, in redshift space (dotted; OT-Gz). For comparison, the gray curve (same as in Fig. 1) shows the $\delta_{IC} - \delta_{NL}$ correlation.

FIG. 5. Initial, evolved and reconstructed power spectra. Top blue curve shows observed monopole of a biased, redshift-space distorted set of tracers (mass-weighted halos more massive than $10^{13}h^{-1}M_\odot$); green and peach curves show the power spectra of the corresponding protohalos and of the initial dark matter field ($P_{IC} \approx P_{Lin}$). Yellow and purple curves show the power spectra of the OT reconstructed halos and Θ_{OT} of the full reconstructed field, and red curve shows the power spectrum of the field which is their sum (OT reconstructed halo fluctuations plus Θ_{OT}). Dashed black line shows the shot noise for mass-weighted halos.

distribute the mass uniformly within the Laguerre cell, with displacement given by $(x_{com} - q_{com}) - (q - q_{com})$.

Of course, in practice, we do not know the true spatial distribution of the dust. If our guess is not accurate, then this will further degrade $r(k)$. Reference [14] discusses a number of approximations which one can use to model the dust. A simple Wiener-filter model, that uses the observed tracers to guide the placement of the dust particles, was shown to work quite well and shall hence be used hereafter. For this model, if δ_b is the fluctuation in the biased tracer field, then

$$\delta_d = \frac{\langle \delta_b \delta_d \rangle}{\langle \delta_b \delta_b \rangle} \delta_b \approx (b_d/b)\delta_b, \qquad (5)$$

where the final expression assumes that $\langle \delta_b \delta_d \rangle \approx bb_d P(k)$ and $\langle \delta_b^2 \rangle \approx b^2 P(k) + \text{noise} \approx b^2 P(k)$ on BAO scales.

In addition, Ref. [14] also studied the impact of redshift-space distortions—the anisotropy induced by the fact that estimated distances along the line of sight are distorted by peculiar velocities—finding that ignoring them altogether was a rather good approximation, at least for restoring the monopole of the two-point correlation to its initial shape.

The curve that is third from top in Fig. 4 (dotted) shows the result of starting from redshift-space distorted positions of biased tracers, using the Wiener-filter model to generate the dust, performing the semidiscrete OT reconstruction, and then measuring $r(k)$ between Θ_{OT} of the full field (i.e. reconstructed halos plus dust) and δ_{init}. Evidently, inaccuracies in the Wiener-filter model for the dust and redshift-space distortions do degrade $r(k)$, but it is still significantly better than δ_{NL} (bottom gray curve, same as the gray curve in Fig. 1).

Figures 5 and 6 show the corresponding $P(k)$ and $\xi(r)$ of the reconstructed protohalos and of Θ_{OT}. We do not subtract shot noise from the measured $P(k)$; this is negligible for the full field, but significant for the halos and protohalos. [We mass weight the halos, so the shot noise is not $\bar{n}_h^{-1}$, but $(\sum_i m_i^2/V)/(\sum_i m_i/V)^2 = \bar{n}_h^{-1}\langle m^2 \rangle/\langle m \rangle^2$.]

The first thing to notice is that $P(k)$ and $\xi(r)$ of the OT reconstructed halos (yellow) are extremely similar to the protohalos (green). This is the agreement previously reported by Ref. [14]. Although Ref. [14] focused on the reconstructed position of biased tracers, the same OT algorithm also outputs the displacements of the dust particles as well. We can use the full set of displacements (mass-weighted halos plus dust) to estimate Θ_{OT} as described in the previous section. This gives the purple curves, which are very close to P_{IC} or ξ_{IC} (pink) on BAO scales. Evidently, neither inaccuracies in the Wiener-filter model for the dust nor redshift-space distortions degrade the agreement substantially compared to the ideal case shown in the previous section. We conclude that our OT methodology successfully reconstructs both the positions of the biased tracers *and* the fluctuations in the full field.

B. Combining the two reconstructions

Figures 5 and 6 show that the amplitude of the clustering signal of the reconstructed biased tracers is smaller than that in the evolved field. The physics which drives this difference—essentially the continuity equation—dictates that if the biased subset initially satisfies $\delta_b = b_L \delta_L$ then the evolved subset satisfies $(b_L + 1)\delta_{NL}$ (e.g.
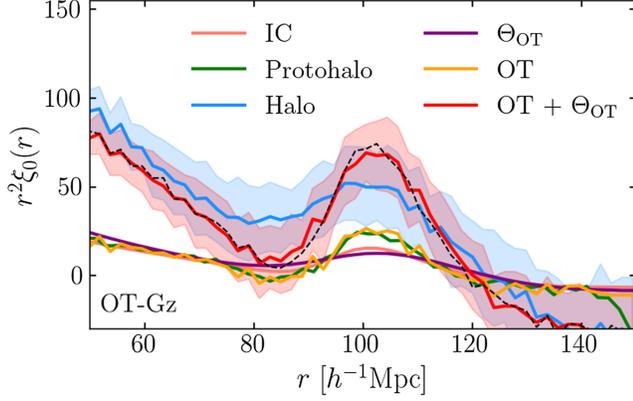
FIG. 6. Correlation functions corresponding to the power spectra shown in Fig. 5. The dashed black curve shows the correlation function of the sum of the actual protohalo and δ_{init} fluctuation fields.

Refs. [12,20]). However, because the amplitude of the reconstructed signal is smaller (yellow curve lies far below blue), the signal to noise of the BAO feature in the reconstructed field is diminished. We can use our knowledge of the physics to correct for this as follows.

On BAO scales, $\delta_{\text{NL}} \approx \delta_{\text{L}} \approx \Theta_{\text{Zel}}$, so the evolved subset $(b_{\text{L}} + 1)\delta_{\text{NL}}$ is approximately $b_{\text{L}}\delta_{\text{L}} + \Theta_{\text{Zel}}$. This suggests that if we add the reconstructed halo and Θ_{OT} fluctuation fields, then the result should have a power spectrum which is given by

$$\langle(b_{\text{L}}\delta_{\text{L}} + \Theta_{\text{Zel}})^2\rangle \approx b_{\text{L}}^2 P_{\text{Lin}} + 2b_{\text{L}}P_{\text{Lin}} + P_{\text{Lin}}$$
$$\approx (b+1)^2 P_{\text{Lin}}. \qquad (6)$$

That is, it should have approximately the same amplitude as the original signal. The red curve shows this result: The overall amplitude of the signal is nearly restored. On scales slightly below the BAO feature, the red curve lies slightly below the blue one. This slight mismatch is reassuring, because the blue curve is the redshift-space distorted monopole, so it is slightly enhanced compared to the real-space correlation [21], which is the quantity that OT aims to reconstruct.

In addition to this slight difference in amplitude, the OT reconstructed signal also has a sharper BAO feature compared to the linear matter field. The reason for this is also understood [12]: most protohalos are associated with peaks in the initial field, and this is due to the scale-dependent nature of the "bias" factor b_{L}, which includes a k^2 term that sharpens the peak amplitude [22,23]. There is further corroborating evidence of the peak-protohalo correspondence: the initial speeds of peaks are predicted to be smaller than the rms speeds of randomly chosen positions, and protohalo speeds are indeed smaller, and well reproduced by peak theory [24]. So, it is reassuring that OT

returns smaller rms displacements for the halo field ($\sim 9.1h^{-1}$ Mpc) than for the total ($9.8h^{-1}$ Mpc).

In the evolved field, the scale-independent piece increases, but the scale-dependent piece decreases (both by unity) [25]. Adding the linear (rather than the evolved) fluctuation field to that of the protohalos increases the scale-independent bias while leaving the scale-dependent part unchanged. This results in an enhanced BAO signal. To check that this is indeed what is going on, the dashed black curve shows the correlation function of the field that is the sum of the actual protohalo and δ_{init} fluctuation fields. Our (red) OT-reconstructed curve is in great agreement, indicating that the reason for the enhanced signal is understood.

Of course, the two reconstructed fields are not independent, so when constraining parameters from the shape/scale of the BAO signal, the covariance between the reconstructed halo and matter (i.e. Θ_{OT}) fields must be taken into account. This raises the question: What combination (linear or nonlinear) of the two fields has the most constraining power ? In this context, the straight linear sum of the two fields is not obviously optimal, but it does illustrate the potential gains from combining the two. We return to this in the next section.

C. Relation to previous work

The straight linear sum has an interesting connection to what is known as "standard: reconstruction [26] (see Ref. [27] for a perturbative analysis and [28] for a summary of related methods). In this approach, the observed nonlinear, biased field $\delta_b(\boldsymbol{k})$ is first smoothed with a filter $G(|\boldsymbol{k}|)$, and then used to define a displacement field $\boldsymbol{S}_G(\boldsymbol{k}) \equiv i(\boldsymbol{k}/k^2) [\delta_b(\boldsymbol{k})/b]G(k)$. The observed particles are shifted by $-\boldsymbol{S}_G(\boldsymbol{k})$, and the overdensity field which results is called $\delta_{\text{displaced}}$. In addition, a uniform distribution of particles covering the same geometry is also shifted by $-\boldsymbol{S}_G(\boldsymbol{k})$, and the resulting field is called δ_{shifted}. The reconstructed field is $\delta_{\text{rec}} \equiv \delta_{\text{displaced}} - \delta_{\text{shifted}}$. This method requires knowledge of b (hence σ_8), and depends on a smoothing filter $G(k)$ to satisfy the linearized version of the continuity equation. Typically $G(k) \approx \exp(-k^2 R^2/2)$ with $R \approx 15h^{-1}$ Mpc or more, mainly calibrated with simulations. If $\delta_{\text{displaced}} \to 0$, then $\delta_{\text{rec}} \approx -\delta_{\text{shifted}}$. If, in addition, these shifts are gradients of a potential field (more likely if the smoothing scale R is large), then $\delta_{\text{shifted}} \approx \boldsymbol{\nabla} \cdot \boldsymbol{S}_G$, which is the moral equivalent of Θ_{OT} (modulo the smoothing window).

The correspondence is even closer. The standard method assumes that $\delta_m = \delta_b/b$ (on sufficiently large scales) whereas OT models the dust. However, for the simplest Wiener-filter dust model [Eq. (5)], if f_b is the mass fraction in biased tracers, then the total mass fluctuation field is given by

$$\delta_m \equiv f_b\delta_b + (1 - f_b)\delta_d$$
$$\approx [f_b + (1 - f_b)(bb_d/b^2)]\delta_b = \delta_b/b, \qquad (7)$$

where the second line assumes that $\delta_d/b_d = \delta_b/b = \delta_m$ on BAO scales (i.e. shot noise can be ignored), and the final expression follows from the fact that $f_b b + (1 - f_b) b_d \equiv 1$. Therefore, unless more sophisticated models for the dust are used, the starting assumptions for OT and the standard approaches are very similar. Of course, OT is *guaranteed* to have $\delta_{\text{displaced}} = 0$ without having to introduce any smoothing window. And, perhaps more importantly, as Ref. [14] notes, OT is able to quantify the systematic error which comes from inaccuracies in the model for the dust.

The OT approach also has a close connection to the work of Ref. [29]. We discuss this in Appendix C. OT differs from this, more standard, and action-minimization approaches by explicitly tracking the dust component separately. As a result, the displacements it returns for the biased tracers are physically intuitive: e.g. for halos, OT returns both how halos have moved, and how their shapes evolved as they formed. In addition, it returns at least two reconstructed fields: the halos and the dust. In the previous section, we showed that this allowed us to provide a reconstructed BAO signal that had higher signal to noise than if we had worked with either field individually. We exploit another benefit of having more than one field in the next section.

IV. MULTIPLE TRACER ANALYSES

Although we have focused on the virtues of OT reconstruction for BAO-related science, there is considerable interest in using the large-scale clustering signal to constrain the statistical properties of the initial fluctuation field.

For example, suppose we wish to find that linear combination of the reconstructed halo and Θ fields, δ_h and Θ_{OT}, for which the difference from the full (initial) matter field δ_m is as small as possible (in a least-squared sense). Then, we want that w_h and w_Θ which minimize

$$\chi^2(k) \equiv \sum_i \left(\delta_m(\boldsymbol{k}) - w_h \delta_h(\boldsymbol{k}) - w_\Theta \Theta_{\text{OT}}(\boldsymbol{k}) \right)^2, \quad (8)$$

where the sum is over all modes with magnitude $|\boldsymbol{k}| = k$. This yields

$$w_h = \frac{P_{mh}}{P_{hh}} \frac{1 - r_{h\Theta}^2(r_{m\Theta}/r_{mh}r_{h\Theta})}{1 - r_{h\Theta}^2}$$

$$w_\Theta = \frac{P_{m\Theta}}{P_{\Theta\Theta}} \frac{1 - r_{h\Theta}^2(r_{mh}/r_{m\Theta}r_{\Theta h})}{1 - r_{h\Theta}^2} \quad (9)$$

with variance
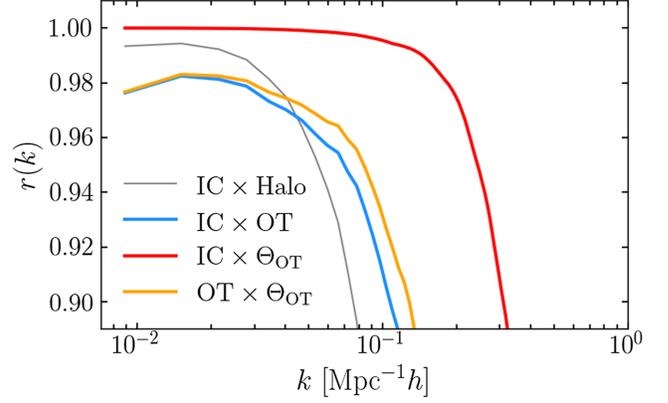
FIG. 7. Correlation coefficients between the initial field and the Θ_{OT} estimate of it, the reconstructed halo field, and the evolved (unreconstructed) halo field, and between the two reconstructed fields. Note that $r(k)$ may not be 1 at large scales due to shot noise.

$$\chi_{\min}^2 = P_{mm} \left(1 - r_{mh}^2 - \frac{(r_{m\Theta} - r_{mh}r_{h\Theta})^2}{1 - r_{h\Theta}^2} \right)$$

$$= P_{mm} \left(1 - r_{m\Theta}^2 - \frac{(r_{mh} - r_{m\Theta}r_{h\Theta})^2}{1 - r_{h\Theta}^2} \right). \quad (10)$$

If we only used one field, δ_h, then the variance would be $P_{mm}(1 - r_{mh}^2)$, so the final term shows how much the variance is reduced by adding the extra field. There is no reduction only if the correlation between δ_m and Θ is entirely due to the correlations of δ_m and δ_h with Θ: i.e. only if $r_{m\Theta} = r_{mh}r_{h\Theta}$. Figure 7 shows the various cross-correlation coefficients: $r_{m\Theta} \neq r_{mh}r_{h\Theta}$ in general.

It is conventional [e.g. Ref. [30]] to use $\sqrt{\chi_{\min}^2/P_{mm}}$ as a measure of the stochasticity between the optimally weighted fields, $w_h \delta_h + w_\Theta \Theta$, with w_h and w_Θ given by Eq. (9), and the one they are trying to represent, δ_m. Figure 8 shows the stochasticity as a function of k. For this (mass-weighted) halo example, the reconstructed halos are closer to the initial conditions (ICs) than the evolved halos—reconstruction was useful—but the reconstructed Θ_{OT} field is much closer, so that adding Θ_{OT} to the halos results in significant gains (i.e. reduction in stochasticity), but not vice versa.

Furthermore, because OT reconstruction does not assume that the initial field was Gaussian, it provides a useful framework for studying departures from Gaussianity. Indeed, Ref. [31] explored the use of the one-point distribution of divergences in the full OT reconstructed field (i.e. without considering the added complication of starting from a biased tracer field) as a tool for quantifying primordial non-Gaussianity. However, other constraints on primordial non-Gaussianity are now sufficiently tight that one-point statistics are not expected to provide competitive constraints; two-point or higher-order statistics are expected to be more informative.

Models where the primordial non-Gaussianity is "local" are a case in point. These have $\phi(\boldsymbol{x}) = \phi_{\text{G}}(\boldsymbol{x}) + f_{\text{NL}}\phi_{\text{G}}^2(\boldsymbol{x})$,
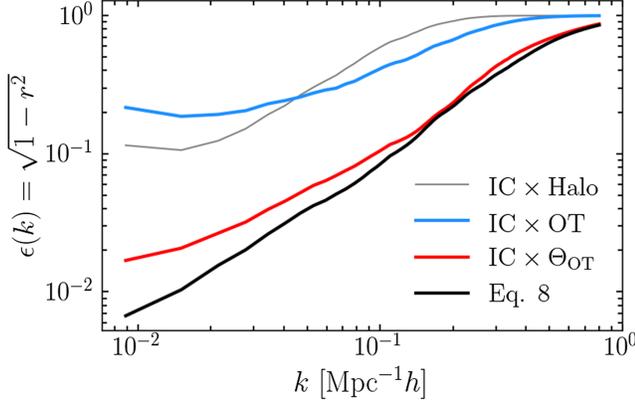
FIG. 8.   Stochasticity between the initial fluctuation field and the evolved halo field, the OT reconstructed halos, the divergence of the OT displacements of the full field, and the optimal linear combination of the two reconstructed fields.

where ϕ_G is an isotropic Gaussian random field. On large scales in these models, biased tracers satisfy $\delta_b(\mathbf{k}) = b_\delta \delta(\mathbf{k}) + f_{NL} b_\phi \phi(\mathbf{k})$. Since $\phi(\mathbf{k}) \propto \delta(\mathbf{k})/k^2$, the clustering of biased tracers has a characteristic scale dependence [32]. However, the scales on which this is dramatic are so large that cosmic variance on the measured power spectrum is a concern. On the other hand, the cosmological information here—in this case the amplitude of f_{NL}—is carried, not by the power spectrum itself, but by the ratio of the biased and unbiased fields, or by the ratio of two differently biased fields.

If both fields suffer the same cosmic variance then, to first approximation, cosmic variance cancels when one studies their ratio [33]. To be useful in practice, the tracer fields must occupy the same survey volume *and* have different bias factors. For two such tracers, A and B, the error on f_{NL} is proportional to $|b_\delta^A b_\phi^B - b_\delta^B b_\phi^A|^{-1}$ [34]. If $b_\phi \propto b_\delta - 1$ then this becomes $(b_\delta^B - b_\delta^A)^{-2}$. We will work with this approximation in what follows, but it is not crucial to the spirit of our analysis.

The main point we wish to make is that OT reconstruction naturally provides us with at least three tracer fields, in addition to the observed, unreconstructed one, which can be used to address such problems: the mass-weighted and unweighted halo reconstructed fields, and Θ (not to mention linear combinations of these fields). Roughly, such studies require $\bar{n}(b_\delta^B - b_\delta^A)^2 P_{mm} \gg 1$. If b_δ^B and b_δ^A refer to halos before and after reconstruction, then $b_\delta^A = b_\delta^B - 1$, so $b_\delta^B - b_\delta^A = 1$, and we require $\bar{n} P_{mm} \gg 1$. Typically, surveys have $\bar{n}(b_\delta^B)^2 P_{mm} \sim 3$, so we would like a large enough number density of tracers that $\bar{n} P_{mm} \gg (3/(b_\delta^B)^2)$. (For example, in Fig. 5, the effective number density is $\sim 10^{-4}$ so the (un)reconstructed halos have $\bar{n} b^2 P \sim (10)3$.) Of course, sample B could be mass-weighted halos at their evolved positions, and sample A could be the unweighted halos at their reconstructed

positions: this would increase $b_\delta^B - b_\delta^A$, and so improve the constraint on f_{NL}.

In this context, it is worth mentioning that there exists a well-developed literature on how to best weight biased tracers such as halos so as to reproduce the full matter distribution with as little stochasticity as possible [30,35]. For halos, the optimal weight (in a linear least-squares sense), is the sum of two terms: the halo mass and a Wiener-filter-like term which involves the halo-dust cross-correlation function. Since OT works most easily with mass-weighted halos, it would be natural to apply these optimal weighting schemes to the evolved *and* reconstructed halo fields, before combining them in a multitracer analysis. We leave this to future work.

V. DISCUSSION AND CONCLUSIONS

Optimal transport reconstruction returns the set of displacements which map a clustered point distribution to a uniform field. If the displacements are gradients of a potential, they are exactly reconstructed. Although the reconstructed point distribution is uniform, the field defined by the divergence of the displacements, $\Theta(\mathbf{q})$ is not. We showed that this field is quite well correlated with the initial density fluctuation field, evolved using linear theory (purple curve in Fig. 1). That is, $\Theta(\mathbf{q})$ represents the OT reconstruction of the linearly evolved field starting from the nonlinearly evolved field, even though OT makes no assumption about the amplitude of linear theory growth. We argued that because OT reconstructs displacements rather than densities or velocities, there is an upper limit to the fidelity with which $\Theta(\mathbf{q})$ can trace the linear theory overdensity field (red or green curves in Fig. 1). Nevertheless, the pair-correlation function of this OT reconstructed field, $\xi(|\mathbf{q}_i - \mathbf{q}_j|) \equiv \langle \Theta(\mathbf{q}_i)\Theta(\mathbf{q}_j)\rangle$, is extremely similar to that of the pair-correlation function in linear theory $\xi_{Lin}(|\mathbf{q}|)$, both in amplitude and shape (Fig. 3), on the large scales that are relevant to BAO physics. This is also true of its power spectrum: $P_{\Theta\Theta}(|\mathbf{k}|) \approx P_{Lin}(|\mathbf{k}|)$ (Fig. 2).

In practice, only observations of the evolved positions of a subset of the field will be available, and these will suffer from redshift-space distortions. Previous work has shown how to estimate the distribution of the remaining mass, the dust, from the observed subset, before the OT reconstruction can be performed. In principle, inaccuracies in this estimate of the dust can impact the fidelity of the OT reconstructions of the biased tracers. In practice, even the simplest Wiener-filter model of the dust allows rather good reconstruction of the displacements of the biased tracers. This remains true for the full field (biased tracers plus dust) as well: although the fidelity of the reconstruction is not as good as in the ideal case, it is still rather good on BAO scales (compare orange and blue curves in Fig. 4). In particular, the reconstructed power spectra and correlation functions are in good agreement with the linear theory shapes and amplitudes, both for the biased subset (yellow and green curves in Figs. 5 and 6

agree) and for the full field (purple and pink curves agree), on BAO scales.

Previous OT work has made the point that the reconstructed BAO feature in the clustering signal of the biased subset can be used to estimate the cosmological distance scale [12]. Our work now allows the full reconstructed field to be used as well. Since the two fields are correlated with one another, this raises the question of how best to combine the two. The amplitude of the correlation function in either of the two fields is smaller than in the original nonlinear field (compare orange and yellow curves with blue in Figs. 5 and 6). However, the signal in the field that is a simple sum of the two fields has the same amplitude as the original measurement, but with an enhanced BAO feature (red curves in Figs. 5 and 6). We discussed why summing the two fields is physically reasonable, and why peak theory provides a useful framework for understanding why this enhances the BAO feature: peak theory explains both this enhancement and the fact that the OT displacements for the halo field are slightly smaller than for the total field. Finally, we discussed how this summed field, and our treatment of the dust, compare to other reconstruction methods in the literature (Sec. III C).

It is encouraging that the OT methodology can be used to estimate the linear theory correlations. Nevertheless, there is room for improvement. We have assumed that divergence of OT displacement is a good estimator for linear theory density, but it may be possible to do better. In addition, our current Wiener-filter model for the dust degrades the fidelity of the reconstruction (orange curve lies below the blue in Fig. 4). This "linear" dust model can be improved by including the fact that the dust-tracer correlation depends on the mass of the tracers and their environment [30,36]. Alternatively, recent machine-learning based methods may provide better "nonlinear" dust assignment models [37,38]. Incorporating these improvements into our approach is the subject of work in progress.

We believe there is good reason for doing so. Some cosmological analyses benefit from having multiple tracer fields over the same survey volume (Sec. IV). OT reconstruction naturally provides us with at least four tracer fields, in addition to the observed, unreconstructed one: the unweighted, mass- and optimally weighted reconstructed halo fields, and Θ (not to mention combinations of these fields). For such studies, the inclusion of higher-fidelity dust models and optimally weighting the reconstructed fields should reap rich dividends.

ACKNOWLEDGMENTS

APPENDIX A: TECHNICAL DETAILS ABOUT ESTIMATING DIVERGENCES

In the initial conditions of the simulations, we have $\boldsymbol{v}(\boldsymbol{r})$ on a grid in configuration space. We do not estimate the divergence in configuration space. Instead, we perform a FFT to obtain $\boldsymbol{v}(\boldsymbol{k})$ on a grid in Fourier space. We then evaluate $i\boldsymbol{k} \cdot \boldsymbol{v}(\boldsymbol{k})$ at each Fourier-space grid point and divide by $(afH)_{\text{init}}$ to get $\Theta_{\text{Zel}}(\boldsymbol{k})$. Finally, if needed, we inverse-FFT back to real space. The power spectrum of the divergence is determined by summing the squared magnitudes of the components of $\Theta_{\text{Zel}}(\boldsymbol{k})$ over all available $\boldsymbol{k}$ modes.

In practice, for OT reconstructions, we estimate the divergence by interpolating the displacements $\boldsymbol{S}_{\text{OT}}(\boldsymbol{r})$ of the reconstructed Laguerre cell barycenters onto a real-space grid. We then perform an FFT to obtain $\boldsymbol{S}_{\text{OT}}(\boldsymbol{k})$ on a grid in Fourier space, evaluate $i\boldsymbol{k} \cdot \boldsymbol{S}_{\text{OT}}(\boldsymbol{k})$ at each Fourier-space grid point, and finally, compute the power spectrum. However, in sparse tracer fields, there may be many empty voxels. Assigning zero velocity to such voxels is incorrect. To address this in our OT reconstruction measurements, we compute the divergence of both halo and dust displacements, since this combined field is not sparse. We also perform different convergence tests, such as varying grid sizes and mass-assignment schemes for the displacement field, to effectively mitigate this issue.

For our analyses, we chose a three-dimensional grid with 512 points along each dimension, resulting in a total 512^3 grid points or voxels. We also used the triangular-shaped cloud (TSC) mass-assignment scheme for assigning mass (or displacement values) to the grid points. The TSC scheme is a higher-order mass-assignment method that distributes the mass of particles over a local neighborhood of grid points, using a weighted average where the weights decrease with distance from the particle's actual position. Compared to simpler schemes such as nearest grid point or cloud-in-cell, TSC offers improved accuracy in representing the density and displacement fields by considering a larger neighborhood around each point. This smoother representation of the field is also advantageous in Fourier-space analysis and in reducing numerical artifacts.

APPENDIX B: EFFECT OF POINT-MASS HALOS ONLY

The main text showed results which reconstruct the initial protohalo and matter fields starting from biased tracers in redshift space, with a simple Wiener-filter model for the dust. This model leads to the lowest of the three $r(k)$ curves shown in Fig. 4. The middle curve there starts from
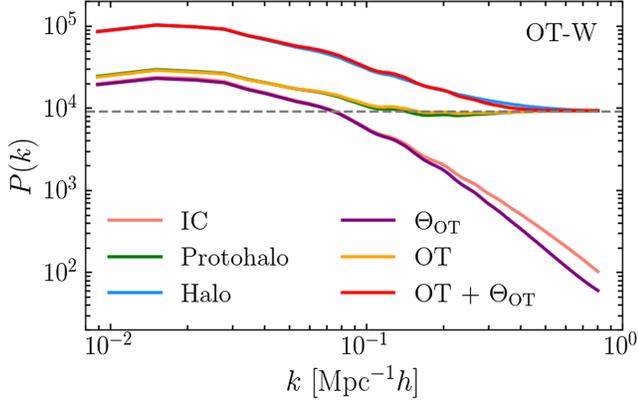
FIG. 9. Same as Fig. 5 but when the OT algorithm starts with the (point-mass) halos in real space and the correct distribution of dust.
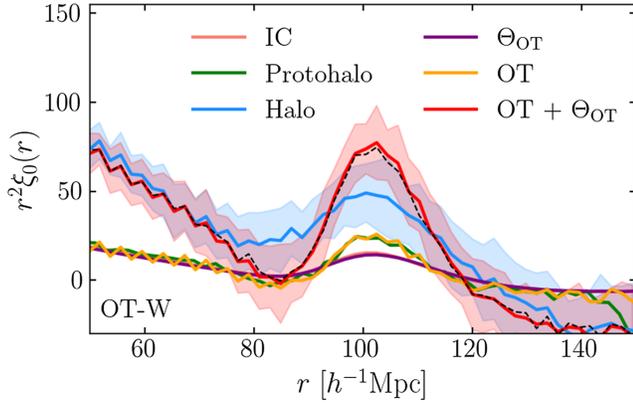
FIG. 10. Correlation functions corresponding to the power spectra shown in Fig. 9. The dashed black curve shows the correlation function of the sum of the actual protohalo and the initial fluctuation fields.

the same halos, but in configuration space, and uses the true dust. This cannot be done in real data, but, for completeness, Figs. 9 and 10 show the associated $P(k)$ and $\xi(r)$. In contrast to Figs. 5 and 6, the reconstructed (red) and input (blue) curves have the same amplitude (because the blue curve is not redshift-space distorted).

APPENDIX C: CONNECTION TO RECONSTRUCTION USING MULTIGRID GAUSS-SEIDEL RELAXATION

Reference [29] describes a multigrid Gauss-Seidel relaxation method for reconstructing the initial field from the evolved one. We discuss how it compares to our method here. Related to this discussion, and as mentioned in the text, the displacement can be equivalently thought of as a function of $\boldsymbol{q}$ or $\boldsymbol{x}$. Here, we also show explicitly the difference between taking the derivative with respect to $\boldsymbol{x}$ and $\boldsymbol{q}$.

1. Displacement potential in the initial field

We start with the gradient of Eq. (2), because it is related to the nonlinear density:

$$1 + \delta(\boldsymbol{x}) = \frac{1}{|d\boldsymbol{x}/d\boldsymbol{q}|}. \tag{C1}$$

If the displacements $\boldsymbol{S}$ are gradients of a potential $\psi(\boldsymbol{q})$, then the nonlinear density is the (inverse of the) Jacobian determinant of the "deformation tensor" of second derivatives of the potential. In this case, we can write the determinant as

$$\frac{d\boldsymbol{x}}{d\boldsymbol{q}} = \det|\delta_{ij} - \psi_{ij}| = 1 - I_1 + I_2 - I_3, \tag{C2}$$

where

$$I_1 = \text{Tr}(\psi_{ii}) = \sum_i \psi_{ii} = -\boldsymbol{\nabla}_q \cdot \boldsymbol{S},$$

$$I_2 = \sum_{j>i} \psi_{ii}\psi_{jj} - \psi_{ij}\psi_{ij},$$

$$I_3 = \det(\psi_{ij}). \tag{C3}$$

Thus,

$$1 + \delta(\boldsymbol{x}) = \frac{1}{|d\boldsymbol{x}/d\boldsymbol{q}|} = 1 + I_1 + \frac{2}{3}I_1^2 + \frac{I_1^2 - 3I_2}{3} + \ldots; \tag{C4}$$

although the nonlinear density is given by the determinant of the deformation tensor, the leading-order term, I_1, is the trace of the deformation tensor. It is useful to think of this trace as the sum of the three eigenvalues of the tensor, because

$$\boldsymbol{\nabla}_q \cdot \boldsymbol{S} = \boldsymbol{\nabla}_q \cdot \boldsymbol{x} - 3 = \sum_{i=1}^3 (1 - \lambda_i) - 3 = -\sum_{i=1}^3 \lambda_i. \tag{C5}$$

That is, the trace equals the divergence (at $\boldsymbol{q}$) of the displacements, which is why this divergence played a leading role in the main text. Note that the divergence evaluated at $\boldsymbol{x}$ (rather than $\boldsymbol{q}$) of the same displacements is

$$\boldsymbol{\nabla}_x \cdot \boldsymbol{S} = 3 - \sum_{i=1}^3 (1 - \lambda_i)^{-1} = -\sum_{i=1}^3 \frac{\lambda_i}{1 - \lambda_i}. \tag{C6}$$

This nonlinear divergence diverges as $\lambda_i \to 1$.

The combinations

$$\delta_{\text{L}} \equiv I_1, \quad q_{\text{L}}^2 \equiv I_1^2 - 3I_2, \quad u_{\text{L}}^3 \equiv \frac{2I_1^3}{9} - I_1 I_2 + 3I_3 \tag{C7}$$

are special. If we define

$$\bar{\lambda}_i \equiv \lambda_i - \frac{\sum_i \lambda_i}{3}, \qquad (C8)$$

then

$$\delta_{\mathrm{L}} = \sum_i \lambda_i, \quad q_{\mathrm{L}}^2 = \frac{\sum_i \bar{\lambda}_i^2}{2/3} \quad \text{and} \quad u_{\mathrm{L}}^3 = \sum_i \bar{\lambda}_i^3. \quad (C9)$$

The analysis above shows that δ_{L} equals the divergence (at the initial position $\boldsymbol{q}$) of displacements $\delta_{\mathrm{L}} = -\boldsymbol{\nabla}_q \cdot \boldsymbol{S}$. The quantity q_{L}^2 is the "traceless shear" [e.g. [39]]. In a Gaussian random field, δ_{L} is a Gaussian variable, q_{L}^2 is distributed as χ_5^2 and is independent of δ_{L}, and $\mu_{\mathrm{L}} \equiv (9u_{\mathrm{L}}^3/2)/q_{\mathrm{L}}^3$ is distributed like the cosine of an angle: it is uniform over $-1 \le \mu_{\mathrm{L}} \le 1$.

In terms of these variables, Eq. (C2) reads

$$\frac{d\boldsymbol{x}}{d\boldsymbol{q}} = (1 - \delta_{\mathrm{L}}/3)^3 (1 - Q_{\mathrm{L}}^2/3 - U_{\mathrm{L}}^3/3), \quad (C10)$$

where $Q_{\mathrm{L}} \equiv q_{\mathrm{L}}/(1 - \delta_{\mathrm{L}}/3)$ and $U_{\mathrm{L}} \equiv u_{\mathrm{L}}/(1 - \delta_{\mathrm{L}}/3)$. This form will be useful below, in part because q_{L} and u_{L} both vanish in spherical symmetry (when all the λ_i are equal then all $\bar{\lambda}_i = 0$), so the first term isolates the spherical contribution, and the second term encodes the corrections from asphericity.

2. Displacement potential in the displaced field

Reference [29] assumes that

$$1 + \delta_{\mathrm{NL}}(\boldsymbol{x}) = \frac{d\boldsymbol{q}}{d\boldsymbol{x}} = \det|\delta_{ij} + \zeta_{ij}|, \quad (C11)$$

where $\boldsymbol{q} = \boldsymbol{x} - \boldsymbol{S}$. The difference in notation compared to Eq. (C1) is intended to highlight the fact that here we will work with derivatives with respect to $\boldsymbol{x}$, whereas until now all derivatives were with respect to $\boldsymbol{q}$.

The second equality follows from assuming that the displacements $\boldsymbol{S}$ are gradients of a potential $\zeta(\boldsymbol{x})$, so the determinant on the right-hand side involves second derivatives (with respect to the evolved position $\boldsymbol{x}$) of $\zeta(\boldsymbol{x})$. However, if the flow is potential, then the displacement $\boldsymbol{S} = \boldsymbol{x} - \boldsymbol{q}$ can also be thought of as the gradient with respect to the initial position $\boldsymbol{q}$ of a (different) potential $\psi(\boldsymbol{q})$: In particular, $\nabla_q \psi = \boldsymbol{S} = \nabla_x \zeta$, and the potentials $\zeta(\boldsymbol{x})$ and $\psi(\boldsymbol{q})$ are Legendre transforms of one another [3]. The main point we make below is that Ref. [29] aims to estimate $\zeta(\boldsymbol{x})$ whereas we estimate $\psi(\boldsymbol{q})$ [compare Eqs. (C11) and (C1)].

To see how the methods are related, let $1 + \Lambda_i$ denote the eigenvalues of the deformation tensor $\delta_{ij} + \zeta_{ij}$. In terms of these, equation (8) of Ref. [29] reads

$$1 + \delta_{\mathrm{NL}} = (1 + \nabla^2\zeta/3)^3 (1 - Q_{\mathrm{NL}}^2/3 + U_{\mathrm{NL}}^3/3), \quad (C12)$$

where Q_{NL} and U_{NL} are the same combinations of the Λ_i that Q_{L} and U_{L} are of the λ_i. A little algebra shows that, for the right-hand sides of Eqs. (C11) and (C1) equal one another,

$$\Lambda_i(\boldsymbol{x}) = \frac{\lambda_i(\boldsymbol{q})}{1 - \lambda_i(\boldsymbol{q})}. \quad (C13)$$

This is just as expected from Eq. (C6), and is most easily shown by using Eq. (C13) in their equation (3).

Since the quantity of interest for reconstruction is δ_{L}, Sec. II. B of Ref. [29] describes the additional steps that must be taken to estimate $\delta_{\mathrm{L}}(\boldsymbol{q})$ once $\zeta(\boldsymbol{x})$ is estimated. However, Eq. (C13) shows that this is not really necessary, since

$$\delta_{\mathrm{L}}(\boldsymbol{q}) \equiv \sum_i \lambda_i(\boldsymbol{q}) = \sum_i \frac{\Lambda_i(\boldsymbol{x})}{1 + \Lambda_i(\boldsymbol{x})}, \quad (C14)$$

where it is understood that $\boldsymbol{x}$ is displaced from $\boldsymbol{q}$ by the gradient of a potential, i.e. their analysis does not exploit the fact that $\zeta(\boldsymbol{x})$ and $\psi(\boldsymbol{q})$ are related by a Legendre transform.

The analysis of Ref. [29], like that of Ref. [10], assumes knowledge of the full evolved density field $\rho(\boldsymbol{x})$. To treat the (more realistic) case in which positions of only a biased subset are known, Ref. [40] argues that one must assume a model for the bias (e.g. $\delta_b = b_1 \delta + b_2 \delta^2/2 + \cdots$), and use this to express δ in terms of δ_b. In effect, this assigns weights to the halo field. This differs from our approach, in which we assume a simple model for the dust (essentially, $\delta_b = b_1 \delta$; while it is potentially interesting to extend the dust model to incorporate high-order bias models, this would come with the cost of introducing more free parameters). Moreover, as Ref. [12] notes, because we use dust rather than a modified weight at the location of each halo, our approach accurately recovers the displacement vectors of halos, as well as the evolution of their shapes, extremely well. This matters because the protohalo centers are *not* uniformly distributed in the initial conditions, so an approach which weights the halo field will have trouble returning a displacement field which is physically intuitive.

An additional benefit of our methodology is that, in contrast to Gauss-Seidel relaxation, the OT algorithm is guaranteed to converge to the unique solution [9]. In Ref. [29] BAO reconstruction problem was studied also because of its mathematical similarities to a particular class of modified gravity theories: the field equation in the "quartic Galileon model" is also a nonlinear elliptical partial differential equation. Having demonstrated the correspondence between the semidiscrete OT algorithm and the Gauss-Seidel relaxation one, our work suggests that it would be interesting to see if semidiscrete OT is useful in the context of quartic Galileons, especially because Gauss-Seidel has linear speed of convergence, whereas for OT it is nearly quadratic.

[1] P. Peebles, *The Large-Scale Structure of the Universe* (Princeton University Press, Princeton, NJ, 1980).

[2] C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics (American Mathematical Society, Providence (RI), 2003).

[3] Y. Brenier, U. Frisch, M. Henon, G. Loeper, S. Matarrese, R. Mohayaee, and A. Sobolevskii, Mon. Not. R. Astron. Soc. **346**, 501 (2003).

[4] Y. Brenier, Commun. Pure Appl. Math. **44**, 375 (1991).

[5] J. Benamou and Y. Brenier, Numer. Math. **84**, 375 (2000).

[6] T. Buchert, Mon. Not. R. Astron. Soc. **267**, 811 (1994).

[7] U. Frisch, S. Matarrese, R. Mohayaee, and A. Sobolevski, Nature (London) **417**, 260 (2002).

[8] B. Lévy, A numerical algorithm for L_2 semi-discrete optimal transport in 3D (2015), https://www.esaim-m2an.org/articles/m2an/abs/2015/06/m2an150137/m2an150137.html.

[9] B. T. Jun Kitagawa and Quentin Mérigot, J. Eur. Math. Soc. **21**, 2603 (2019).

[10] S. von Hausegger, B. Lévy, and R. Mohayaee, Phys. Rev. Lett. **128**, 201302 (2022).

[11] B. Levy, R. Mohayaee, and S. von Hausegger, Mon. Not. R. Astron. Soc. **506**, 1165 (2021).

[12] F. Nikakhtar, R. K. Sheth, B. Lévy, and R. Mohayaee, Phys. Rev. Lett. **129**, 251101 (2022).

[13] N. Padmanabhan, X. Xu, D. J. Eisenstein, R. Scalzo, A. J. Cuesta, K. T. Mehta, and E. Kazin, Mon. Not. R. Astron. Soc. **427**, 2132 (2012).

[14] F. Nikakhtar, N. Padmanabhan, B. Lévy, R. K. Sheth, and R. Mohayaee, Phys. Rev. D **108**, 083534 (2023).

[15] R. Hada and D. J. Eisenstein, Mon. Not. R. Astron. Soc. **478**, 1866 (2018).

[16] A. Nusser and E. Branchini, Mon. Not. R. Astron. Soc. **313**, 587 (2000).

[17] E. Sarpa, C. Schimd, E. Branchini, and S. Matarrese, Mon. Not. R. Astron. Soc. **484**, 3818 (2019).

[18] F. Villaescusa-Navarro, A. Banerjee, N. Dalal, E. Castorina, R. Scoccimarro, R. Angulo, and D. N. Spergel, Astrophys. J. **861**, 53 (2018).

[19] A. N. Taylor and A. J. S. Hamilton, Mon. Not. R. Astron. Soc. **282**, 767 (1996).

[20] H. J. Mo and S. D. M. White, Mon. Not. R. Astron. Soc. **282**, 347 (1996).

[21] N. Kaiser, Mon. Not. R. Astron. Soc. **227**, 1 (1987).

[22] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, Astrophys. J. **304**, 15 (1986).

[23] V. Desjacques and R. K. Sheth, Phys. Rev. D **81**, 023526 (2010).

[24] R. K. Sheth and A. Diaferio, Mon. Not. R. Astron. Soc. **322**, 901 (2001).

[25] V. Desjacques, M. Crocce, R. Scoccimarro, and R. K. Sheth, Phys. Rev. D **82**, 103529 (2010).

[26] D. J. Eisenstein, H.-J. Seo, E. Sirko, and D. Spergel, Astrophys. J. **664**, 675 (2007).

[27] C. Hikage, K. Koyama, and A. Heavens, Phys. Rev. D **96**, 043513 (2017).

[28] X. Chen and N. Padmanabhan, arXiv:2311.09531.

[29] Y. Shi, M. Cautun, and B. Li, Phys. Rev. D **97**, 023505 (2018).

[30] Y.-C. Cai, G. Bernstein, and R. K. Sheth, Mon. Not. R. Astron. Soc. **412**, 995 (2011).

[31] R. Mohayaee, H. Mathis, S. Colombi, and J. Silk, Mon. Not. R. Astron. Soc. **365**, 939 (2006).

[32] N. Dalal, O. Doré, D. Huterer, and A. Shirokov, Phys. Rev. D **77**, 123514 (2008).

[33] P. McDonald and U. Seljak, J. Cosmol. Astropart. Phys. 10 (2009) 007.

[34] A. Barreira and E. Krause, J. Cosmol. Astropart. Phys. 10 (2023) 044.

[35] N. Hamaus, U. Seljak, and V. Desjacques, Phys. Rev. D **84**, 083509 (2011).

[36] F. Fang, Y.-C. Cai, Z. Li, S. Yue, W. Zhu, and L. Feng, arXiv:2311.15903.

[37] P. Ganeshaiah Veena, R. Lilow, and A. Nusser, Mon. Not. R. Astron. Soc. **522**, 5291 (2023).

[38] X. Chen, F. Zhu, S. Gaines, and N. Padmanabhan, Mon. Not. R. Astron. Soc. **523**, 6272 (2023).

[39] R. K. Sheth, K. C. Chan, and R. Scoccimarro, Phys. Rev. D **87**, 083002 (2013).

[40] J. Birkin, B. Li, M. Cautun, and Y. Shi, Mon. Not. R. Astron. Soc. **483**, 5267 (2019).