# BAO reconstruction: a swift numerical action method for massive spectroscopic surveys

E. Sarpa [1,2]★ C. Schimd,[1] E. Branchini[3,4,5] and S. Matarrese[2,6,7,8]

[1]*Aix Marseille Univ, CNRS, CNES, LAM, 38, Rue Joliot-Curie, F-13388 Marseille Cedex 13, France*
[2]*Dipartimento di Fisica e Astronomia 'Galileo Galilei', Università degli studi di Padova, Via F. Marzolo 8, I-35131 Padova, Italy*
[3]*Dipartimento di Matematica e Fisica, Università degli studi Roma Tre, Via della Vasca Navale 84, I-00146 Roma, Italy*
[4]*INFN - Sezione di Roma Tre, via della Vasca Navale 84, I-00146 Roma, Italy*
[5]*INAF - Osservatorio Astronomico di Roma, via Frascati 33, I-00040 Monte Porzio Catone (RM), Italy*
[6]*INFN, Sezione di Padova, via F. Marzolo 8, I-35131 Padova, Italy*
[7]*INAF-Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122 Padova, Italy*
[8]*Gran Sasso Science Institute, Viale F. Crispi 7, I-67100 L'Aquila, Italy*

## ABSTRACT

A new fully non-linear reconstruction algorithm for the accurate recovery of the baryonic acoustic oscillations (BAO) scale in two-point correlation functions is proposed, based on the least action principle and extending the Fast Action Minimisation method by Nusser & Branchini ([2000]). Especially designed for massive spectroscopic surveys, it is tested on dark matter halo catalogues extracted from the DEUS-FUR Lambda cold dark matter simulation (Reverdy et al. [2015]) to trace the trajectories of up to ∼207 000 haloes backward in time, well beyond the first-order Lagrangian approximation. The new algorithm successfully recovers the BAO feature in real and redshift space in both the monopole and the anisotropic two-point correlation function, also for anomalous samples showing misplaced or absent signature of BAO. In redshift space, the non-linear displacement parameter $\Sigma_{NL}$ is reduced from $11.8 \pm 0.3\,h^{-1}\,\mathrm{Mpc}$ at redshift $z = 0$ to $4.0 \pm 0.5\,h^{-1}\,\mathrm{Mpc}$ at $z \simeq 37$ after reconstruction. A comparison with the first-order Lagrangian reconstruction is presented, showing that these techniques outperform the linear approximation in recovering an unbiased measurement of the acoustic scale.

**Key words:** methods: numerical – cosmological parameters – large-scale structure of Universe.

## 1 INTRODUCTION

Before recombination and on small scales, the acoustic oscillations of the primordial baryon–photon plasma propagate at relativistic speed driven by photon pressure. On large scales they are standing waves with the fundamental mode set by the sound horizon $r_s$ (Peebles & Yu [1970]; Sunyaev & Zeldovich [1970]) and overtones damped on scales $\lesssim 8\,h^{-1}\,\mathrm{Mpc}$ because of the non-perfect coupling between baryons and photons (Silk [1968]). At recombination and afterwards, while photons freely stream, the residual baryonic fluctuations play as additional seeds of clustering at the characteristic scale $r_s^* \simeq 150$ Mpc, progressively driving an excess of clustering of the dominant, collisionless dark matter (Hu & Sugiyama [1996]; Eisenstein et al. [1998]; Eisenstein & Hu [1998]). On these scales the gravitational clustering is well described by the linear approximation until today. Moreover, $r_s^*$ depends on

the primordial baryon-to-photon ratio and on the matter density, extremely well constrained by the cosmic microwave background observations (Planck Collaboration et al. [2018]). The baryonic acoustic oscillations (BAO) scale $r_s^*$ therefore provides a powerful and robust standard ruler to strongly constrain the expansion rate of the Universe and dark energy (Blake & Glazebrook [2003]; Hu & Haiman [2003]; Seo & Eisenstein [2003]; White [2005]), though requiring large volumes to be convincingly measured (Cole et al. [2005]; Eisenstein et al. [2005]).

At low redshift the non-linear evolution of structures, the bias of tracers, and their peculiar velocities determine a mild degradation of the acoustic signature in the two-point correlation functions (Eisenstein, Seo & White [2007a]; Mehta et al. [2011]) that must be reduced in order to enhance the signal-to-noise ratio and achieve sub- per cent precision cosmology goals. The resulting broadening and shift of the acoustic peak in the correlation function, or equivalently the damping and phase shift of the higher harmonics in the power spectrum, can indeed be alleviated by means of so-called reconstruction techniques designed to recover the initial, linear

★ E-mail: elena.sarpa@lam.fr

density field using a Lagrangian perturbative scheme (Eisenstein et al. 2007b; Padmanabhan, White & Cohn 2009), which provides a good reconstruction also when applied to biased mass tracers such as galaxies (Noh, White & Padmanabhan 2009). The most widely used technique (Padmanabhan et al. 2012) is based on the displacement of the tracers backward in time using the Zel'dovich approximation calculated from the local density field, in which a linear correction is applied to remove redshift-space distortion (RSD) that actually occurs on both small and large scales (Kaiser 1987). Since its first application on the Sloan Digital Sky Survey (SDSS) DR7 LRG sample at redshift $z \simeq 0.35$ (Padmanabhan et al. 2012), this reconstruction technique has been routinely adopted for BAO studies based on the data releases DR9 to DR12 using both the LOWZ and CMASS galaxy samples across the redshift range $z = 0.34$–$0.7$ (Anderson et al. 2012; Anderson et al. 2014a, b; Ross et al. 2014; Burden et al. 2014; Tojeiro et al. 2014; Cuesta et al. 2016; Gil-Marín et al. 2016) and with the WiggleZ galaxies up to redshift $z \simeq 1$ (Kazin et al. 2014; Beutler et al. 2016), resulting in a substantial improvement on the measurement of the BAO scale.

The main limitation of this technique is the approximate treatment of the non-linear dynamics that may degrade the quality of the reconstruction. More sophisticated techniques are required in next-generation redshift surveys, such as those realized by PFS-SuMIRe (Takada et al. 2014), eBOSS (Dawson et al. 2016), 4MOST (de Jong et al. 2016), DESI (Martini et al. 2018), Euclid (Laureijs et al. 2011), or WFIRST (Doré et al. 2018), which will sample a large range of galaxy overdensities within cosmological volumes. Indeed, in large overdensity regions (i.e. small scales) the linear approximation for the RSD correction adopted so far fails, the Zel'dovich approximation not being an exact solution (Nusser et al. 1991; Burden, Percival & Howlett 2015). Moreover, although optimal in recovering the BAO signature in the monopole of the correlation function or in the power spectrum also from large realistic samples (Keselman & Nusser 2017), the efficiency of the Lagrangian perturbative schemes to recover the BAO feature from higher order multipoles and from the full two-dimensional correlation functions, and simultaneously to account for the small scales for biased tracers, is less evident (White 2014).

Several reconstruction techniques that adopt different approaches alternative to the standard one have been proposed. One group is represented by methods that still rely on (higher order) Lagrangian perturbation theory. They implement iterative schemes, such as PIZA (Croft & Gaztanaga 1997) or ZTRACE (Monaco & Efstathiou 1999), include the gravitational tidal-field tensor and some linearization scheme (Kitaura et al. 2012; Kitaura & Angulo 2012), apply a local transform to the density field (Falck et al. 2012; McCullagh et al. 2013), Wiener filtering to the large-scale forward displacement (Tassev & Zaldarriaga 2012), or an iterative fast Fourier transform method (Burden et al. 2015). The second group is that of reconstruction methods based on Bayesian theory, which have been tested on galaxy mock catalogues (Kitaura & Enßlin 2008; Jasche & Wandelt 2013; Wang et al. 2013). Finally, there are the fully non-linear techniques, in which simplifying assumptions are not made on the dynamical state of the system but rather of the orbits of the objects. One example is based on the optimal Monge–Ampère–Kantorovic (MAK) mass transportation problem (Frisch et al. 2002; Brenier et al. 2003; Mohayaee et al. 2003). Successfully applied to reconstruct the peculiar velocity field of the Two Micron All-Sky Redshift Survey catalogue (Lavaux et al. 2010), the MAK technique is however computationally expensive and therefore very likely limited to reconstructions within few hundreds Mpc (Mohayaee

et al. 2006). Recently, an efficient non-linear reconstruction method has been proposed by Keselman & Nusser (2017), based on a forward iterative scheme that uses standard *N*-body techniques for exactly solving the equation of motions.

In this paper, we shall focus on the non-linear method based on the cosmological least action principle (LAP; Peebles 1989), which yield the full trajectory of tracers along with their velocities as a byproduct. Further developed by Peebles (1994, 1995) and Shaya, Peebles & Tully (1995) to trace the dynamics of galaxies in the Local Universe, the LAP has been investigated against cold dark matter *N*-body simulations to estimate the collapsing history of haloes by Branchini & Carlberg (1994) and Dunn & Laflamme (1995). An efficient version dubbed Fast Action Minimization (FAM) method has been then implemented by Nusser & Branchini (2000, hereafter NB00) to reconstruct the dynamics of galaxies in the nearby Universe, finally accounting for RSD (Branchini, Eldar & Nusser 2002) and tidal-field effects (Romano-Díaz, Branchini & van de Weygaert 2005).

In this paper, we present an extended version of the original FAM algorithm, dubbed eFAM, coded in C++ language and designed for massive spectroscopic surveys such as Euclid, able to reconstruct the trajectories of $O(10^6)$ objects in generic background cosmologies and that we specifically apply to BAO reconstructions. Since the method provides multiple solutions for the orbits of the particles in virialized regions, the impact of non-linearities is minimized by reconstructing the trajectories of only collapsed haloes considered as point-like tracers, hence neglecting their internal structure altogether.

The plan of the paper is as follows: in the next section, we present the eFAM algorithm. Section 3 describes the analysis of large dark matter halo catalogues extracted from DEUS-FUR Lambda cold dark matter (ΛCDM) simulations and the modelling of the RSD. The analysis of reconstructed haloes' orbits is performed and discussed in Section 4, focusing on the monopole moment of the two-point correlation function in real- and redshift space and the anisotropic correlation function in redshift space. The efficiency of the eFAM algorithm to recover the BAO scale and a comparison with a reconstruction based on the simple Zel'dovich approximation is also discussed. Section 5 is dedicated to the conclusions.

## 2 THE RECONSTRUCTION METHOD

### 2.1 The extended Fast Action Minimisation (eFAM) method

The original FAM algorithm (NB00) has been developed to recover the past orbits of point-like particles both in real- and redshift space in an Einstein-De Sitter universe. Here, FAM is extended to a generic cosmology defined by the Hubble parameter $H$ and the linear growth factor $D$ or the linear growth rate $f$, as functions of the scale factor $a$. The action of a set of $N$ collisionless equal-mass point-like particles with comoving coordinates $\{\mathbf{x}_i\}_{i=1,...,N}$ sampling in an unbiased way a volume $V$ and interacting only by gravity is, in the weak-field limit,

$$S = \sum_{i=1}^{N} \int_0^{D_{\mathrm{obs}}} \mathrm{d}D \left[ f E D a^2 \frac{1}{2} \left( \frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}D} \right)^2 \right.$$
$$\left. + \frac{3\Omega_{\mathrm{m0}}}{8\pi f E D a} \left( \frac{1}{\bar{n}_{\mathrm{obs}} a_{\mathrm{obs}}^3} \frac{1}{2} \sum_{j \neq i, j=1}^{N} \frac{1}{|\mathbf{x}_i - \mathbf{x}_j|} + \frac{2}{3} \pi \mathbf{x}_i^2 \right) \right], \quad (1)$$

where $D$ is used as time variable and $E = H/H_0$ the dimensionless Hubble parameter. Here the subscript '0' denotes quantities at

present time and 'obs' at redshift $z = z_{obs}$. The mean number density of objects $\bar{n}_{obs} = N/V$ determines the mean mass density in the volume, $\bar{\rho}$, and its mass density parameter $\Omega_{m0}$. The trajectories of objects are the solutions of the equations of motion deduced from a stationary action, $\delta S = 0$, subject to mixed boundary conditions as in FAM.

Differently from the original FAM algorithm that used a tree-code to calculate the gravitational force-field, eFAM uses FALCON (Dehnen 2002), a very efficient Poisson solver that optimally combines a tree-code and the fast multipole method (FMM). The FMM implements an improved multipole-acceptance criterion for the splitting or execution of the cell–cell interaction, and a symmetric calculation of the cell–cell interactions that conserves the total momentum. The computational complexity is finally reduced to $O(N)$. This is the major improvement of eFAM over FAM since it allows its application to larger data sets than its predecessor and, more specifically, capable of targeting the BAO reconstruction goal.

## 2.2 Orbits parametrization in generic cosmology

As in NB00, the trajectories $\{\mathbf{x}_i(D)\}_i$ are described by a linear combination of $M$ time-dependent basis functions $\{q_n(D)\}_n$ with unknown coefficients $\mathbf{C}_{i,n}$, viz.

$$\mathbf{x}_i(D) = \mathbf{x}_{i,obs} + \sum_{n=0}^{M} \mathbf{C}_{i,n} q_n(D). \tag{2}$$

The $q_n(D)$ are polynomials defined such that their derivatives $p_n(D) \equiv dq_n(D)/dD$ are the Jacobi polynomials satisfying the orthogonality condition

$$\int_0^{D_{obs}} dD\, w(D) p_n p_m = \delta_{nm}^K A_n, \tag{3}$$

with $\delta_{nm}^K$ the Kronecker delta and $A_n$ determined by the recurrence relations. Different from NB00, the weight function $w(D) \equiv fEDa^2$ can fit a generic background cosmology reasonably close to $\Lambda$CDM (see Appendix A); a wrong guess does not bias the trajectories, which are almost independent of the cosmological parameters (Nusser & Colberg 1998).

For every term $n$, the mixed boundary conditions deduced from $\delta \mathbf{x}_i(D_{obs}) = 0$ and $\lim_{D \to 0} \dot{D}\theta_i(D) = 0$ are

$$q_n(D_{obs}) = 0, \quad \lim_{D \to 0} af HD p_n(D) = 0, \tag{4}$$

in which the dependence on $D$ is omitted for clarity for all but the basis functions and their derivatives. Denoting $\theta_i \equiv d\mathbf{x}_i/dD = \mathbf{v}_i/fDH$ the rescaled peculiar velocity of the $i$-th particle and

$$\mathbf{g}_i \equiv -\frac{1}{\bar{n}_{obs} a_{obs}^3} \frac{1}{2} \sum_{j \neq i, j=1}^{N} \frac{\mathbf{x}_i - \mathbf{x}_j}{|\mathbf{x}_i - \mathbf{x}_j|^3} + \frac{4}{3}\mathbf{x}_i \tag{5}$$

its peculiar acceleration, the stationary variations of the action with respect to $\mathbf{C}_{i,n}$ give

$$0 = \frac{\partial S}{\partial \mathbf{C}_{i,n}} = \int_0^{D_{obs}} dD\, w\boldsymbol{\theta}_i p_n + \int_0^{D_{obs}} dD\, \frac{3\Omega_{m,0}}{8\pi f EDa} \mathbf{g}_i q_n$$

$$= [w\boldsymbol{\theta}_i q_n]_0^{D_{obs}} - \int_0^{D_{obs}} dD\, \left[\frac{d(w\boldsymbol{\theta}_i)}{dD} - \frac{3\Omega_{m0}}{8\pi f EDa}\mathbf{g}_i\right] q_n. \tag{6}$$

With the boundary conditions (equation 4), these $N \times M$ equations correspond to the Euler-Lagrange equations obtained from $\delta S(\mathbf{x}_1, \ldots, \mathbf{x}_N, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N) = 0$. This assures that the search for the stationary point of the action with respect to the coefficients $\mathbf{C}_{i,n}$ is equivalent to the one with respect to the whole trajectories.

## 2.3 Redshift space

In redshift space we introduce the comoving redshift coordinates of the $i$-th object as

$$\mathbf{s}_{i,obs} = \frac{H_0 a_0}{c} \mathbf{x}_{i,obs} + \frac{a_0(fDH)_{obs}}{c} \boldsymbol{\theta}_{i,obs}^{\parallel}, \tag{7}$$

in which $\boldsymbol{\theta}_i^{\parallel}$ denotes the component of the peculiar velocity along the line of sight. The additional term proportional to $\boldsymbol{\theta}_{i,obs}^{\parallel}$, absent in real-space, breaks the isotropy of the $\mathbf{s}_{i,obs}$ introducing a preferential direction along the line of sight; a Cartesian decomposition of the orbits coefficients $\mathbf{C}_{i,n}$ is therefore not convenient anymore. Instead, the coefficients $\mathbf{C}_{i,n}$ can be split into two components perpendicular and parallel to the line of sight. In the extended version of FAM we implemented the approach illustrated in Schmoldt & Saha (1998) by assigning to each object a Cartesian coordinate system with one axis aligned to the line of sight and the observer's position as the common origin of the galaxies' frames. In this way, the correction for the RSD is confined to one single axis, parallel to the radial velocity. Note that although the objects move, their coordinate frames do not.

We adopted the orbits parametrization introduced in NB00 [see their equation (20) and Appendix B] with

$$Q_n(D_{obs}) = -(fDE)_{obs} p_n(D_{obs}), \quad n = 1, \ldots, M. \tag{8}$$

To preserve the time-averaged equations of motion, a kinetic energy term corresponding to a degree of freedom parallel to the line of sight is added to the action of the system; the resulting action in redshift space to be minimized is

$$S = S + \frac{1}{2} (wfDE)_{obs} \left(\boldsymbol{\theta}_{i,obs}^{\parallel}\right)^2. \tag{9}$$

## 2.4 Minimization procedure and first guess

The very hard minimization problem in $3N \times M$ dimensions, which can be as large as $10^7$ for several millions of objects and $M \simeq 10$ polynomials, is carried out using the same non-linear conjugate gradient method with the Polak-Ribière formula as in NB00. However, this method is locally optimal while the action $S$ can have many minima corresponding to different solutions of the time-averaged equations of motion. Indeed, as pointed out by Peebles (1989) and Giavalisco et al. (1993) the solution of this mixed boundary-value problem is not unique, because the boundary conditions prescribe the time dependence of the velocities near the initial time $D = 0$ but do not specify their amplitude. Since we are here interested in the large-scale dynamics, the minimum should correspond to orbits that do not significantly deviate from the Zel'dovich approximation. The first guess of the iterative minimization is then chosen as the one prescribed by the Zel'dovich approximation: the peculiar gravitational acceleration (equation 5) linearly scales with the growth factor, $\mathbf{g}_i(D) = \frac{D/D_{obs}}{a/a_{obs}} \mathbf{g}_i(D_{obs})$, and all but the zeroth-order coefficient are vanishing, i.e. $\mathbf{C}_{i,n} = 0$ for all $n > 0$, corresponding to straight-line orbits.

For both the real-space ($r$-space) and the perpendicular components in the redshift space, the orthogonality condition (equation 3) yields

$$\mathbf{C}_{i,n}^{r-space} = \mathbf{C}_{i,n}^{\perp} = -\frac{1}{A_n} \frac{a_{obs}}{D_{obs}} \mathbf{g}_i(D_{obs}) \int_0^{D_{obs}} dD\, \frac{3\Omega_{m0}}{8\pi f Ea^2} q_n, \tag{10}$$

while for the parallel component in redshift space

$$\mathbf{C}_{i,n}^{\parallel} = \frac{-a_{obs}\mathbf{g}_i^{\parallel}(D_{obs})}{[A_n + (wfDEp_n^2)_{obs}]D_{obs}} \int_0^{D_{obs}} dD\, \frac{3\Omega_{m0}}{8\pi f Ea^2} q_n. \tag{11}$$

## 3 BAO RECONSTRUCTION

### 3.1 DEUS Full Universe Run haloes

The eFAM method is formulated to reconstruct the complete trajectory of objects potentially well into the non-linear regime. Accordingly, its accuracy needs then to be tested using simulated catalogues produced by fully non-linear *N*-body experiments rather than those obtained with approximated schemes like PTHALOS (Scoccimarro & Sheth 2002; Manera et al. 2013), COLA (Tassev, Zaldarriaga & Eisenstein 2013; Howlett, Manera & Percival 2015), PINOCCHIO (Monaco, Theuns & Taffoni 2002; Munari et al. 2017), PATCHY (Kitaura, Yepes & Prada 2014), EZMOCKS (Chuang et al. 2015), or HALOGEN (Avila et al. 2015), which are less accurate on small scales. Moreover, all these algorithms add random peculiar velocities to galaxies, typically drawn from a Maxwellian distribution scaled on the underlying matter density; this stochastic recipe is not compatible with the deterministic nature of the eFAM algorithm.

Aiming at assessing the quality of the reconstruction on large scales in the presence of significant non-linear effects, we have considered various sets of dark matter haloes at redshift $z = 0$, extracted from the Dark Energy Universe Simulation – Full Universe Run (DEUS-FUR; Reverdy et al. 2015), a flat $\Lambda$CDM simulation set on the WMAP-7 best-fitting cosmology (Spergel et al. 2007)[1] employing $8192^3$ dark matter particles with formal mass and spatial resolution, respectively, of $1.2 \times 10^{12}\,h^{-1}\,\mathrm{M_\odot}$ and $40\,h^{-1}\,\mathrm{kpc}$, in a cubic volume of $(21\,h^{-1}\,\mathrm{Gpc})^3$. The friend-of-friend (FoF) halo catalogue includes only haloes with more than 100 particles, amounting to more than 144 millions haloes at $z = 0$. We stress that this is a demanding but somewhat unrealistic test, since next generation surveys will probe higher redshifts where non-linear effects are less prominent.

The large volume of DEUS-FUR, which encompasses the Hubble horizon thus enabling cosmic-variance limited predictions at the BAO scale (Rasera et al. 2014), allows us to extract 512 cubic sub-volumes of length $2\,h^{-1}\,\mathrm{Gpc}$ separated by a buffer region of $500\,h^{-1}\,\mathrm{Mpc}$, a distance above the scale of homogeneity (see e.g. Ntelis et al. 2017). The sub-volumes can therefore be considered as effectively independent, allowing for a Monte Carlo estimation of the covariance (Norberg et al. 2009).

The peculiar velocities of haloes are not supplied with the DEUS-FUR FoF catalogue. To emulate the RSD and perform the reconstruction in redshift-space, eFAM is run first time on every sub-volume to assign the peculiar velocities to haloes. Then for every sub-volume a reference frame is fixed in its centre of mass, the redshift coordinates of haloes are computed using equation (7), and the action (9) is minimized. The accuracy of this procedure is discussed and quantitatively assessed in Section 4.4.

### 3.2 Gravitational tidal field

In its basic formulation the FAM method does not assume any external gravitational field, treating the sampled density field as an isolated system. This assumption does not apply in a cosmological context, however the impact of the external density field can be minimized by choosing a spherical geometry for the sample. Here the reconstruction is applied to spherical domains $\mathcal{D}$ of radius $990\,h^{-1}\,\mathrm{Mpc}$ extracted from each sub-volume, each containing about 56 000 haloes on average.

A direct computation of the external gravitational field $\Phi_{\mathrm{tidal}}(\mathbf{x})$ affecting the dynamics inside the sub-region $\mathcal{D}_{\mathrm{int}}$ can be performed by extending the reconstruction to a larger domain $\mathcal{D}$ embedding $\mathcal{D}_{\mathrm{int}}$, as done by Shaya et al. (1995). To assess the impact of $\Phi_{\mathrm{tidal}}(\mathbf{x})$, we applied eFAM to haloes in several concentric spherical domains with radius ranging from 100 to $300\,h^{-1}\,\mathrm{Mpc}$, extracted from DEUS-648, a smaller simulation for which the peculiar velocities are provided. We then performed a point-wise comparison between the Cartesian components of reconstructed and real velocities of the haloes that are in the common domain $\mathcal{D}_{\mathrm{int}}$ of radius $100\,h^{-1}\,\mathrm{Mpc}$, and assessed the impact of $\Phi_{\mathrm{tidal}}(\mathbf{x})$ by the offset $q$ of the linear regression $v_{\mathrm{FoF}} = mv_{\mathrm{eFAM}} + q$. It turns out that the reconstruction within $\mathcal{D}_{\mathrm{int}}$ is improved by the inclusion of a buffer region consisting of a spherical shell of at least $200\,h^{-1}\,\mathrm{Mpc}$, reducing the offset from $q \sim 100\,\mathrm{km\,s^{-1}}$ when no external field is considered to $q \sim 10\,\mathrm{km\,s^{-1}}$, regardless of the Cartesian component. Therefore, in the DEUS-FUR-based BAO reconstruction, we shall only analyse the regions within $700\,h^{-1}\,\mathrm{Mpc}$, each containing about 23 000 haloes on average, and ignore the outskirts that extend to $990\,h^{-1}\,\mathrm{Mpc}$.

## 4 FITTING THE ACOUSTIC FEATURE

### 4.1 Measuring, modelling, and fitting the correlation function

For each sub-sphere $\mathcal{D}_{\mathrm{int}}$ the monopole of the two-point correlation function $\xi(r)$ is computed in the separation range 30–$200\,h^{-1}\,\mathrm{Mpc}$ with linear binning of $10\,h^{-1}\,\mathrm{Mpc}$ using the minimum variance Landy & Szalay (1993) estimator, with 50 times random objects homogeneously distributed within $\mathcal{D}_{\mathrm{int}}$. The measurement is repeated for all sources at their pre- and post-reconstructed positions at 14 different redshifts, starting from $z = 0$ and up to the maximum redshift $z_{\mathrm{max}}$ allowed by reconstruction, defined as the redshift for which $\Sigma_{\mathrm{NL}}$ attains the minimum positive value.

The results are fitted using the Xu et al. (2012) model

$$\xi(r; z) = B^2 \xi_{\mathrm{m}}(\alpha r; z) + A(r), \tag{12}$$

where $B$ is a multiplicative constant bias and $\xi_{\mathrm{m}}(r)$ is the Fourier transform of

$$P(k) = [P_{\mathrm{lin}}(k) - P_{\mathrm{smooth}}(k)]\,e^{-k^2 \Sigma_{\mathrm{NL}}^2/2} + P_{\mathrm{smooth}}(k), \tag{13}$$

with $P_{\mathrm{lin}}(k)$ the actual linear power spectrum and $P_{\mathrm{smooth}}(k)$ its de-wiggled limit, both computed by CAMB (Lewis & Bridle 2002) using the same cosmological model adopted in the DEUS-FUR simulations. The broad-band term

$$A(r) = A_0 + \frac{A_1}{r} + \frac{A_2}{r^2} \tag{14}$$

can be interpreted as an effective description of mode-coupling (e.g. Crocce & Scoccimarro 2008) not affecting the BAO scale but biasing its measurement if not accounted for properly. This additive term can also help to alleviate the effects of assuming a wrong cosmological model. The broadening and shift of the BAO feature due to the non-linear growth of structures are described by the parameter $\Sigma_{\mathrm{NL}}$, accounting for the Lagrangian displacements, and by the scale dilation parameter $\alpha$. The model has therefore six free parameters, $(B^2, \alpha, \Sigma_{\mathrm{NL}}, A_0, A_1, A_2)$, and fixed cosmological parameters.

The data are fitted using a Markov chain Monte Carlo technique with Gaussian likelihood and flat priors. The covariance matrix is calculated from the $N_{\mathrm{mocks}} = 512$ mocks as

$$\mathbf{C}_{i,j} = \frac{1}{N_{\mathrm{mocks}} - 1} \sum_{n=1}^{N_{\mathrm{mocks}}} [\xi_n(r_i) - \bar{\xi}(r_i)][\xi_n(r_j) - \bar{\xi}(r_j)] \tag{15}$$

---

[1]Cosmological parameters: $\Omega_{\mathrm{b}} = 0.04356$, $\Omega_{\mathrm{m}} = 0.2573$, $\Omega_{\mathrm{r}} = 0.000049$, $h = 0.72$, $n_{\mathrm{s}} = 0.9630$, $\sigma_8 = 0.8010$.
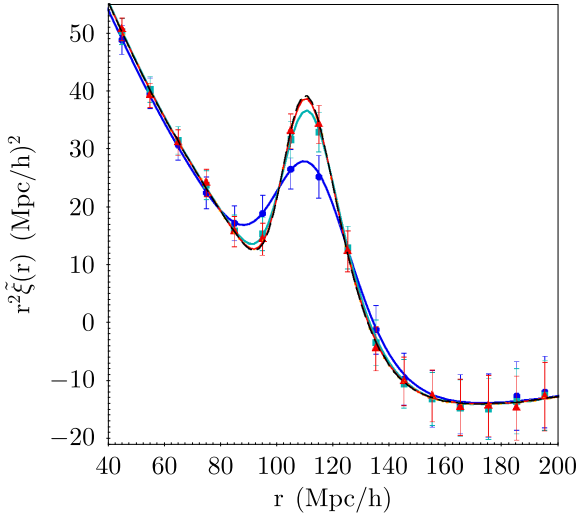
**Figure 1.** Rescaled two-point correlation function in real-space of DEUS-FUR ΛCDM dark-matter haloes before and after reconstruction by eFAM ($M = 10$), averaged over the 512 mocks. Before reconstruction at redshift $z = 0$ (circles/blue line) the acoustic feature is broadened by non-linear evolution and peculiar velocities. After reconstruction at $z = 2.7$ and $z = 6.5$ (squares/green and triangles/red lines, respectively) the measured correlation function gets progressively closer to the linear prediction (black-dashed), indicating the quality of the eFAM algorithm. The error bars from the diagonal of the covariance matrix.

where $\bar{\xi}(r) = \sum_{n=1}^{N_{mocks}} \xi(r)/(N_{mocks} - 1)$ is the mean two-point correlation function.

Measurements, modelling, and fitting are performed using the routines implemented in the COSMOBOLOGNALIB (Marulli, Veropalumbo & Moresco 2016).

### 4.2 Monopole in real space

An ideal reconstruction pushed at early time, before the non-linear clustering became effective, would yield $\Sigma_{NL} \to 0$ and $\alpha \to 1$. As shown in Fig. 1, using the eFAM algorithm with $M = 10$ basis functions (hereafter quoted as eFAM$_{10}$) the template model $\xi_m(r)$ based on the linear theory is closely approached at redshift $z = 2.7$ and almost fully restored at $z = 6.5$ (to ease the comparison between the correlation functions at different redshifts and enhance the acoustic feature, the rescaled monopole $r^2\bar{\xi}(r; z) = r^2 [\xi(r; z) - A(r)] [B^2 D(z)^2]^{-1}$ is plotted). The error

bars are the rms-variance obtained from the diagonal elements of the covariance matrix (15). Remarkably enough, the errors on the monopole calculated pre-reconstruction do not increase after reconstruction. More interestingly, the correlation matrix becomes definitely more diagonal going towards higher redshift; see Fig. 2. This indicates that the reconstruction de-correlates the signal in all bins.

The fact that the linear regime is almost fully restored already at $z = 6.5$ rather than at higher redshifts, as expected, reflects that eFAM overestimates the amplitude of peculiar velocities of objects in high-density environments (small scales), where linear theory fails, but predicts their correct direction. As a result, the back in time displacement of these objects is overestimated, the density contrast is more efficiently reduced and, consequently, linear theory restored at lower redshifts than expected. Such overcorrection does not affect the quality of the BAO reconstruction as long as it does not lead to an unphysical compression of the BAO peak, namely below the Silk scale.

To estimate the impact of the reconstruction on the dilation $\alpha$ and its error, we performed the same analysis as in Padmanabhan et al. (2012): the two-point correlation function is fitted for every mock, using flat priors for all the parameters but $\Sigma_{NL}$, for which a Gaussian prior centred on the best fit obtained for the average $\bar{\xi}(r)$ and with the same variance. The scatterplot in the left-hand panel of Fig. 4 compares the values of $\alpha$ pre- and post-reconstruction for all the 512 halo catalogues. The non-linear eFAM method improves the measurement of the BAO scale reducing the standard deviation of the probability distribution function of $\alpha$, without introducing any statistical bias. The improvement of the precision on $\alpha$ (right-hand panel) is significant, the eFAM algorithm yielding $\sigma_{\alpha, rec} < \sigma_{\alpha, unrec}$ for the 69 per cent of mocks.

The eFAM reconstruction is superior to the standard Zel'dovich one. To quantify the improvement we repeated the reconstructions using eFAM with $M = 1$ (i.e. eFAM$_1$), that is forcing straight orbits. This is not strictly Zel'dovich approximation since velocities change along the orbit, but is a good approximation to it. As shown in Fig. 3, the averaged correlation function has a very large variance, with the acoustic peak systematically shifted towards larger values, $\alpha = 0.98 \pm 0.01$. The poor reliability of the results achieved by eFAM$_1$ is mainly due to the low value of its maximum allowed redshift, $z = 3.7$, above which the best-fitting value of $\Sigma_{NL}$ becomes unphysical; see table 1. This also explains why the results obtained at $z = 2.7$ are better than those at $z = 6.5$, which is the maximum allowed redshift for eFAM$_{10}$. Instead, owing to the larger number of degrees of freedom, at the same redshift the fully non-linear eFAM$_{10}$
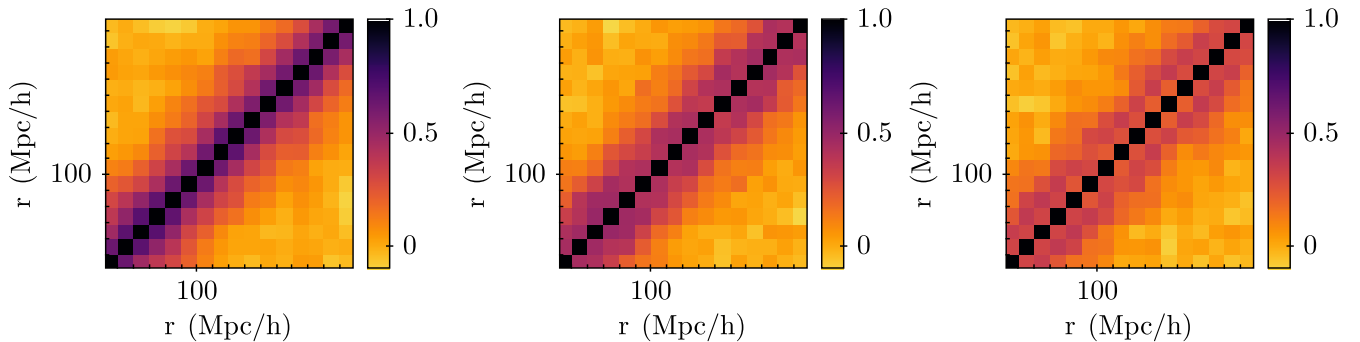


**Figure 2.** Correlation matrix of the two-point correlation function (monopole, real space) around the BAO scale. Left: from non-reconstructed halo catalogues. Middle and right: from reconstructed haloes, using eFAM$_{10}$, at $z = 2.7$ and $z = 6.5$, respectively. The correlation matrix becomes more diagonal at higher redshifts, proving that the correlation functions of the different mocks tend to converge to the mean form when reconstructed.
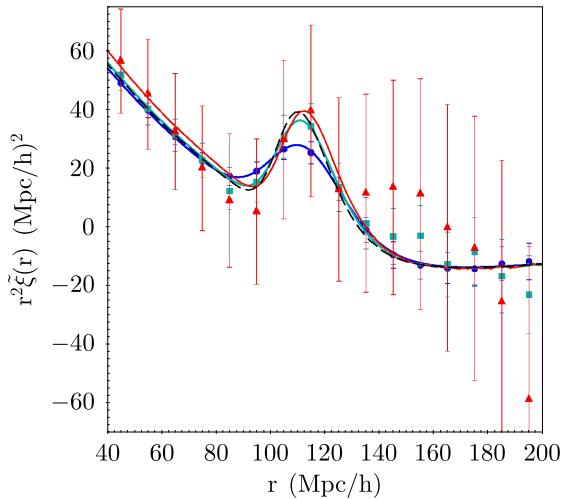
**Figure 3.** Similar to Fig. 1 (real-space), but with reconstruction by the Zel'dovich approximation as provided by the eFAM with $M = 1$. The very large errors indicate the non-reliability of this method, which moreover returns a biased estimation of the acoustic scale towards larger values.

**Table 1.** Fit results to the average correlation function before and after reconstruction in real-space by eFAM$_{10}$ and eFAM$_1$. The parameters not shown are marginalized over. A negative (unphysical) best-fitting value of $\Sigma_{NL}$ indicates that $z_{max}$ has been attained, values for $z > z_{max}$ (marked by $*$) are shown just for comparison.

| Method | Redshift | $\alpha$ | $\Sigma_{NL}$ ( $h^{-1}$ Mpc) |
|---|---|---|---|
| pre-recon | $z = 0$ | $1.007 \pm 0.002$ | $9.0 \pm 0.2$ |
| eFAM$_{10}$ | $z = 2.7$ | $0.999 \pm 0.001$ | $3.5 \pm 0.4$ |
| | $z = 3.7$ | $0.998 \pm 0.001$ | $2.3 \pm 0.7$ |
| | $z = 4.3$ | $0.999 \pm 0.001$ | $1.6 \pm 0.7$ |
| | $z = 6.5$ | $1.000 \pm 0.001$ | $1.2 \pm 0.7$ |
| eFAM$_1$ | $z = 2.7$ | $0.996 \pm 0.003$ | $3.8 \pm 1.2$ |
| | $z = 3.7$ | $0.997 \pm 0.003$ | $2.0 \pm 1.4$ |
| | $z = 4.3*$ | $0.927 \pm 0.018$ | $-3.4 \pm 4.3$ |
| | $z = 6.5*$ | $0.980 \pm 0.010$ | $2.0 \pm 2.1$ |

method ensures a non-biased measurement of the acoustic scale, $\alpha = 1.000 \pm 0.001$, and a non-linear broadening reduced to $\Sigma_{NL} = 1.2 \pm 0.7\, h^{-1}$ Mpc. This value is smaller by a factor $\sim 1.7$ than the one obtained with eFAM$_1$, which is moreover totally dominated by errors.

### 4.3 Recovering the BAO signal in statistically anomalous samples

The analysis of the two-point correlation function before reconstruction shows that, in some sample, the scale of the BAO is badly constrained. We identified two types of samples: those returning a wrong best-fitting $\alpha_{unrec}$ typically deviating from the actual value $\alpha = 1$ more than $\sigma_\alpha$ (dubbed type-I samples), with a corresponding $\chi^2(\alpha)$ with a minimum significantly shifted from the true value; and samples without a clear acoustic feature (type-II samples), often yielding a $\chi^2(\alpha)$ with a very shallow minimum. The eFAM method is remarkably able to recover the correct $\alpha$ value from both types of anomalous samples.

For illustrative purposes, two representative examples of type-I anomalous samples are shown in Fig. 5. The correct shift of the

BAO peak in the two-point correlation function from its incorrect position pre-reconstruction (left column) to the right position around $\sim 110\, h^{-1}$ Mpc post-reconstruction (central column) is obtained, and consequently the correct value of $\alpha$ is measured. Also, the precision of the $\alpha$ estimate increases after the reconstruction. The shift of the minimum $\chi^2$ (right column) clearly illustrates and quantifies the de-biasing effect of the reconstruction. Analogously for type-II anomalous samples, three typical examples are shown in Fig. 6. Here the eFAM algorithm displays its remarkable ability to sharpen and shift the minimum, and consequently to significantly increase the statistical significance of the BAO peak, the ability to reconstruct it at its expected position, and to improve the precision of the measured $\alpha$ value. The sharpening of the BAO peak is paralleled by the decrease of the $\chi^2$ minimum, which was either not present or not unique before the reconstruction. BAO false positive are not produced, as tested by applying eFAM on Poisson samples.

### 4.4 RSD modelling: consistency tests

The velocities of DEUS-FUR haloes are not supplied. Since these are necessary to set the initial condition of the eFAM reconstruction in redshift space, we use the velocity predicted by eFAM itself. There is of course a certain degree of circularity in this procedure that may, in principle, artificially increase the accuracy of the reconstruction itself. To investigate this issue we run a specific test in which we considered a set of haloes extracted from the DEUS-648 simulation already used in Section 3.2 using a standard FoF algorithm that returns mass, position, and centre of mass velocity of each object. In the test we run two FAM reconstructions: one in which FoF velocities are used to set the initial conditions and the other in which we use FAM velocities. We then compared the monopole of the two-point correlation function computed from the haloes with known velocities, with the one obtained using the output of the reconstruction; see top panel of Fig. 7. At small separation the strong two-point correlation between the velocities of particles increases the amplitude of the correlation function in redshift-space, $\xi(s)$. This effect is artificially magnified when eFAM velocities are considered. However, this effect decreases with increasing relative separation when the velocities of the pairs become less correlated, asymptotically tending to zero at large scales; see bottom panel of Fig. 7. We therefore expect to obtain unbiased correlation functions at the BAO scale, which supports the robustness of our procedure. It is worth noticing that the number density of haloes in this test is approximately 100 times higher than in the BAO reconstructions, making this a very demanding consistency test. Moreover, no additional smoothing has been applied to mitigate non-linear effects and the same geometry and treatment of the tidal field as in the BAO tests has been adopted.

A more demanding test consists of comparing halo-by-halo the Cartesian components of the true velocities $\mathbf{v}_{i,Nbody}$ with those reconstructed by eFAM$_{10}$, $\mathbf{v}_{i,eFAM}$. In real space (Fig. 8, left-hand panel), regardless of the Cartesian component, the reconstructed velocities are typically overestimated by $20-25$ per cent irrespective of the extension of the outskirt accounting for the tidal field, with small and constant dispersion for velocities $\lesssim 1000\,$km s$^{-1}$. This is a well-known effect, already pointed out by Branchini et al. (2002) due to the biased density field that we use to compute the gravitational potential. While in linear theory this overestimation can be approximately undone by normalizing the reconstructed velocities by a factor $1/b$, with $b$ the effective halo bias at $z = 0$, the correction for FAM velocities is more complicated since this method goes beyond the linear theory. To correct for this
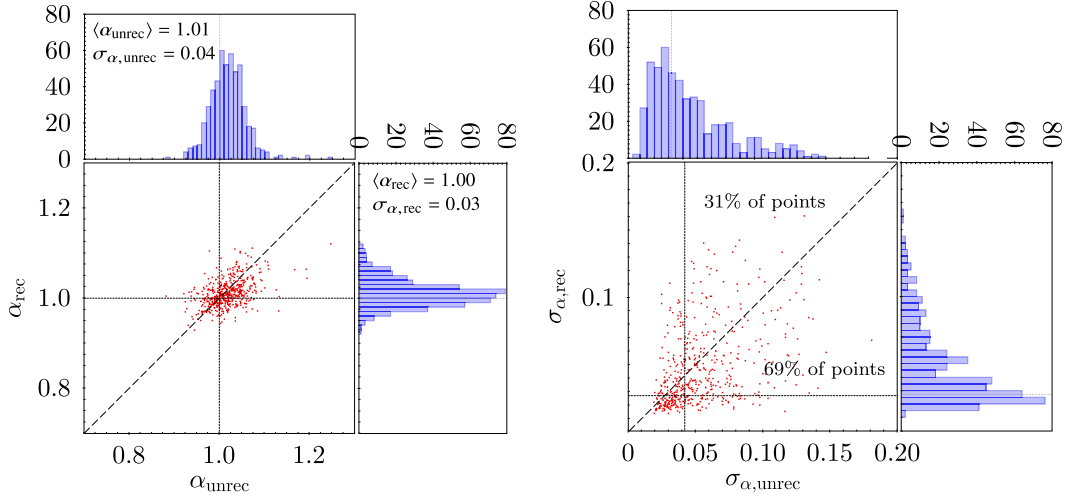
**Figure 4.** Comparison of scale dilation parameter $\alpha$ (left-hand panel) and its error (right-hand panel) estimated from the real-space mocks pre- ('unrec') and post-reconstruction ('rec'), together with the marginalized histograms, for the 512 mocks. *Left:* the black-dotted lines show the true values ($\alpha = 1$), the diagonal black-dashed line marks the perfect reconstruction ($\alpha_{\rm rec} = \alpha_{\rm unrec}$); the eFAM reconstruction reduces the scatter in these values without introducing any statistical bias. *Right:* short dashed line marks the equality of errors on the scale dilation parameter pre- and post-reconstruction, $\sigma_{\alpha,\,\rm rec} = \sigma_{\alpha,\,\rm unrec}$, the black-dotted lines correspond to the median values for the marginalized distributions; for 69 per cent of mocks the error $\sigma_\alpha$ decreases after reconstruction.
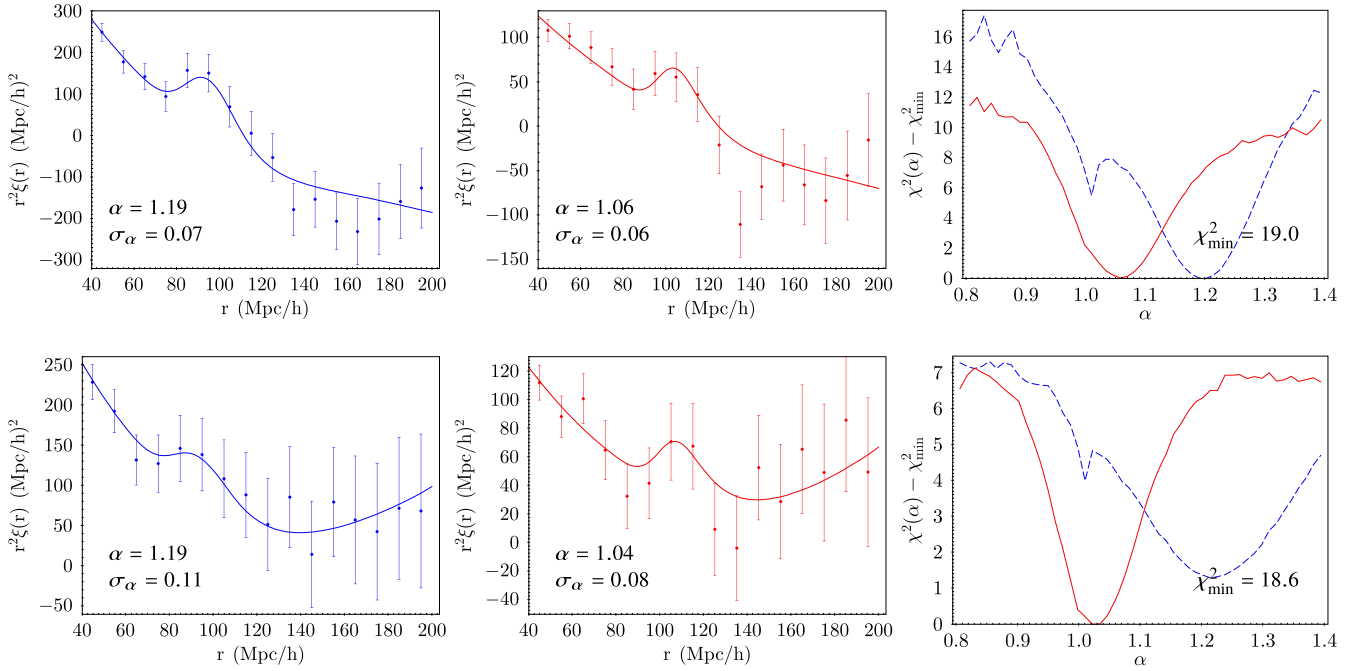


**Figure 5.** Fit results from the fiducial model for two type-I anomalous samples. *Left column:* Monopole of the two-point correlation function before reconstruction; the acoustic peak is shifted towards small scales. *Middle column:* Correlation function after reconstruction; the peak location is now compatible with the expected value within one standard deviation $\sigma_\alpha$. *Right column:* $\Delta\chi^2(\alpha) = \chi^2(\alpha) - \chi^2_{\rm min}$ before reconstruction (blue-dashed line) and after reconstruction (red-solid line); the shift of the best-fitting $\alpha$ towards the expected value $\alpha = 1$ reflects the shift of the peak in the correlation function.

effect, we weighted each halo mass by the number of dark matter particles within it. Though not accurate, this recipe does provide an approximate correction accounting for the haloes as biased tracers of the mass distribution. Besides, note that *N*-body velocities contain incoherent non-linear components that are not captured by the FAM reconstructed velocities. The small amplitude of the offset, reduced to about $10\,\rm km\,s^{-1}$ after considering an external buffer of thickness $200\,h^{-1}\,\rm Mpc$, indicates a not exact though largely sufficient modelling of the negligible bulk-flow offset. It

is worth noting that an overestimation of velocities by $\sim$20 per cent in amplitude shall result in an error on the redshift coordinate of about $\Delta s = 0.2 v_{\rm Nbody}/c \sim 2.7 \times 10^{-4}$ for the typical value $v_{\rm Nbody} = 400\,\rm km\,s^{-1}$, which is less than the usual error on the spectroscopic measurement of redshift.

In redshift space (Fig. 8, right-hand panel) the overestimation of reconstructed velocities is reduced to $\sim$8 per cent but slightly more scattered, with similar bulk flow as in real space. This is an effect of the Fingers-of-God, which act as a natural smoothing of the density
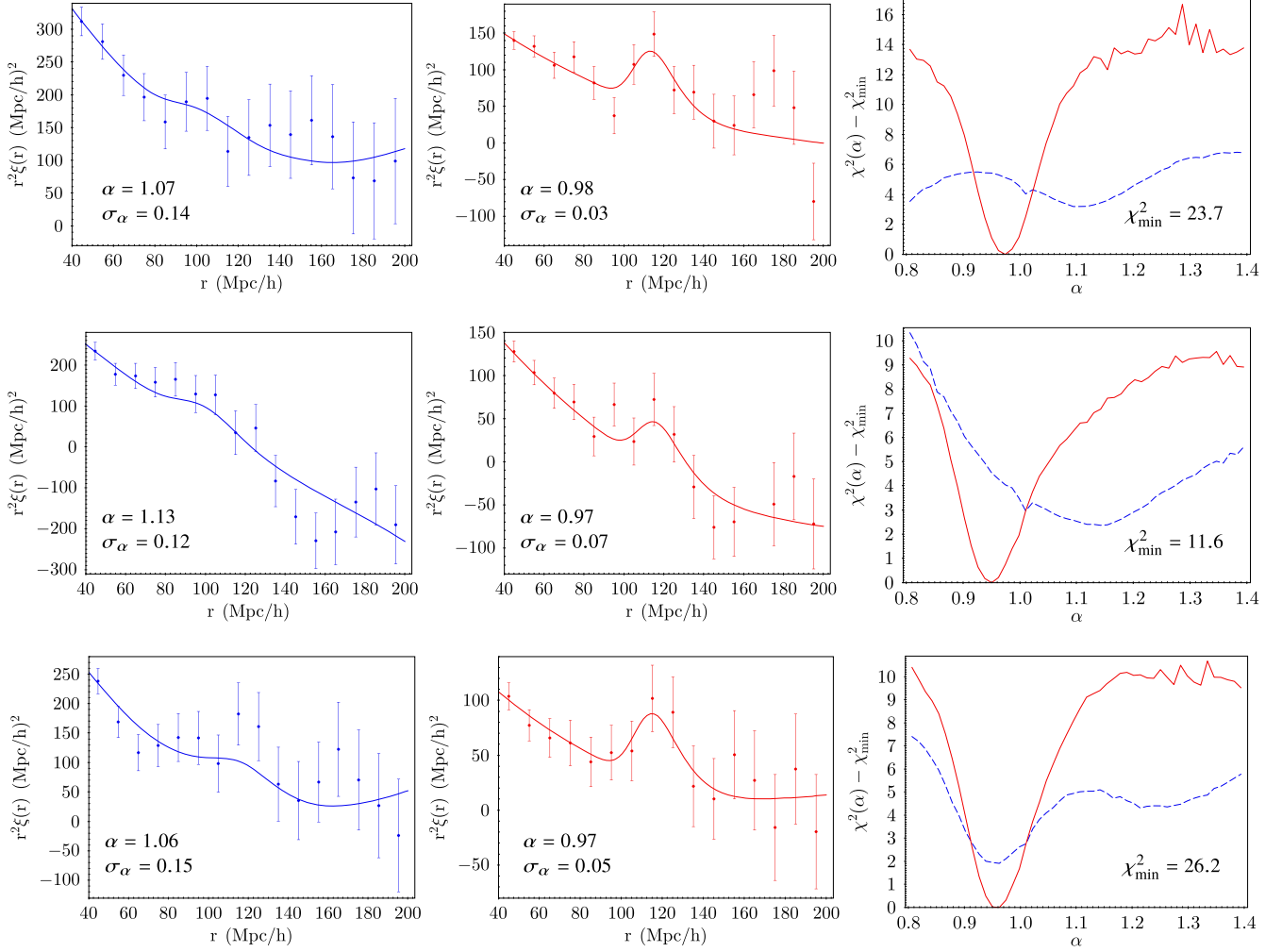
**Figure 6.** Fit results from the fiducial model for the type-II anomalous samples. *Left column:* Monopole of the two-point correlation function before reconstruction; the acoustic peak is not visible. *Middle column:* Correlation function after reconstruction; the peak is now clearly visible and the fit returns an unbiased value of $\alpha$. *Right column:* $\Delta\chi^2(\alpha) = \chi^2(\alpha) - \chi^2_{min}$ before reconstruction (blue-dashed line) and after reconstruction (red-solid line); the minimum of $\chi^2(\alpha)$ curve that pre-reconstruction was either unclear or not unique becomes well defined post-reconstruction.

field dumping the amplitudes of peculiar velocities. Accordingly, smaller peculiar velocities allow a reconstruction pushed at earlier time, from $z \sim 7$ in real space to $z \sim 40$ in redshift space using eFAM$_{10}$.

## 4.5 Monopole, quadrupole, and anisotropic correlation function in redshift space

The attractive feature of the eFAM technique is to recover the peculiar velocities of objects at their observed redshift, allowing for a non-parametric modelling of the RSD. This is illustrated in Fig. 9, which shows the density plots of the rescaled anisotropic two-point correlation function $s^2\bar{\xi}(s^{\parallel}, s^{\perp})$ as function of the longitudinal ($s^{\parallel}$) and transverse ($s^{\perp}$) components of the separation vector **s**, averaged over the 512 mocks. If the reconstruction is successful, $s$ represents the cosmological redshift with no peculiar velocity component in it. Before reconstruction (left-hand panel), the isotropy of the correlation function is broken by the RSD, which compresses the BAO ring at $\sim 110\,h^{-1}$ Mpc along the line of sight and split it into two arcs. This deformation is almost completely removed after correcting the density field in redshift space by subtracting

the longitudinal displacement due to the peculiar velocities, as estimated by eFAM at the same redshift of objects (middle panel). The BAO ring is further sharpened by reconstructing the density field at higher redshift (right-hand panel).

The power of the eFAM technique to improve the measurements of the acoustic scale enhancing the BAO signature becomes glaring looking at the monopole of the correlation function, $\tilde{\xi}(s)$ after the non-linear evolution is maximally reversed; see Fig. 10 top panel. This can be achieved with eFAM at order $M = 10$, reaching $z = 36.6$ (red line) when the monopole substantially matches the linear model (dashed line). If only the monopole is considered, the eFAM success to correct for RSD at $z = 0$ results limited (green line), only slightly improving the measurement of the acoustic scale, moderately sharpening and shifting the BAO peak towards the expected value. The results of the model fitting listed in Table 2 show that this reconstruction does not bias the measurements of the acoustic scale and reduces the value of the non-linear broadening $\Sigma_{NL}$ by 66 per cent. The efficiency of the reconstruction in the redshift-space, smaller than in real-space case, could be the result of the lack of precision in the assignment of the initial comoving redshift coordinates, as described in subsection 3.1
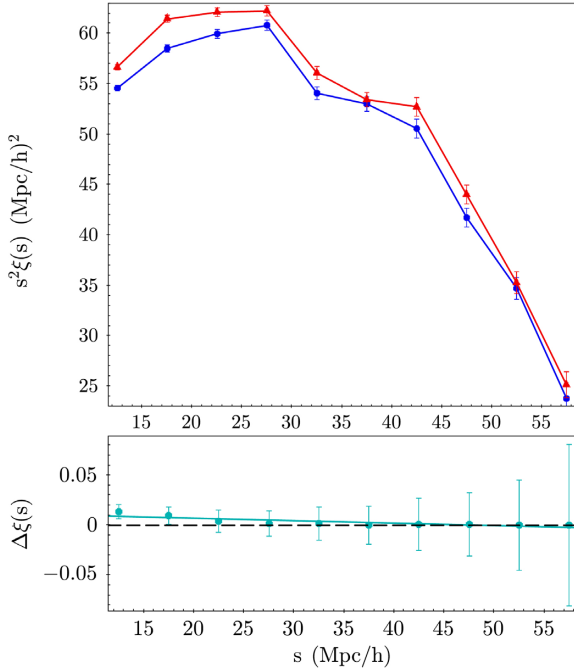
**Figure 7.** Accuracy test of eFAM$_{10}$ for the clustering statistics in redshift-space. *Top:* Rescaled monopole of the two-point correlation function computed from the haloes with known velocities (circles/blu line) and with eFAM peculiar velocities (triangles/red line). *Bottom:* Residuals of the correlation function. Assigning eFAM velocities to build the halo catalogues in redshift space results in an overestimation of the amplitudes of the correlation function, which decreases with increasing separation.

Although not fully recovering the clustering signal at the BAO scale at $z = 0$, eFAM efficiently restores statistical isotropy already at this redshift, as shown by the vanishing quadrupole of the correlation function a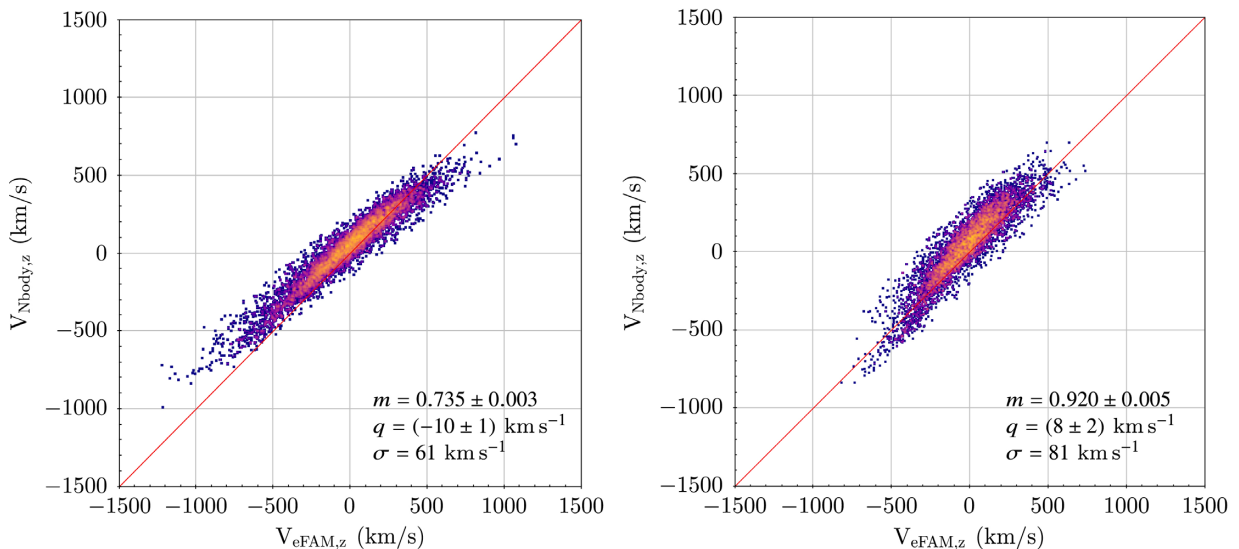fter reconstruction; see Fig. 10, bottom panel. To ease the comparison at different redshifts pre- and post-reconstruction, the rescaled quadrupole of the two-point correlation function is shown, $\tilde{\xi}_2(s) = 5[B D(z)]^{-2} \int_0^1 L_2(\mu) \xi(s, \mu) d\mu$, with $\mu$ the cosine of the angle between the separation vector and the line of sight and $L_2$ the Legendre polynomial of order two. Before reconstruction, the RSD brake the isotropy of the correlation function returning a non-zero value for $\tilde{\xi}_2$. The deviation from the isotropy is almost completely restored after correcting for the peculiar velocities at the observed redshift $z = 0$ and is further improved at small scales when the density field is reconstructed at higher redshift, $z = 36.6$.

As done for the real-space analysis, the impact of the reconstruction on the dilation parameter and its error is estimated by a point-wise comparison between the fitted values of $\alpha$ pre- and post-reconstruction from the 512 mocks; see Fig. 11. The distribution of $\alpha_{\rm rec}$ is more centred on the actual value, though its dispersions is only mildly improved. The improvement of the precision of $\alpha$ is less significant in redshift space, here the eFAM algorithm yielding $\sigma_{\alpha, \rm rec} < \sigma_{\alpha, \rm unrec}$ for the 61 per cent of mocks.

## 5 CONCLUSIONS

An extended version of the Fast Action Minimisation method (Nusser & Branchini 2000), dubbed eFAM, is presented, intended for applications with the next-generation massive spectroscopic surveys designed to observe billions of objects. Based on the Peebles (1989) LAP, the new algorithm coded in C++ reconstructs the trajectories of collisionless mass tracers in generic background cosmologies, owing to a parametrization of the orbits based on Jacobi polynomials, and works both in real and redshift-space with a new implementation. It further implements the powerful Poisson solver FALCON (Dehnen 2002), whose linear scaling with the number of particles realistically allows for application to large spectroscopic samples with $O(10^6)$ objects (eventually per redshift bin in deep surveys).

For the first time a numerical action method is used for BAO reconstruction. Primarily interested in large scales, where the
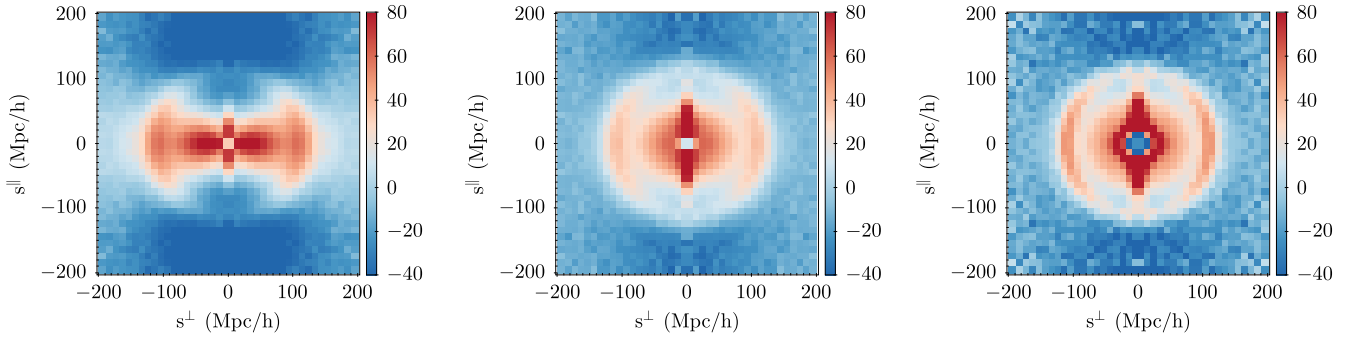


**Figure 8.** Accuracy tests of eFAM$_{10}$ in real space (left) and redshift space (right). Reconstructed versus true peculiar velocities of haloes at $z = 0$ for one Cartesian component (results are similar for other components). A perfect reconstruction would give a linear regression $V_{\rm eFAM} = mV_{\rm Nbody} + q$ with slope $m = 1$ (solid line), no residual bulk velocity ($q = 0$), and no scatter. The reconstructed peculiar velocities in real space are slightly overestimated though well correlated with the true ones. In redshift space the reconstructed peculiar velocities are definitely less biased but more scattered as expected.

**Figure 9.** Rescaled two-point correlation function in redshift-space, averaged over the 512 mocks, as function of the transverse ($s^{\perp}$) and line-of-sight ($s^{\parallel}$) separation. *Left-hand panel:* Before reconstruction, from objects at $z = 0$; the isotropy of the acoustic feature, which should be visible as a ring in the ($s^{\parallel}$, $s^{\perp}$) plane, is broken by RSD. *Middle panel:* Correlation function after correcting the density field for the RSD, again at the observed redshift $z = 0$; the isotropy of the acoustic feature is almost completely restored. *Right-hand panel:* Correlation function after reconstruction at $z = 33.6$; the BAO feature is sharper and symmetric, indicating the quality of the reconstruction.
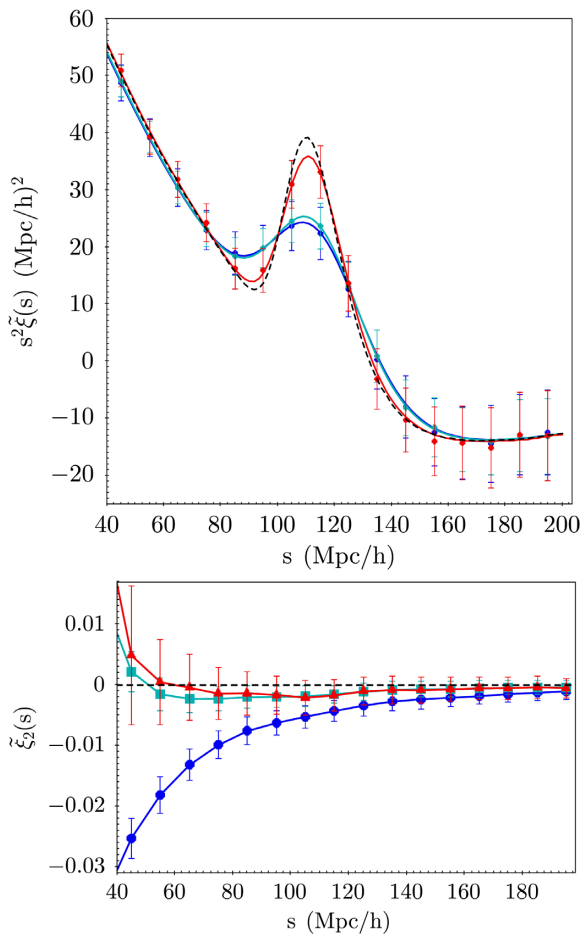


**Figure 10.** Rescaled monopole (top) and quadrupole (bottom) of the two-point correlation function in redshift-space pre- and post-reconstruction by eFAM$_{10}$, averaged over the 512 mocks; error bars from the diagonal of the covariance matrix. Before reconstruction at observed redshift $z = 0$ (circles/blue line) the acoustic feature in the monopole is broadened by non-linear evolution and peculiar velocities. The correction for RSD is effective at $z = 0$ (squares/green line) as shown in the quadrupole, but the BAO peak is only slightly enhanced. Pushing the reconstruction at the highest redshift possible, $z = 33.6$ (triangles/red line), the monopole well approximates the linear model (dashed line) and the quadrupole is still consistent with zero, showing the efficiency of the eFAM method in both sharpening the peak and correcting for RSD.

**Table 2.** Fit results to average mock correlation functions in redshift space using eFAM$_{10}$.

| Type | $\alpha$ | $\Sigma_{NL}$ ($h^{-1}$ Mpc) |
|---|---|---|
| pre-recon, $z = 0$ | $1.007 \pm 0.002$ | $11.8 \pm 0.3$ |
| post-recon, $z = 0$ | $1.005 \pm 0.002$ | $11.0 \pm 0.3$ |
| post-recon, $z = 36.6$ | $0.997 \pm 0.001$ | $4.0 \pm 0.5$ |

complexities of galaxy formation and fully non-linear clustering are mitigated, eFAM is probed with mock haloes whose large-scale (quasi-Newtonian) dynamics only mildly deviate from the Hubble flow. A future study will extend this method to mock and real galaxy (rather than halo) catalogues, accounting for the bias and selection function of tracers as done in NB00.

We have tested the eFAM algorithm on 512 independent halo catalogues extracted from the DEUS-FUR $\Lambda$CDM simulation in real space, each with about 56 000 dark matter haloes of mass larger than $1.2 \times 10^{14} \, h^{-1} \, M_{\odot}$, typical value for galaxy clusters, in a spherical volume of radius $\sim 1 \, h^{-1}$ Gpc. Because of the petabyte size of the parent catalogue, the peculiar velocities of the FoF haloes were not available. The catalogues in redshift space are therefore built by modelling the comoving redshift coordinates from the peculiar velocities as reconstructed by eFAM in real-space. Both in real and redshift-space, the reconstructed trajectories are finally trusted only in spheres of radius $\sim 700 \, h^{-1}$ Mpc, each containing about 23 000 haloes, using the mass in the external shell to model the tidal forces by direct computation; this assures a correct estimation of velocities within a $\sim 10$ per cent error in redshift space.

We first evaluated the performances of the fully non-linear reconstruction by eFAM, namely using an orbit expansion at 10th order (eFAM$_{10}$), in recovering the linear model of the monopole of the two-point correlation function in real space. The eFAM$_{10}$ algorithm successfully recovers the linear correlation function at the BAO scale, reducing the non-linear broadening of the acoustic feature $\Sigma_{NL}$ by 87 per cent from $9.0 \pm 0.2 \, h^{-1}$ Mpc at $z = 0$ to $1.2 \pm 0.7 \, h^{-1}$ Mpc at $z = 6.5$. Moreover, eFAM$_{10}$ returns an unbiased and improved position of the acoustic scale as measured by the dilation parameter, $\alpha_{rec} = 1.000 \pm 0.001$, reducing its associated error $\sigma_{\alpha}$ in 69 per cent of the mocks. Instead, the reconstruction achieved by the Zel'dovich approximation obtained from the first guess, i.e. eFAM$_1$, is not equally powerful; the huge errors in the average correlation function post-reconstruction yields
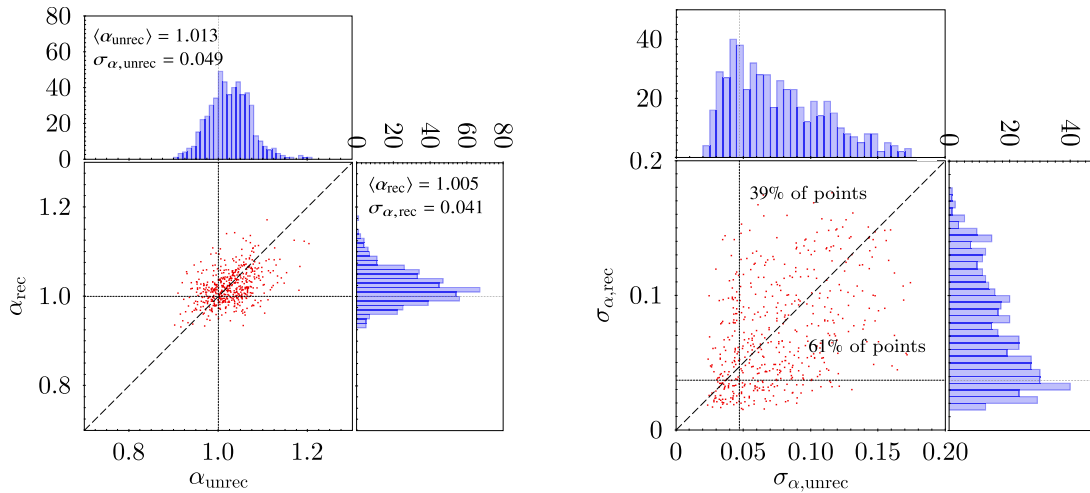
**Figure 11.** As Fig. 4 but in redshift space. The quality of the reconstruction mildly worsen with respect to real space, however this data analysis is based non-informative flat priors.

a value of $\Sigma_{NL}$ larger by a factor $\sim 1.7$ and a biased estimation of the BAO scale with 3–10 times larger error. Moreover, this first-order approximation eFAM$_1$ only allows for linear trajectories that quickly undergo unphysical crossing, limiting the reconstruction at much lower redshift than eFAM$_{10}$, which more easily removes the effects of non-linear clustering.

Allowing for a reconstruction pushed to very high redshift, the eFAM algorithm is extremely powerful in recovering the BAO feature from anomalous samples that, without reconstruction, would show a wrong location of the BAO peak in monopole of the two-point correlation function (real-space), or no BAO signal at all. Using eFAM$_{10}$ the BAO feature is correctly restored and the signal increased with high statistical significance.

In redshift space, the fully non-linear eFAM algorithm successfully corrects for the RSD. By correcting the comoving redshift coordinates of objects using the reconstructed peculiar velocities, eFAM$_{10}$ already restores the isotropy of the two-dimensional correlation function at the observed redshift, here $z = 0$. Performing the non-linear reconstruction at the highest redshift possible before shell-crossing, here $z = 33.6$, the acoustic ring is efficiently restored. The BAO feature in the azimuthally averaged two-point correlation function $\xi(s)$ is correspondingly well sharpened, with a 66 per cent reduction of the $\Sigma_{NL}$ broadening parameter from $11.8 \pm 0.3$ to $4.0 \pm 0.5 \, h^{-1}$ Mpc. Although not reproducing the internal dynamics in virialized haloes, the fully non-linear eFAM technique achieves a very good accuracy in reconstructing the dynamics down to scales comparable to the mean inter-halo separation, i.e. $\sim 10 \, h^{-1}$ Mpc, as shown by point-wise comparison of real (simulated) and reconstructed velocities of haloes from small DEUS simulations. This opens the possibility of a non-parametric modelling of RSD, possibly exploring the effect of local environment on the reconstruction (Achitouv & Blake 2015); this is left for a future study.

A final remark on the computational load. The CPU time is driven by the force computation, which scales linearly with the number of particles $N$, and not by the minimization procedure. Indeed, the computational efficiency of the code is almost independent of the dimension $M$ of the basis set; for DEUS-648 $M = 4$ is large enough to reliably reconstruct the orbits from the linear to the non-linear

regime. The computational complexity increases by a factor of $\sim 5$ in redshift space; eFAM being an iterative reconstruction algorithm, the number of iteration required to relax to a minimum of the action is significantly larger in redshift space, where the initial conditions of the particles are set by the observed redshifts rather than positions.

## ACKNOWLEDGEMENTS

## REFERENCES

Abramowitz M., Stegun I. A., 1964, Handbook of Mathematical Functions : With Formulas, Graphs, and Mathematical Tables Applied mathematics series, National Bureau of Standards, Washington, DC, p. 295
Achitouv I., Blake C., 2015, Phys. Rev., 92, 083523
Anderson L. et al., 2012, MNRAS, 427, 3435
Anderson L. et al., 2014a, MNRAS, 439, 83
Anderson L. et al., 2014b, MNRAS, 441, 24
Avila S., Murray S. G., Knebe A., Power C., Robotham A. S. G., Garcia-Bellido J., 2015, MNRAS, 450, 1856
Beutler F., Blake C., Koda J., Marín F. A., Seo H.-J., Cuesta A. J., Schneider D. P., 2016, MNRAS, 455, 3230
Blake C., Glazebrook K., 2003, ApJ, 594, 665
Branchini E., Carlberg R. G., 1994, ApJ, 434, 37
Branchini E., Eldar A., Nusser A., 2002, MNRAS, 335, 53
Brenier Y., Frisch U., Hénon M., Loeper G., Matarrese S., Mohayaee R., Sobolevskiǐ A., 2003, MNRAS, 346, 501
Burden A., Percival W. J., Manera M., Cuesta A. J., Vargas Magana M., Ho S., 2014, MNRAS, 445, 3152

Burden A., Percival W. J., Howlett C., 2015, MNRAS, 453, 456

Chuang C.-H., Kitaura F.-S., Prada F., Zhao C., Yepes G., 2015, MNRAS, 446, 2621

Cole S. et al., 2005, MNRAS, 362, 505

Crocce M., Scoccimarro R., 2008, Phys. Rev., 77, 023533

Croft R. A. C., Gaztanaga E., 1997, MNRAS, 285, 793

Cuesta A. J. et al., 2016, MNRAS, 457, 1770

Dawson K. S. et al., 2016, AJ, 151, 44

De Jong Roelof S, et al., 2012, Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, , SPIE, Bellingham. p.

Doré O. et al., 2018, preprint (arXiv:1804.03628)

Dunn A. M., Laflamme R., 1995, ApJ, 443, L1

Eisenstein D. J. et al., 2005, ApJ, 633, 560

Eisenstein D. J., Hu W., 1998, ApJ, 496, 605

Eisenstein D. J., Hu W., Silk J., Szalay A. S., 1998, ApJ, 494, L1

Eisenstein D. J., Seo H.-J., White M., 2007a, ApJ, 664, 660

Eisenstein D. J., Seo H.-J., Sirko E., Spergel D. N., 2007b, ApJ, 664, 675

Falck B. L., Neyrinck M. C., Aragon-Calvo M. A., Lavaux G., Szalay A. S., 2012, ApJ, 745, 17

Frisch U., Matarrese S., Mohayaee R., Sobolevski A., 2002, Nature, 417, 260

Giavalisco M., Mancinelli B., Mancinelli P. J., Yahil A., 1993, ApJ, 411, 9

Gil-Marín H. et al., 2016, MNRAS, 460, 4210

Howlett C., Manera M., Percival W. J., 2015, Astron. Comput., 12, 109

Hu W., Haiman Z., 2003, Phys. Rev., 68, 063004

Hu W., Sugiyama N., 1996, ApJ, 471, 542

Jasche J., Wandelt B. D., 2013, MNRAS, 432, 894

Kaiser N., 1987, MNRAS, 227, 1

Kazin E. A. et al., 2014, MNRAS, 441, 3524

Keselman J. A., Nusser A., 2017, MNRAS, 467, 1915

Kitaura F.-S., Angulo R. E., 2012, MNRAS, 425, 2443

Kitaura F.-S., Angulo R. E., Hoffman Y., Gottlöber S., 2012, MNRAS, 425, 2422

Kitaura F.-S., Yepes G., Prada F., 2014, MNRAS, 439, L21

Kitaura F. S., Enßlin T. A., 2008, MNRAS, 389, 497

Landy S. D., Szalay A. S., 1993, ApJ, 412, 64

Laureijs R. et al., 2011, preprint (arXiv:1110.3193)

Lavaux G., Tully R. B., Mohayaee R., Colombi S., 2010, ApJ, 709, 483

Lewis A., Bridle S., 2002, Phys. Rev., 66, 103511

Manera M. et al., 2013, MNRAS, 428, 1036

Martini P. et al., 2018, in Christopher J. E. , Luc S., Hideki T., eds, Proc. SPIE Conf. Ser., Vol. 10702, Ground-based and Airborne Instrumentation for Astronomy VII. SPIE, Bellingham, p. 107021F

Marulli F., Veropalumbo A., Moresco M., 2016, Astron. Comput., 14, 35

McCullagh N., Neyrinck M. C., Szapudi I., Szalay A. S., 2013, ApJ, 763, L14

Mehta K. T., Seo H.-J., Eckel J., Eisenstein D. J., Metchnik M., Pinto P., Xu X., 2011, ApJ, 734, 94

Mohayaee R., Frisch U., Matarrese S., Sobolevskii A., 2003, A&A, 406, 393

Mohayaee R., Mathis H., Colombi S., Silk J., 2006, MNRAS, 365, 939

Monaco P., Efstathiou G., 1999, MNRAS, 308, 763

Monaco P., Theuns T., Taffoni G., 2002, MNRAS, 331, 587

Munari E., Monaco P., Sefusatti E., Castorina E., Mohammad F. G., Anselmi S., Borgani S., 2017, MNRAS, 465, 4658

Noh Y., White M., Padmanabhan N., 2009, Phys. Rev., 80, 123501

Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, MNRAS, 396, 19

Ntelis P. et al., 2017, J. Cosmol. Astropart. Phys., 6, 019

Nusser A., Branchini E., 2000, MNRAS, 313, 587

Nusser A., Colberg J. M., 1998, MNRAS, 294, 457

Nusser A., Dekel A., Bertschinger E., Blumenthal G. R., 1991, ApJ, 379, 6

Padmanabhan N., White M., Cohn J. D., 2009, Phys. Rev., 79, 063523

Padmanabhan N., Xu X., Eisenstein D. J., Scalzo R., Cuesta A. J., Mehta K. T., Kazin E., 2012, MNRAS, 427, 2132

Peebles P. J. E., 1989, ApJ, 344, L53

Peebles P. J. E., 1994, ApJ, 429, 43

Peebles P. J. E., 1995, ApJ, 449, 52

Peebles P. J. E., Yu J. T., 1970, ApJ, 162, 815

Planck Collaboration, 2018, preprint (arXiv:1807.06209)

Rasera Y., Corasaniti P.-S., Alimi J.-M., Bouillot V., Reverdy V., Balmès I., 2014, MNRAS, 440, 1420

Reverdy V. et al., 2015, Int. J. High Perform. Comput. Appl., 29, 249

Romano-Díaz E., Branchini E., van de Weygaert R., 2005, A&A, 440, 425

Ross A. J. et al., 2014, MNRAS, 437, 1109

Schmoldt I. M., Saha P., 1998, AJ, 115, 2231

Scoccimarro R., Sheth R. K., 2002, MNRAS, 329, 629

Seo H.-J., Eisenstein D. J., 2003, ApJ, 598, 720

Shaya E. J., Peebles P. J. E., Tully R. B., 1995, ApJ, 454, 15

Silk J., 1968, ApJ, 151, 459

Spergel D. N. et al., 2007, ApJS, 170, 377

Sunyaev R. A., Zeldovich Y. B., 1970, Ap&SS, 7, 3

Takada M. et al., 2014, PASJ, 66, R1

Tassev S., Zaldarriaga M., 2012, J. Cosmol. Astropart. Phys., 10, 006

Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, J. Cosmol. Astropart. Phys., 6, 036

Tojeiro R. et al., 2014, MNRAS, 440, 2222

Wang H., Mo H. J., Yang X., van den Bosch F. C., 2013, ApJ, 772, 63

White M., 2005, Astropart. Phys., 24, 334

White M., 2014, MNRAS, 439, 3630

Xu X., Padmanabhan N., Eisenstein D. J., Mehta K. T., Cuesta A. J., 2012, MNRAS, 427, 2146

## APPENDIX A: BASIS FUNCTIONS AND JACOBI POLYNOMIALS

The Jacobi polynomials $p_n^{(\alpha,\beta)}(x)$, defined for $n = 0, 1, ...$ and $\alpha, \beta > 1$, satisfy the orthogonality condition

$$\int_{-1}^{1} dx (1-x)^\alpha (1+x)^\beta p_n^{(\alpha,\beta)}(x) p_m^{(\alpha,\beta)}(x) = h_n \delta_{nm} \qquad (A1)$$

that can be determined using the recurrence relation

$$p_{n+1}^{(\alpha,\beta)}(x) = (A_n x + B_n) p_n^{(\alpha,\beta)}(x) - C_n p_{n-1}^{(\alpha,\beta)}(x). \qquad (A2)$$

For the expression of the coefficients $h_n$, $A_n$, $B_n$, and $C_n$, see Abramowitz & Stegun (1970).

For $\Lambda$CDM and closer cosmologies the weight function $w(D) = f(D)E(D)Da^2(D)$ in equation (3) is almost indistinguishable from a power law. Defining $x = 2(D/D_{obs}) - 1$, $w(D)$ can be fitted by the weight function $K(1-x)^\alpha(1+x)^\beta = K(2D/D_{obs} - 2)^\alpha(2D/D_{obs})^\beta$ that settles the orthogonality condition (A2). The best-fitting values for $K$, $\alpha$, and $\beta$ only mildly depend on the specific background cosmology; for a standard cold dark matter model the exact values $(\alpha, \beta) = (0, 3/2)$ are recovered, while for the WMAP-7 $\Lambda$CDM model the best fit is $(\alpha, \beta) \approx (0, 1.53)$.

Once the values of $\alpha$ and $\beta$ are fixed, the functions $q_n^{(\alpha,\beta)}$ for $n \geq 1$ are given by

$$q_n^{(\alpha,\beta)}(D) = \int_0^{D_{obs}} dD \, p_n^{(\alpha,\beta)}(D) = \frac{D_{obs}}{n + \alpha + \beta} p_{n+1}^{(\alpha-1,\beta-1)}, \qquad (A3)$$

where the relation

$$\frac{dp_n^{(\alpha,\beta)}}{dx} = \frac{1}{2}(n + \alpha + \beta + 1) p_{n-1}^{(\alpha+1,\beta+1)} \qquad (A4)$$

has been used. The asymptotic limit of the $p_n^{(\alpha,\beta)}(x)$'s for $x \to -1$, i.e. $D \to 0$, guarantees the vanishing of the initial peculiar velocities while the constraints on the observed positions is satisfied choosing the integration constant in (A3).

## APPENDIX B: PARAMETRIZATION OF ORBITS IN REDSHIFT SPACE

The parametrization of the $i$-th particle's trajectory parallel and perpendicular to the line of sight, explicitly accounting for the cosmological dependence, in redshift space reads

$$\mathbf{x}_i^{\parallel}(D) = \frac{c\mathbf{s}_{i,0}}{H_0 a_0} + \sum_{n=0}^{M} \mathbf{C}_{i,n}^{\parallel} Q_n(D) \tag{B1}$$

$$\mathbf{x}_i^{\perp}(D) = \sum_{n=0}^{M} \mathbf{C}_{i,n}^{\perp} q_n(D) \tag{B2}$$

where $Q_n(D) \equiv q_n(D) - (fDE)_{\mathrm{obs}} p_{n,\,\mathrm{obs}}$.

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