Estimating cosmic velocity fields from density fields and tidal tensors

Francisco-Shu Kitaura,1,2⋆ Raul E. Angulo,2 Yehuda Hoffman3 and Stefan Gottlöber1
1Leibniz-Institut für Astrophysik (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany
2Max-Planck Institut für Astrophysik (MPA), Karl-Schwarzschildstr. 1, D-85748 Garching, Germany
3Racah Institute of Physics, The Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel

Accepted 2012 June 25. Received 2012 June 21; in original form 2011 November 29

ABSTRACT
In this work we investigate the non-linear and non-local relation between cosmological density and peculiar velocity fields. Our goal is to provide an algorithm for the reconstruction of the non-linear velocity field from the fully non-linear density. We find that including the gravitational tidal field tensor using second-order Lagrangian perturbation theory based upon an estimate of the linear component of the non-linear density field significantly improves the estimate of the cosmic flow in comparison to linear theory not only in the low density, but also and more dramatically in the high-density regions. In particular we test two estimates of the linear component: the lognormal model and the iterative Lagrangian linearization. The present approach relies on a rigorous higher order Lagrangian perturbation theory analysis which incorporates a non-local relation. It does not require additional fitting from simulations being in this sense parameter free, it is independent of statistical–geometrical optimization and it is straightforward and efficient to compute. The method is demonstrated to yield an unbiased estimator of the velocity field on scales \( \gtrsim 5 \, h^{-1} \) Mpc with closely Gaussian distributed errors. Moreover, the statistics of the divergence of the peculiar velocity field is extremely well recovered showing a good agreement with the true one from \( N \)-body simulations. The typical errors of about 10 km s\(^{-1}\) (1σ confidence intervals) are reduced by more than 80 per cent with respect to linear theory in the scale range between 5 and 10 \( h^{-1} \) Mpc in high-density regions (\( \delta > 2 \)). We also find that iterative Lagrangian linearization is significantly superior in the low-density regime with respect to the lognormal model.

Key words: catalogues – galaxies: clusters: general – galaxies: statistics – large-scale structure of Universe.

1 INTRODUCTION
Gravitational instability is one of the key ingredients to explain the rich hierarchy of structures we observe today in the Universe. It has amplified small mass fluctuations produced after inflation to give rise from large galaxy clusters to huge voids. Simultaneously, the same local gravitational field imprinted ‘peculiar’ velocities in galaxies; deviations from the overall expansion of the Universe which is responsible for the Hubble flow.

The peculiar velocity of galaxies is a valuable quantity in cosmology since it contains similar but complementary information to that enclosed in the galaxies position. For instance, by requiring isotropy in the measured galaxy clustering, the cosmological mass density parameter and even the nature of gravity can be constrained (see e.g. Davis, Nusser & Willick 1996; Willick & Strauss 1998; Branchini et al. 2001; Guzzo et al. 2008). In addition, these motions can be used to reconstruct the properties of the universe at an earlier time, in principle, even at recombination where perturbations were completely linear (Nusser & Dekel 1992; Gramann 1993a; Croft & Gaztanaga 1997; Narayanan & Weinberg 1998; Monaco & Efstathiou 1999; Frisch et al. 2002).

A method able to accurately determine the peculiar velocity field can be used in many different applications; ranging from bias studies combining galaxy redshift surveys with measured peculiar velocities (see e.g. Fisher et al. 1995; Zaroubi, Hoffman & Dekel 1999; Courtois et al. 2011), baryon acoustic oscillations reconstructions (see e.g. Eisenstein et al. 2007), determination of the growth factor, to estimates of the initial conditions of the Universe which in turn can be used for constrained simulations (see e.g. Klypin et al. 2003; Gottloeber, Hoffman & Yepes 2010). A particularly well suited application regards the topological methods to detect the kinematic Sunyaev–Zel’dovich effect.

© 2012 The Authors
Monthly Notices of the Royal Astronomical Society © 2012 RAS

*E-mail: kitaura@aip.de
†Karl-Schwarzschild fellow.
There is in addition recent interest in the measurement of large-scale flows. Some authors claim to have detected a so-called ‘dark flow’ caused by superhorizon perturbations (see e.g. Kashlinsky et al. 2009; Kashlinsky, Astero-Barandela & Ebeling 2011). Others discuss such flows as a challenge to the standard cosmological model as a whole (see e.g. Watkins, Feldman & Hudson 2009).

Unfortunately, the apparent shift in spectral features of a galaxy is also affected by the expansion of the universe, therefore, it is not possible to directly measure the peculiar velocity field in spectroscopic surveys. For this reason, one has to resort to indirectly infer it from the mass density fluctuations (but see Nusser, Branchini & Davis 2011, for a recent alternative method). However, this is not a trivial procedure due the highly non-linear state of the density fluctuations today and due to its non-local relationship with the velocity field.

The simplest approach is given by the linear continuity equation, which is routinely used in clustering studies. However, it has a range of applicability only limited to very large scales (e.g. Angulo et al. 2008). Alternative methods devised to improve upon linear theory can be separated into three areas. The first one consists on developing non-linear relationships with higher order perturbation theory (Bernardeau 1992; Chodorowski et al. 1998; Bernardeau et al. 1999; Kudlicki et al. 2000), with spherical collapse models (Bilicki & Chodorowski 2008) or based on empirical relations found in cosmological N-body simulations (Nusser et al. 1991; Willick et al. 1997).

Another idea is to solve the boundary problem of finding the initial positions of a set of particles governed by the Eulerian equation of motion and gravity with the least action principle (see Peebles 1989; Nusser & Branchini 2000; Branchini et al. 2002). A similar approach consists on relating the observed positions of matter tracers (e.g. galaxies) in a geometrical way to a homogeneous distribution by minimizing a cost function, which combined with the Zel’dovich approximation (Zel’dovich 1970) provides an estimation of the velocity field (see Mohayaee & Tully 2005; Lavaux et al. 2008).

The diversity of strategies and approximations for obtaining the velocity from the density field hint at the difficulty of the problem. Some approaches are simply not accurate enough and some are computationally very expensive. This sets the agenda for an improvement in the field. Any new method should be accurate, unbiased, computationally efficient and applicable to observational data.

A further shortcoming of the existing approaches is that they are mostly particle based, which is not applicable for more general matter tracers like the Lyman α forest or the 21-cm line, nor they can be combined with optimal density field estimators (see Jasche & Kitaura 2009; Kitaura, Jasche & Metcalf 2010a; Kitaura, Gallerani & Ferrara 2010b).

In this paper we investigate a different approach based on higher order Lagrangian perturbation theory (LPT), and it is motivated by the pioneering work of Gramann (1993b) and further extended by Hivon et al. (1995) and Monaco & Efstathiou (1999). The theoretical basis for LPT was carefully worked out by Moutarde et al. (1991), Buchert & Ehlers (1993), Buchert (1994), Bouchet et al. (1995) and Catelan (1995) (for further references see Bernardeau et al. 2002).

Contrary to classical applications of LPT, in which the properties of an evolved distribution are predicted from a linear density field in Lagrangian coordinates (e.g. in the generation of initial conditions for N-body simulations or of galaxy mock catalogues; Scoccimarro & Sheth 2002; Jenkins 2010), our starting point is an evolved field in Eulerian coordinates (e.g. the present-day galaxy distribution). The key realization of our approach is that it is possible to decompose a non-linear density field into a Gaussian and non-Gaussian term, which are related to each other through LPT (see similar approaches Gramann 1993a,b; Monaco & Efstathiou 1999). In other words, it is possible to find a closely Gaussian field which would evolve, under the assumption of LPT, into the measured density field today. This Gaussian field can then naturally be used to predict the corresponding velocity field today in LPT.

Our method combines (i) the equations of motion and continuity for a fluid under self-gravity in second (third) order LPT (2LPT (3LPT)) with (ii) the idea that the present-day galaxy distribution can be expanded into a closely linear-Gaussian field and a highly skewed non-linear component consistent with 2LPT (or 3LPT) (see Kitaura & Angulo 2012). The former aspect makes our approach physically motivated and also captures the non-local nature of the density–velocity relation via the gravitational tidal field tensor. The latter aspect self-consistently minimizes the impact of the approximations of a second-order formulation of gravitational evolution, but more importantly, it enables the application of LPT to non-linear fields.

We note that the use of the lognormal transformation (including the subtraction of a mean field) to obtain an estimate of the linear field was proposed by Kitaura et al. (2010b) and has been recently confirmed to give already a good estimate of the divergence of the displacement field (Falck et al. 2012). To estimate the velocity field one needs, however, to go to higher order perturbation theory as we show in this work.

In the next section we recap LPT and derive the velocity–density relation to second and third order. In Section 2.2 we will present the method to compute the peculiar velocity field from the non-linear density field. In Section 3 we will present our numerical tests based on the Millennium Run simulation. Here we will analyse the goodness of the recovered velocity divergence as compared to the true one and the same for each velocity component. A study of the errors in our method is also presented. Finally, we present our conclusions.

2 VELOCITY–DENSITY RELATION

The first part of this section presents the relation between density and velocity fields in 2LPT, and how it can be applied to an evolved field. In the second part, we outline a practical implementation of this method.

2.1 Second-order Lagrangian perturbation theory

The basic idea in LPT is that an initial homogeneous field expressed in Lagrangian coordinates \( q \) can be related to a final field in Eulerian coordinates \( x \) trough a unique mapping: \( x = q + \Psi(q) \) determined by a displacement field \( \Psi \) (see e.g. Bernardeau et al. 2002). The linear solution for this expression is the well-known Zel’dovich approximation, in which the displacement field is given by the Laplacian of the gravitational potential at \( q \). This result has been successfully applied to many aspects of cosmology, but it fails to describe the dynamics of a non-linear field where shell crossings and curved trajectories are common.

An improvement is found by expanding the displacement field and considering higher order terms (see e.g. Buchert, Melott & Weiss 1994; Bouchet et al. 1995; Melott, Buchert & Weib 1995). For instance, the displacement field to second order is given by

\[
\Psi(q) = -D_1 \nabla_q \phi^{(1)}(q) + D_2 \nabla_q \phi^{(2)}(q),
\]

where \( \phi^{(1)} \) and \( \phi^{(2)} \) are the first and second order linear density fields, respectively. The coefficients \( D_1 \) and \( D_2 \) are determined by matching the linear and non-linear functions at the boundaries of the shell crossings.

The key to the approach is that the second-order displacements are obtained from the solutions of the linear fields, which are computationally much cheaper and more accurate than the non-linear fields. This allows for a more accurate estimation of the velocity field, especially in regions with shell crossings.

In the next section, we will present a practical implementation of this method, including numerical tests and comparisons with observed data.
where $D_1$ is the linear growth factor normalized to unity today, $D_2$ the second-order growth factor, which is given by $D_2 \simeq -3/7 \Omega^{-1/4} \Omega_1^2$ (see Buchert & Ehlers 1993; Bouchet et al. 1995). The potentials $\Phi^{(1)}(q)$ and $\Phi^{(2)}(q)$ are obtained by solving a pair of Poisson equations: $\nabla^2 \Phi^{(1)}(q) = \delta^{(1)}(q)$, where $\delta^{(1)}$ is the linear overdensity, and $\nabla^2 \Phi^{(2)}(q) = \delta^{(2)}(q)$.

It is important to realize that these terms are not independent of each other. The second-order non-linear term $\delta^{(2)}$ is fully determined by the linear overdensity field $\delta^{(1)}$ through the following quadratic expression:

$$\delta^{(2)}(q) \equiv \mu^{(2)}(\Phi^{(1)}(q)) = \sum_{i,j} \left( \langle \Phi^{(1)}(q)_i, \Phi^{(1)}(q)_j \rangle - [\delta^{(1)}(q)]_{i,j}^2 \right),$$

where we use the following notation $\Phi^{(1)}(q)_i \equiv \Phi^{(1)}(q) / \partial q_i$ and the indices $i,j$ run over the three Cartesian coordinates.

Similarly, the particle comoving velocities $v$ are given to second order by

$$v(q) = -D_1 f_1 H \nabla \Phi^{(1)}(q) + D_2 f_2 H \nabla \Phi^{(2)}(q),$$

where $f_i = \ln D_1 / \ln a$, $H$ is the Hubble constant and $a$ is the scale factor. For flat models with a non-zero cosmological constant, the following relations apply: $f_1 \equiv \Omega^{5/6} f_2 \approx 2 \Omega^{5/6} f_1$ (see Bouchet et al. 1995), where $\Omega(z)$ is the matter density at a redshift $z$.

We note that going to third order in LPT provides modest improvements (see Buchert & Ehlers 1993; Bouchet et al. 1995; Catelan 1995; Melott et al. 1995; Bernardeau et al. 2002).

To apply the Lagrangian framework to a density field in the Eulerian frame $\delta_\text{E}(x)$ one must be careful. Under the assumption that there is no shell-crossing, one can write the inverse equation for the linear overdensity field $\delta^{(1)}$ as $\mu^{(1)}(\Phi^{(1)}(q))$ and the following relations apply:

$$\delta^{(1)}(x) = \int \mu^{(1)}(\Phi^{(1)}(q)) \, dq = \int \delta^{(1)}(q) \, dq.$$

Expanding the Jacobian $J(x)$ one finds (Kitaura & Angulo 2012)

$$\delta_M(x) = 1 + \nabla \cdot \Phi^{(1)}(x) - 1 \equiv -\nabla \cdot \Phi^{(1)}(x) + \mu^{(2)}(\Theta(x)) + \mu^{(3)}(\Theta(x)).$$

The displacement or velocity field derived from this expression will automatically be expressed in Eulerian coordinates as we will show below. We note that the same expression is found as a function of the Lagrangian coordinate $q$ when expanding the inverse of the corresponding Jacobian $J(q) \equiv |\partial q / \partial x|$ (see Kitaura & Angulo 2012).

Using the displacement field (equation 1) in Eulerian coordinates, the final density field can be written in terms of the linear and non-linear fields:

$$\delta_M(x) = D_1 \delta^{(1)}(x) - D_2 \mu^{(2)}(\Phi^{(1)}(x)) + \mu^{(2)}(\Theta(x)) + \mu^{(3)}(\Theta(x)) = \delta^{(1)}(x) + \delta^{\text{NL}}(x),$$

with $\Theta(x)$ being the potential associated with the divergence of the displacement field: $\Theta(x) = -\nabla \cdot \Phi^{(1)}(x)$ and $\delta^{\text{NL}}(x) = -D_2 \mu^{(2)}(\Phi^{(1)}(x)) + \mu^{(2)}(\Theta(x)) + \mu^{(3)}(\Theta(x))$. The third-order term in the Jacobian expansion is given by

$$\mu^{(3)}(\Theta(x)) = \partial_i (\Theta(x)_i).$$

From now on the Eulerian coordinate dependence $(x)$ is left out.

Assuming that any primordial vorticity has no growing modes associated (the first vorticity terms appear in 3LPT, see Appendix A) implies that the velocity field today is fully characterized by its divergence $(\nabla \cdot v)$, or, for convenience, by the scaled velocity divergence, defined as

$$\theta \equiv -f(\Omega, \Lambda, z)^{-1} \nabla \cdot v,$$

with $f(\Omega, \Lambda, z) \equiv D_1 / D_2 = f_1 / (\Omega, \Lambda, z) H(z)$.

Combining equations (7) and (3) truncated to quadratic terms in $\Phi^{(1)}$, $\delta^{\text{NL}} = (D_1^2 - D_2^2)\theta$, one gets

$$\theta = \delta_M - \left[ D_1^2 + \left( \frac{f_2}{f_1} - 1 \right) D_2 \right] \theta^{(2)}.$$  

This expression is very similar to the one found by Gramann (1993b) (see Table 1). Note, however, that using directly the evolved field as the source for the second-order term $\theta^{(2)}$ is a good approximation for the true velocity field only on very large scales (where $\delta_M$ is close to unity), as we will see in Section 3 and Fig. 1, breaking down on scales of even $10 h^{-1}$ Mpc for both estimations of the non-linear field and of the velocity divergence.

In this paper we follow a different approach, and solve iteratively the following equation:

$$\theta = D_1 \delta^{(1)} - D_2 \frac{f_2}{f_1} \theta^{(2)},$$

which results from taking the divergence of equation (3). For this, we rely on an estimation of the linear component of the present-day density field, which in turn can be calculated by solving iteratively equation (7). Note that the Eulerian description forces one to expand the inverse of the Jacobian (see equations (4)–(7) and Kitaura & Angulo 2012). This is the main difference with respect to the work of Monaco & Efstathiou (1999) in which first the Jacobian is expanded and then the inverse of it is taken and could explain why we avoid problems caused by artificial Lagrangian caustics in low-density regions as reported in their work.

In practice, a good approximation for the linear term, $\delta^{(1)}$, is simply given by the logarithm of the density field: $\delta^{(1)} = D_1 \ln (1 + \delta_M) - \mu$, with $\mu = \ln (1 + \delta_M)$, as shown by Neyrinck, Szapudi & Szalay (2009) and Kitaura & Angulo (2012). Note that this expression is essentially the lognormal approximation for the matter field (Coles & Jones 1991). This local transformation has the advantage of being computationally inexpensive.

In summary, approach finds a linear field which when plugged into 2LPT expressions produces the observed matter field (or third order, see Appendix A). If gravity worked only at a second order level, then this linear field would be identical to the actual linear field that originated $\delta_M$, but naturally in reality it is just an approximation. Thus, it is important to characterize the performance and accuracy of the method, which we do in Section 3. But first, we discuss a practical implementation of our method in the next subsection.

### 2.2 Method

The method to estimate the peculiar velocity field from the non-linear density field is straightforward and fast to compute. We now outline the steps to be followed for its implementation. For this, we have assumed that there is an unbiased and complete estimation of the matter field $\delta_M$. The extra layer of complication introduced by shot noise, a survey mask, biasing and redshift space distortions is out of the scope of this paper and will be addressed in a future publication.

© 2012 The Authors, MNRAS 425, 2422–2435

Monthly Notices of the Royal Astronomical Society © 2012 RAS
The parameters $\alpha$, $\beta$, $\gamma$ are either derived from first principles (Bernardeau 1992; Gramann 1993b; Bilicki & Chodorowski 2008; this work) or from fitting to simulations being different for each case. These parameters also depend on the scale of interest.

The parameters which are in parenthesis are fully determined by the chosen cosmology and the theoretical model. The variance of the density field is given by $\sigma_{\delta}^2 = \langle \delta^2 \rangle$. PT stands for perturbation theory and is applied only in the univariate case (local relation). 2LPT stands for second-order Lagrangian perturbation theory and yields the only non-local (and non-linear) relation from the list. Hivon et al. (1995) and Monaco & Efstathiou (1999) proposed 2LPT to correct for redshift distortions. Let us mention here the least-action principle methods to determine the peculiar velocities from mass tracer objects (galaxies) introduced by Peebles (1989) and further extended by Nusser & Branchini (2000), Branchini, Eldar & Nusser (2002) and Lavaux et al. (2008).

| Works                        | $\theta-\delta$ relation | Parameters | Methodology                        |
|------------------------------|---------------------------|------------|-----------------------------------|
| Yahil et al. (1991)          | $\theta = \delta$         | $\{p\}$    | Linear theory: LPT               |
| Nusser et al. (1991)         | $\theta = \delta(1 + a\delta)$ | $\alpha, \{p\}$ | Empirical approximation          |
| Bernardeau (1992)            | $\theta = a\{(1 + \delta)^{\beta} - 1\}$ | $\alpha, \beta, \{p\}$ | PT                 |
| Gramann (1993b)              | $\theta = \delta - aD_1^2 \mu^{(2)}(\phi_g)$ | $\alpha, \{p\}$ | Approx. 2LPT                     |
| Willick et al. (1997)        | $\theta = \left(1 + a^2 \sigma^2 + a\sigma_g^2\right)/(1 + a\delta)$ | $\alpha, \sigma^2, \{p\}$ | Empirical approximation          |
| Chodorowski et al. (1998)    | $\theta = \left[\sigma + a\left(\frac{\sigma^2}{\delta} + \beta \delta\right)\right]$ | $\alpha, \beta, \gamma, \sigma^2, \{p\}$ | PT+N-body                       |
| Bernardeau et al. (1999)     | $\theta = \left[\frac{\sigma}{\delta} + a\left(\frac{\sigma^2}{\delta^2} + \beta \delta\right)\right]$ | $\alpha, \beta, \gamma, \sigma^2, \{p\}$ | PT+N-body                       |
| Kudlicki et al. (2000)       | $\theta = \alpha \left(1 + \delta^{1/3} - 1\right)/(2\alpha^2)\sigma^2$ | $\alpha, \sigma^2, \{p\}$ | PT+Eulerian grid-based code      |
| Bilicki & Chodorowski (2008) | $\theta = \delta^{1/3} \left(1 + \delta^{1/3} - 1\right)$ | $\alpha, \beta, \gamma, \{p\}$ | Spherical collapse model        |
| This work                    | $\theta = D_1 \delta^{(1)} - D_2 f_1 f_2 \beta^{(2)}(\phi_g)$ | $\{p\}$ | 2LPT                              |

Figure 1. Statistics of the scaled velocity divergence ($\theta_k$; true from the simulation $\delta^{N-body}$ and different reconstructed ones $\theta^{rec}$ as explained below) with different smoothing scales: left: $r_S = 5\ h^{-1}\ Mpc$; right: $10\ h^{-1}\ Mpc$. Curves from left to right in the order of appearance: yellow: approximate 2LPT solution from the non-linear density field (GRAM) (see Gramann (1993b) in Table 1); cyan: logarithm of the density field (LOG); blue dashed: true scaled velocity divergence at $z = 0$ the Millennium Run ($N$-body); black: 2LPT estimate from the logarithm of the density field (LOG–2LPT) (equation 11); red: 2LPT estimate from the iterative solution (2LPT) (see Section 2.2); green: linear theory (LIN) (density field).

(i) **Linear density field**
One starts by computing an estimate of the linear component of the density field. We propose two alternatives for this:

(a) lognormal model

$$\delta^{(1)}_{\text{LOG}} = D_1 \delta^{(1)} = \ln(1 + \delta_M) - \mu; \quad (12)$$

(b) Gaussianization with LPT (Kitauro & Angulo 2012)

$$\delta^{(1)}_{\text{LPT}} = D_1 \delta^{(1)} = \delta_M - \delta^{\text{NL}}. \quad (13)$$

(ii) **Linear potential**
Then the Poisson equation is solved to obtain the linear potential:

$$\phi^{(1)} = \nabla^{-2} \delta^{(1)}. \quad (14)$$

(iii) **Non-linear second-order density field**

The tidal field contribution to second order is calculated from the linear potential:

$$\delta^{(2)}(\phi^{(1)}) = \sum_{i,j} \left(\phi^{(1)}_{ij} (\phi^{(1)}_{ij}) - (\phi^{(1)}_{ij})^2\right). \quad (15)$$

(iv) **Scaled velocity divergence**
One can now construct the second-order divergence of the velocity field taking both the linear and the second-order contribution:

$$\theta = D_1 \delta^{(1)} - D_2 f_1 f_2 \delta^{(2)}. \quad (16)$$

(v) **Peculiar velocity field**
Finally, one obtains the 3D velocity field:

$$v = -f(\Omega, \Lambda, z) \nabla \nabla^{-2} \theta. \quad (17)$$
Please note that the equations in steps (ib), (ii), (iii) and (v) can be solved with fast Fourier transform algorithms (FFTs). Details of the Gaussianization step with LPT can be found in Kitaura & Angulo (2012).

3 TESTING THE METHOD WITH N-BODY SIMULATIONS

In this section we test the performance of the method outlined above by comparing the velocity field directly extracted from a N-body simulation with our estimation based on the respective non-linear density field.

With this purpose, we employ the Millennium simulation which tracks the non-linear evolution of more than 10 billion particles in a box of comoving side length 500 $h^{-1}$ Mpc (Springel et al. 2005). In particular, we use the output corresponding to redshift 0, which is where the most non-linear structures are present.

At such output we first compute the velocity and density field by mapping the information of dark matter particles on to a grid of 2563 cells using the nearest-grid-point (NGP) assignment scheme, which gives a spatial resolution of about 2 $h^{-1}$ Mpc. We then apply the algorithm presented in Section 2.2 to infer the velocity divergence on a grid of identical dimensions. In the next two subsections we present our results and explore the accuracy when applied on two scales: 10 $h^{-1}$ Mpc on which linear theory is usually assumed to perform well, and 5 $h^{-1}$ Mpc which enters into the mildly non-linear regime.

3.1 The velocity divergence

The first ingredient in our algorithm is the linear component of the density field. We stress that this field is not the ‘initial conditions’ of the universe, since structures have moved from its Lagrangian position to the Eulerian ones at $z = 0$. In contrast to the linear term, the probability distribution function (PDF) of the non-linear component is highly skewed.

Fig. 1 compares the PDF of the velocity divergence $\theta$ as given by different estimators, with the one directly measured from the simulation (blue dashed line). The left-hand panel shows the fields smoothed with a 5 $h^{-1}$ Mpc, whereas the right-hand panels do so with a larger smoothing of 10 $h^{-1}$ Mpc. In both cases, the predictions of linear perturbation theory (i.e. $\theta = \delta$) display the worst performance of all – it overestimates systematically the number of volume elements with large value of $\theta$ and underestimates the ones with low values of $\theta$ (green curve). This is a consequence of linear theory breaking down even in mildly under- or overdense regions.

Using linear theory together with the linear component of the non-linear density field given by the logarithm of the field (ln (1 + $\delta_M$) − $\mu$, see cyan curve in Fig. 1) is a poor description of the PDF, suggesting the need of higher order corrections. Indeed, third-order PT appears to perform better than 2LPT for underdense regions (see Appendix A). In spite of this, on both 5 and 10 $h^{-1}$ Mpc our method is clearly superior to any of the other methods we investigated here, as well as the PDF is concerned and for any value of $\theta$. The iterative LPT solution yields a moderate overestimation for high values of $\theta$ due to the laminar flow approximation used in LPT which does not fully capture non-linear structure formation. Nevertheless, the PDF of $\theta$ using this solution is clearly superior to the linear approximation.

We now continue with a more detailed testing of our method. In Fig. 2 we plot the predicted velocity divergence, $\theta_{\text{LPT}}^{\text{LR}}$ and $\theta_{\text{LPT}}^{\text{LR}2}$ (based on $\delta_{\text{LOG}}^{\text{LR}}$ and $\delta_{\text{LPT}}^{\text{LR}}$, respectively), for each of the 2563 cells in our mesh, as a function of the value measured in the simulation. As in the previous plot, we display results on two different scales: 5 $h^{-1}$ Mpc on the left-hand panels and 10 $h^{-1}$ Mpc on the right-hand panels. For comparison we also provide the results using linear perturbation theory $\theta_{\text{LPT}}^{\text{LR}} = \delta$.

The values of $\theta$ are remarkably well predicted by LPT. In fact, measurements lie around the 1:1 line in the LOG–2LPT case, implying that there are no appreciable biases in our estimation over all the range probed by the Millennium simulation (with the exception of the low values of $\theta$, for an improvement on this see Appendix A). In contrast, the linear theory prediction presents overestimations of up to a factor of 3 for the 5 $h^{-1}$ Mpc smoothing and of 2 for 10 $h^{-1}$ Mpc.

The iterative solution (2LPT) produces smaller dispersions but also a slight overestimation of $\theta$ for high values, as we already mentioned before.

The distribution of differences in our method is well approximated by a Gaussian function, whereas in linear theory there are

1 We have also checked that this is true on 3, 4, 6, 7 and 8 $h^{-1}$ Mpc.

© 2012 The Authors, MNRAS 425, 2422–2435

Monthly Notices of the Royal Astronomical Society © 2012 RAS
Figure 2. Cell-to-cell comparison between the true and the reconstructed scaled velocity divergence, $\theta^{N\text{-body}}$ and $\theta^\text{rec}$, respectively. Left-hand panels show results on scales of $5 \, h^{-1}\text{Mpc}$ and right-hand panels on scales of $10 \, h^{-1}\text{Mpc}$. Upper panels: LIN; middle panels: LOG-2LPT; lower panels: 2LPT. The dark colour code indicates a high number and the light colour code a low number of cells.
Figure 3. Difference between the true and the estimated scaled velocity divergence. On scales of $r_S = 5$ (left-hand panels) and $10 \, h^{-1}$ Mpc (right-hand panels) with different colour ranges. Upper panels: LIN; middle panels: LOG–2LPT; lower panels: 2LPT.
Figure 4. Difference between the true and the estimated speeds (velocity magnitude). On scales of $r_S = 5$ (left-hand panels) and $10 h^{-1}$ Mpc (right-hand panels) with different colour ranges. Upper panels: LIN; middle panels: LOG–2LPT; lower panels: 2LPT. The colour code is in units of $10^{-2} \text{s}^{-1} \text{ km}$. 
significant extended tails, we will return to this in more detail in Section 3.2. Overall, this plot suggests that our method not only performs adequately on a statistical basis, but also on predicting the actual average value of \(\theta\) in a given volume element. Although not displayed by the figure, our method also performs better than the other methods shown in Fig. 1. In particular, the classic application of 2LPT recovers \(\theta\) quite well for the range \(-0.5 < \theta < 0\) (2LPT) (equation 11) and red: 2LPT estimate from the iterative solution (2LPT) (see Section 2.2). Note that in the case of \(10 h^{-1}\) Mpc smoothing we have multiplied the lines by a factor of 2 for clarity.

As a final crucial check, we compute the power spectra of the scaled velocity divergence \(v\) for green: linear theory (LIN), black: 2LPT estimate from the logarithm of the density field (LOG–2LPT) and red: 2LPT estimate from the iterative solution (2LPT) (see Section 2.2). We see that the 2LPT solutions perform remarkably well in a wide range going down to scales of \(k \approx 0.3\) and \(\approx 0.5 h\) Mpc \(^{-1}\) for \(r_S = 10\) and \(5 h^{-1}\) Mpc, respectively. We can also see that there is a systematic deviation originated by the lognormal transformation. The LPT estimate of the linear component corrects these and the results are extremely close to the actual power spectrum over most of the \(k\) range shown.

### 3.2 The full 3D peculiar velocity field

We have calculated each component of the peculiar velocity field \((v_x, v_y, v_z)\) from the inferred velocity divergence assuming \(\nabla \times \mathbf{v}_l\) divided by the standard deviation of the velocity divergence \(\sigma(\nabla \cdot \mathbf{v})\) in per cent as a function of the scale \(r_S\). The mean of both the curl and the divergence of the peculiar field is very close to zero.
Figure 7. Cell-to-cell comparison between the true velocity $v_{N\text{-body}}^N$ and the reconstructed one $v_{\text{rec}}^N$ for the $x$-component. Left-hand panels on scales of $5\ h^{-1}\ Mpc$, and right-hand panels on scales of $10\ h^{-1}\ Mpc$. Upper panels: LIN; middle panels: LOG–2LPT; lower panels: 2LPT (see Section 2.2).
Figure 8. Statistics of the errors in the reconstructed velocity for the $x$-component (green: LIN; black: LOG–LPT; red: 2LPT) with different smoothing scales $r_S = 5$ (left-hand panels) and $10 \, h^{-1} \text{Mpc}$ (right-hand panels). Upper panels in the entire $\delta$ range; middle panels: for the range $\delta > 2$; lower panels: for the range $\delta < 0.5$. 
very long tails. Such outliers are not present in our method (see Fig. 8).

As we have discussed along this paper, the larger improvement of our method concerns velocities in high-density regions. For regions with $\delta > 1$ and $\delta > 2$ we find significant differences between linear and 2LPT. At $10 \, h^{-1} \, \text{Mpc}$ and cells with $\delta > 2$ the 1σ errors with linear theory are about 70 km s$^{-1}$ and are reduced with 2LPT to $\sim 13$ km s$^{-1}$, i.e. 81 per cent smaller. The corresponding 2σ confidence intervals are about 160 and 28 km s$^{-1}$, respectively, i.e. an error reduction of about 83 per cent. One can see that the 2σ confidence intervals are about double as large as the 1σ confidence intervals for the 2LPT estimation. However, this is not the case for the linear estimates as these are not Gaussian distributed.

4 CONCLUSIONS AND DISCUSSION

In this paper we presented an improved method to reconstruct the peculiar velocity field from the density field. It builds from 2LPT and the realization that the density field can be split into a linear plus a non-linear term. The latter is the key concept, which enables the application of LPT to an evolved field in Eulerian coordinates. This in turn, creates an approach that is non-linear and non-local by including the information of the gravitational tidal field tensor.

We have shown that this approach is efficient and accurate over the dynamical range probed by the Millennium simulation. When the reconstruction is carried out on $5 \, h^{-1} \, \text{Mpc}$, each component of the velocity field can be recovered to an accuracy of about 10 km s$^{-1}$. On $10 \, h^{-1} \, \text{Mpc}$ this figure is reduced to about 7 km s$^{-1}$. If we consider high-density regions, the typical uncertainty is of 13 km s$^{-1}$, which improves dramatic over linear perturbation theory; typical errors are 81 per cent smaller. An accurate description of the velocity divergence, both in terms of its PDF and on a point-by-point basis, is also achieved. In addition, we have shown that the one- and two-point statistics of the scaled velocity divergence are extremely well recovered, being almost indistinguishable from the true ones. Contrarily, linear theory dramatically overestimates the velocity divergence. This especially on the mildly non-linear scale of $5 \, h^{-1} \, \text{Mpc}$ where our method shows more clearly its advantages. Finally, we highlight that our method does not require calibration nor free parameters to predict the velocity divergence field.

There are a number of simplifications and assumptions that we have adopted throughout our paper. First, our analyses focused on the peculiar velocity averaged over a volume. Another aspect is that we have neglected the rotational component of the velocity field. This, however, does not seem to be relevant at the scales we have investigated (larger than $5 \, h^{-1} \, \text{Mpc}$). Another simplification is performing our comparison assuming that there is an unbiased estimation of the underlying real-space density field. But, of course, densities measured in a survey are in redshift space. The transformation can be done, but not trivially. One alternative is to apply the Gibbs sampling method suggested by Kitaura & Enßlin (2008) and Kitaura et al. (2010b) to correct for redshift-space distortions. In this, the Gaussian distribution of errors in our method is highly convenient, since it permits to model the uncertainties in the peculiar velocity field including observational errors in a straightforward way. One should consider also classical iterative approaches pioneered by Yahil et al. (1991) based on linear theory. In a similar way they have been applied with different approximations by different groups (see Croft & Gaztanaga 1997; Monaco et al. 2000; Nusser & Branchini 2000; Branchini et al. 2002; Mohayaee & Tully 2005; Lavaux et al. 2008; Wang et al. 2012). Here it is crucial to have an accurate description relating the peculiar velocity field to the density field (or galaxy distribution) which subject of study is central in the work presented here. We expect that the improved relation found in this work with respect to linear Eulerian or linear LPT yields better estimates of the peculiar velocity field. Further studies need to be done to test the performance including redshift-space distortions.

We would like to note that our comparison and the uncertainties quoted here were based on the present-day output of the Millennium Run. Such simulation was carried out with a value for $\sigma_8$ about 10 per cent higher than the current best estimations (see Angulo & White 2010, for a method to correct for this). Therefore, our uncertainties should be regarded as an upper limit of the reachable uncertainties for a hypothetical spectroscopic survey, which should contain a less non-linear underlying dark matter distribution.

We finalize this paper by emphasizing that the method presented here can potentially be used in many different applications, and should be further developed and tested to perform bias studies combining galaxy redshift surveys with measured peculiar velocities, baryon acoustic oscillation reconstructions, determination of the growth factor, to estimates of the initial conditions of the Universe.

ACKNOWLEDGMENTS

F-SK and REA thank Simon White for encouraging conversations. The authors appreciate Francisco Prada's comments on the manuscript. We are indebted to the German Astrophysical Virtual Observatory (GAVO) and the MPA facilities for providing us the Millennium Run simulation data. The work of REA was supported by Advanced Grant 246797 ‘GALFORMOD’ from the European Research Council. YH has been partially supported by the ISF (13/08).

REFERENCES

Angulo R. E., White S. D. M., 2010, MNRAS, 405, 143
Angulo R. E., Baugh C. M., Frenk C. S., Lacey C. G., 2008, MNRAS, 383, 755
Bernardeau F., 1992, ApJ, 390, L61
Bernardeau F., Chodorowski M. J., Lokas E. L., Stompor R., Kadlicki A., 1999, MNRAS, 309, 543
Bernardeau F., Colombi S., Gaztaña E., Scoccimarro R., 2002, Phys. Rep., 367, 1
Bilicki M., Chodorowski M. J., 2008, MNRAS, 391, 1796
Bouchet F. R., Colombi S., Hivon E., Juszkiewicz R., 1995, A&A, 296, 575
Branchini E. et al., 2001, MNRAS, 326, 1191
Branchini E., Eldar A., Nusser A., 2002, MNRAS, 335, 53
Buchert T., 1994, MNRAS, 267, 811
Buchert T., Ehlers J., 1993, MNRAS, 264, 375
Buchert T., Melott A. L., Weiss A. G., 1994, A&A, 288, 349
Catelan P., 1995, MNRAS, 276, 115
Chodorowski M. J., Lokas E. L., Pollo A., Nusser A., 1998, MNRAS, 300, 1027
Coles P., Jones B., 1991, MNRAS, 248, 1
Courtois H. M., Hoffman Y., Tully R. B., Gottloeber S., 2011, ApJ, 744, 43
Croft R. A. C., Gaztaña E., 1997, MNRAS, 285, 793
Davis M., Nusser A., Willick J. A., 1996, ApJ, 473, 22
Eisenstein D. J., Seo H.-J., Spergel D. N., 2007, ApJ, 664, 675
Falck B. L., Neyrinck M. C., Aragon-Calvo M. A., Lavaux G., Szalay A. S., 2012, ApJ, 745, 17
APPENDIX A: THIRD-ORDER LAGRANGIAN PERTURBATION THEORY

For the sake of completeness we investigate here third-order LPT. Following Buchert (1994) and Catelan (1995) one can write the displacement field to third order as

\[
\Psi = -D_1 \nabla \phi^{(1)} + D_2 \nabla \phi^{(2)} + D_{3a} \nabla \phi^{(3)} + D_{3b} \nabla \times A^{(3)},
\]

(A1)

where \( D_{3a}, D_{3b}, D_3 \) are the third-order growth factors corresponding to the gradient of two scalar potentials \( \phi^{(3)}, \phi^{(3)} \) and the curl of a vector potential \( A^{(3)} \). Particular expressions for the irrotational third-order growth factors \( D_{3a}, D_{3b} \) can be found in Bouchet et al. (1995), the growth factor of the rotational term \( D_3 \) is given in Catelan (1995).

We assume that the fields are curl-free on scales \( \gtrsim 5 \, h^{-1} \text{Mpc} \) (see Bouchet et al. 1995, and the comparison between the velocity divergence and the curl in the Millennium Run in Fig. 5). We therefore consider only the scalar potential terms \( \phi^{(3)} \) and \( \phi^{(3)} \).

The first term is cubic in the linear potential:

\[
\delta^{(3)} \equiv \mu^{(3)}(\phi^{(1)}) = \text{det}(\phi^{(1)}) ,
\]

(A2)

and the second term is the interaction term between the first- and the second-order potentials:

\[
\delta^{(3)} \equiv \mu^{(2)}(\phi^{(1)}, \phi^{(2)}) = \frac{1}{2} \sum_{i<j} (\phi^{(1)}_{ij} \phi^{(2)}_{ij} - \phi^{(2)}_{ij} \phi^{(1)}_{ij})
\]

(A3)

(see Buchert 1994; Bouchet et al. 1995; Catelan 1995). In order to ensure that the term \( \delta^{(3)} \) is curl-free one has to introduce some proper weights in the expression (A3) (see Buchert 1994; Catelan 1995).

From the displacement field one can derive the expression for the velocity field:

\[
v = -D_1 f_1 H \nabla \phi^{(1)} + D_2 f_2 H \nabla \phi^{(2)} + D_{3a} f_{3a} H \nabla \phi^{(3)} + D_{3b} f_{3b} H \nabla \phi^{(3)} ,
\]

(A4)

with \( f_i = \frac{d\ln D_i}{d\ln a} \) (particular expressions for \( f_{3a} \) and \( f_{3b} \) can be found in Bouchet et al. 1995). As we can see from equation (A4) one can construct all components from the linear potential \( \phi^{(1)} \).

We consider here two models for the linear potential. The first model relies on the local lognormal estimate (see Section 2.2) from which the scaled velocity divergence can be calculated in the following way to third-order LPT:

\[
\theta = D_1 \delta^{(1)} - \frac{f_2}{f_1} D_2 \mu^{(2)}(\phi^{(1)}) - \frac{f_{3a}}{f_1} D_{3a} \mu^{(3)}(\phi^{(1)}) - \frac{f_{3b}}{f_1} D_{3b} \mu^{(2)}(\phi^{(1)}, \phi^{(2)}).
\]

(A5)

The second model yields a non-local estimate of \( \phi^{(1)} \) to third order given by

\[
\delta_L = \delta - \delta^{NL},
\]

(A6)

with \( \delta^{NL} = -D_2 \mu^{(2)}(\phi^{(1)}) - D_{3a} \mu^{(3)}(\phi^{(1)}) - D_{3b} \mu^{(2)}(\phi^{(1)}, \phi^{(2)}) + \mu^{(2)}(\Theta) + \mu^{(3)}(\Theta) \). Note that the potential \( \Theta \) is also different and is determined by equation (A1) recalling the relation to the displacement field \( \Psi = -\nabla \Theta \).

Using the latter expression one can write the \( \theta-\delta \) relation to third order in LPT as

\[
\theta = \delta - \left( \frac{f_2}{f_1} - 1 \right) D_2 \mu^{(2)}(\phi^{(1)}) - \left( \frac{f_{3a}}{f_1} - 1 \right) D_{3a} \mu^{(3)}(\phi^{(1)}) - \left( \frac{f_{3b}}{f_1} - 1 \right) D_{3b} \mu^{(2)}(\phi^{(1)}, \phi^{(2)}) - \mu^{(2)}(\Theta) - \mu^{(3)}(\Theta).
\]

(A7)

Note that equation (A6) should be solved iteratively as we do in the 2LPT case (see Section 2.2 and Kitaura & Angulo 2012). Nevertheless to find a fast solution we plug-in the lognormal model for \( \phi^{(1)} \) into equation (A5) yielding a non-local estimate of the linear field in one iteration. To minimize the deviation between LPT and the full non-linear evolution we have additionally smoothed the density \( \delta \) in equation (A7) with a Gaussian kernel of \( 2 \, h^{-1} \text{Mpc} \) radius.

We do not find an improvement in the determination of the velocity divergence with respect to 2LPT including both curl-free terms using any of the estimates of the linear component. This could be due to an inaccurate determination of the interaction term \( \phi^{(3)} \), since uncertainties in the linear component \( \phi^{(1)} \) propagate more dramatically than in terms which depend only on the initial conditions \( \phi^{(2)}, \phi^{(3)} \).

© 2012 The Authors, MNRAS 425, 2422–2435

Monthly Notices of the Royal Astronomical Society © 2012 RAS
One may neglect the interaction term $\delta^{(3)}_b$ and consider only the cubic term $\delta^{(3)}_a$ to third order. Such a truncated 3LPT model includes the main body of the perturbation sequence with the rest of the sequence being made up of interaction terms (Buchert 1994). Our calculations show in this case better results than 2LPT for low values of the velocity divergence (see Fig. A1). However, for large values of $\nabla \cdot v$ we obtain larger dispersions which could be also due to numerical errors in the estimate of the linear density component $\delta^L$. The errors in the velocity estimation are only moderately reduced with respect to the 2LPT case (see Section 3.2).

This paper has been typeset from a \TeX/\LaTeX file prepared by the author.