A first step towards a direct inversion of the Lyman forest in QSO spectra

Adi Nusser[1]* and Martin Haehnelt[2]

[1]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, D–85740 Garching bei München, Germany*
[2]*Institute of Astronomy, Madingley Road, Cambridge CB3 0HA*

ABSTRACT

A method for the recovery of the real space line-of-sight mass density field from Lyman absorption in QSO spectra is presented. The method makes use of a Lucy-type algorithm for the recovery of the H I density. The matter density is inferred from the H I density assuming that the absorption results from a photoionized intergalactic medium that traces the mass distribution as suggested by recent numerical simulations. Redshift distortions are corrected iteratively from a simultaneous estimate of the peculiar velocity. The method is tested with mock spectra obtained from N-body simulations. The density field is recovered reasonably well up to densities where the absorption features become strongly saturated. The method is an excellent tool with which to study the density probability distribution and clustering properties of the mass density in the (mildly) non-linear regime. Combined with redshift surveys along QSO sightlines, the method will make it possible to relate the clustering of high-redshift galaxies to the clustering of the underlying mass density. We further show that accurate estimates for $(\Omega_{bar}h^2)^2 J^{-1} H(z)^{-1}$ and higher order moments of the density probability function can be obtained despite the missing high-density tail of the density distribution if a parametric form for the probability distribution of the mass density is assumed.

Key words: intergalactic medium – quasars: absorption lines – cosmology: observations – cosmology: theory – dark matter – large-scale structure of Universe.

1 INTRODUCTION

The Lyman forest in QSO absorption spectra is now generally believed to be caused by absorption by large-scale neutral hydrogen (H I) density fluctuations of moderate amplitude in a warm photo-ionized intergalactic medium (IGM). This relatively new paradigm for the forest differs considerably from the conventional 'cloud' picture, which had been advocated for two decades and in which the Lyman forest is due to a superposition from discrete absorbers with small cross-sections (see Rauch 1998 for a review). The new picture had been investigated using analytical calculations (e.g. Bond, Szalay & Silk 1988; McGill 1990; Bi, Börner & Chu 1992) but was only generally accepted after the coherence length of the absorbing structure could be measured accurately and turned out to be of the order of several hundred kpc (Dinshaw et al. 1994; e.g. Bechtold et al. 1994). The picture was then further sustained by (hydrodynamical) cosmological simulations of gas in dark matter dominated universes (Cen et al. 1994; Petitjean, Mücket & Kates 1995; Zhang, Anninos & Norman 1995; Hernquist et al. 1996; Miralda-Escudé et al. 1996). Important results of these simulations are a tight correlation between the H I and the dark matter distribution (on scales larger than the Jeans length of the IGM) and a simple

temperature–density relation for the IGM, which depends only on the reionization history of the Universe (e.g. Hui & Gnedin 1997; Haehnelt & Steinmetz 1998).

As a consequence of this new picture the Lyman forest can be used to probe the distribution and clustering of dark matter at high redshift. For example, Bi & Davidsen (1997) have developed a simple analytic model for the IGM to generate artificial QSO absorption spectra for a variant of the cold dark matter (CDM) cosmogony. They were able to reproduce the characteristic properties of observed absorption spectra. Croft et al. (1998) showed that absorption spectra provide important information on the shape and the amplitude of the power spectrum of mass fluctuations. Croft et al. used the following procedure. They first obtained a linearized flux distribution by applying the Gaussianization scheme proposed by Weinberg (1992). They then inferred the shape of the linear power spectrum of dark matter density fluctuation from the power spectrum of the linearized flux and used mock spectra from numerical simulations to determine the amplitude of the power spectrum. Gnedin & Hui (1998) used mock spectra generated from simulations of collisionless particles run with a particle-mesh code modified to mimic pressure effects of the gas to investigate the effect of amplitude and the power spectrum of dark matter fluctuations on the column density distribution of absorption systems.

In this paper a complementary approach is taken. We propose to use an analytical model of the IGM for a direct inversion of the absorption features in QSO spectra. This approach has the

*Present address: Physics Department, Technion–Israel Institute of Technology, Haifa 32000, Israel.

advantage that no particular cosmological model has to be assumed. Furthermore, the actual real space density along the LOS and its probability distribution can be studied and a direct link to other observations is posssible.

The paper is organized as follows. In Section 2 we describe the assumed model for the Lyman absorption. Section 3 presents the inversion algorithm for recovering the *real* space dark matter (DM) density along the line of sight, concentrating on the forest. In Section 4 we test the inversion procedure with mock spectra generated from numerical simulations of collisional dark matter and show how to estimate higher order moments of the density probability distribution and $(\Omega_{\text{bar}}h^2)^2 J^{-1} H(z)^{-1}$ [the parameter combination of baryonic density $\Omega_{\text{bar}}h^2$, ionizing flux J and Hubble constant $H(z)$ that determines the mean flux level in QSO absorption spectra]. In Section 5 we discuss possible applications and give our conclusions.

2　LYMAN ABSORPTION BY A PHOTOIONIZED INTERGALACTIC MEDIUM

Absorption spectra are normally presented in the form of a normalized flux which (neglecting noise and instrumental broadening) can be related to the optical depth as

$$F(w) = I_{\text{obs}}(w)/I_{\text{cont}}(w) = e^{-\tau(w)}, \tag{1}$$

where τ is the optical depth, w is the redshift space coordinate, $I_{\text{obs}}(w)$ is the observed flux and $I_{\text{cont}}(w)$ is the flux emitted from the quasar that would be observed without intervening absorption and which has to be estimated from the data as well.

The optical depth in redshift space due to resonant scattering is related to the neutral hydrogen density, $n_{\text{H\,I}}$, along the line of sight in real space as (Gunn & Peterson 1965; Bahcall & Salpeter 1965)

$$\tau(w) = \sigma_0 \frac{c}{H(z)} \int_{-\infty}^{\infty} n_{\text{H\,I}}(x)\, \mathcal{H}[w - x - v_{\text{p}}(x), b(x)]\, dx, \tag{2}$$

where σ_0 is the effective cross-section for resonant line scattering, $H(z)$ is the Hubble constant at redshift z, x is the real space coordinate (in km s^{-1}), $\mathcal{H}$ is the Voigt profile normalized such that $\int \mathcal{H}\, dx = 1$, $v_{\text{p}}(x)$ is the line of sight peculiar velocity, and $b(x)$ is the Doppler parameter due to thermal/turbulent broadening. For moderate optical depths the Voigt profile is well approximated by a Gaussian, $\mathcal{H} = 1/(\sqrt{\pi}\,b) \times \exp\{-[w - x - v_{\text{p}}(x)]^2/b^2)\}$. Assuming that hydrogen is highly ionized and in photoionization equilibrium ($n_{\text{H\,I}} \propto n_{\text{H}}^2 J^{-1}$) we have

$$n_{\text{H\,I}} = \hat{n}_{\text{H\,I}} \left[\frac{n_{\text{H}}(\boldsymbol{x})}{\bar{n}_{\text{H}}}\right]^{\alpha} = \hat{n}_{\text{H\,I}} \left[\frac{\rho(\boldsymbol{x})}{\bar{\rho}}\right]^{\alpha}, \tag{3}$$

where $\hat{n}_{\text{H\,I}}$ is the neutral hydrogen density at the mean total gas density and $1.56 < \alpha < 2$ (Hui & Gnedin 1997). The second relation assumes that the gas density traces the dark matter density. At the low densities considered here shock heating is not important and the gas is at the photoionization equilibrium temperature. The Doppler parameter can then be related to the total gas density and temperature as

$$b(x) = 13 \left(\frac{\hat{T}}{10^4\,\text{K}}\right)^{0.5} \left[\frac{\rho_{\text{DM}}(\boldsymbol{x})}{\bar{\rho}_{\text{DM}}}\right]^{\beta}\, \text{km s}^{-1}, \tag{4}$$

where $\hat{T}$ is the temperature at the mean density and $0 < \beta < 0.31$ (Hui & Gnedin 1997).

Combining equations (2), (3) and (4) we get

$$\tau(w) = \mathcal{A}(z) \int_{-\infty}^{\infty} \left[\frac{\rho_{\text{DM}}(\boldsymbol{x})}{\bar{\rho}_{\text{DM}}}\right]^{\alpha} \mathcal{H}[w - x - v_{\text{p}}(x), b(x)]\, dx, \tag{5}$$

$$\mathcal{A}(z) = \sigma_0 \frac{c}{H(z)}\, \hat{n}_{\text{H\,I}}$$

$$\sim 0.12 h^{-1}\, \Omega_{\text{mat}}^{-0.5} \left(\frac{\Omega_{\text{bar}}h^2}{0.0125}\right)^2$$

$$\times \left(\frac{\Gamma}{10^{-12}\,\text{s}^{-1}}\right)^{-1} \left(\frac{\hat{T}}{10^4\,\text{K}}\right)^{-0.7} \left(\frac{1+z}{4}\right)^{4.5}, \tag{6}$$

where Ω_{bar} and Ω_{mat} are the baryonic and total matter density in terms of the critical density and Γ is the photoionization rate per hydrogen atom (Γ is related to the flux of ionizing radiation J_ν as $\Gamma = 4\pi \int d\nu\, \sigma_\nu\, J_\nu/h\nu$, where σ_ν is the hydrogen absorption cross-section). To obtain the second relation in equation (6) we have have taken $\alpha_{\text{rec}} = 4.7 \times 10^{-13}(T/10^4\,\text{K})^{-0.7}\,\text{cm}^{-3}\,\text{s}^{-1}$ for the hydrogen recombination coefficient, $\sigma_0 = 4.5 \times 10^{-18}\,\text{cm}^2$ as the effective cross-section for resonant scattering and used the high-redshift approximation for the Hubble constant, $H(z) = H_0\, \Omega_{\text{mat}}^{1/2}\, (1 + z)^{3/2}$.

3　THE INVERSION ALGORITHM

3.1　The basic iterative scheme

The proposed scheme for recovering the line of sight density from the flux is motivated by Lucy's method (Lucy 1974). In order to demonstrate the method let us, for the time being, consider the case where $\tau \ll 1$ and $v_{\text{p}} = 0$. In this case, we have

$$1 - F(w) = \int n_{\text{H\,I}}(x) G(w, x)\, dx. \tag{7}$$

This is a linear integral equation with a positive kernel $G(w, x) = \mathcal{H}[w - x, b(x)]\sigma_0 c/H(z)$ and can be solved for $n_{\text{H\,I}}$ using Lucy's iterative method. This is done in the following way. Divide w into equal bins of size Δw (the data are actually given in bins of w anyway) and set $f_i = 1 - F_i$. Provide an initial guess for $(n_{\text{H\,I}})_j$, say $(n_{\text{H\,I}})_j^0 = 1 - F_j H(z)/(\sigma_0 c)$. Denote the values of $(n_{\text{H\,I}})_j$ at the rth iteration by $(n_{\text{H\,I}})_j^r$ and evaluate the sum

$$f_i^r = \sum_j (n_{\text{H\,I}})_j^r\, G_{ij}^r\, \Delta w, \tag{8}$$

where

$$G_{ij}^r = \mathcal{H}[w_i - x_j, b_j^r]\, \frac{\sigma_0\, c}{H(z)}. \tag{9}$$

The $(r + 1)$th estimate of $(n_{\text{H\,I}})_j$ is then

$$(n_{\text{H\,I}})_i^{r+1} = \left[\frac{1}{2m + 1} \sum_{k=i-m}^{i+m} (n_{\text{H\,I}})_k^r\right] \frac{\sum_j (f_j/f_j^r)\tilde{G}_{ij}}{\sum_j \tilde{G}_{ij}}, \tag{10}$$

where $\tilde{G}$ is some kernel that in principle may be chosen to be different from G. Lucy's method, however, uses $\tilde{G} = G$. The averaging over $2m + 1$ adjacent bins mitigates the effect of noise in the data. We work with $m = 3$.

Now we generalize the iteration scheme to the case where τ is not necessarily small. In this case, the relation between F and τ is non-linear. However, since the relation is monotonic, we can still use a modified version of equation (10),

$$f_i^r = 1 - \exp[-\sum_j (n_{\text{H\,I}})_j^r\, G_{ij}^r\, \Delta w]. \tag{11}$$

How do we decide when to stop the iterations? Suppose that in the jth bin, the 1σ error in the measurement of F_j is σ_j. Let us define the quantity

$$\chi^2 = \sum_j \frac{1}{\sigma_j^2} \left(f_j - f_j^r\right)^2. \tag{12}$$

We choose to stop at the *r*th iteration when the value of χ^2 drops below the number of data bins. It implies that the observed flux is consistent with being a noisy realization of the reconstructed flux.

3.2 Recovery of the matter density in redshift space

We write (2) in terms of the redshift coordinate $s \equiv x + v_p(x)$, instead of *x*. If the coordinates *x* and *s* are related by a one-to-one mapping, then the velocity and the real-space coordinate can be expressed uniquely as functions of the redshift coordinate *s*. Working with *s*, equation (2) takes the following form:

$$\tau(w) = \sigma_0 \frac{c}{H(z)} \int_{-\infty}^{\infty} n_{H_I}[s - v_p(s)] \left[1 - \frac{dv_p(s)}{ds}\right]$$
$$\times \mathcal{H}\{w - s, b[s - v_p(s)]\} \, ds \qquad (13)$$

For brevity, we have maintained the same notation for the peculiar velocity as a function of *s*. According to the continuity equation, the density in redshift space, $n_{H_I}^s$, is related to the density in real space by

$$n_{H_I}^s(s) = n_{H_I}[s - v_p(s)] \left[1 - \frac{dv_p(s)}{ds}\right]. \qquad (14)$$

Therefore (13) becomes

$$\tau(w) = \int_{-\infty}^{\infty} K(w, s) \, n_{H_I}^s(s) \, ds, \qquad (15)$$

where $K(w, s) = \mathcal{H}\{w - s, b[s - v_p(s)]\}\sigma_0 \, c/H(z)$ and the iterative scheme described in the last section can be applied to obtain the redshift-space density along the line of sight The parameter *b* depends on $(s - v_p)$ implicitly via n_{H_I}. Because *b* is a weak function of n_{H_I}, we assume that *b* is constant in the reconstruction algorithm. This will greatly simplify the application of the algorithm. Tests of the algorithm, however, are done using mock spectra generated with varying *b* according to equation (4).

To relate the H I density to the matter density we use equations (3) and (6). We still need to know the value of $\mathcal{A}$ to obtain $\rho_{DM}/\bar{\rho}_{DM}$. $\mathcal{A}$ has been estimated by adjusting the mean flux of mock spectra generated from hydrodynamical simulations to match the mean flux of observed QSO absorption spectra (Rauch et al. 1997). The accuracy and possible biases of this determination have not yet been investigated in detail (see also Weinberg et al. 1997), but we consider the value of $\mathcal{A}$ $[\propto (\Omega_{bar}h^2)^2 J^{-1} H(z)^{-1}]$ to be presently known with about 30 per cent accuracy. For most of the paper we therefore assume that we know the value of $\mathcal{A}$. In Section 4.3 we show how the value of $\mathcal{A}$ $[\propto (\Omega_{bar}h^2)^2 J^{-1} H(z)^{-1}]$ can be estimated directly from the distribution of the recovered quantity $\mathcal{A}^{1/\alpha}(\rho_{DM}/\bar{\rho}_{DM})$ by assuming a parametric form for the probability distribution of the matter density.

3.3 From redshift space to real space

The correction for redshift distortions requires knowledge of the velocity field. The three-dimensional velocity and mass density fluctuations are tightly related. However, non-linear velocity–density relations, which are easy to implement, generally relate real-space quantities, as does the relation (3) between the H I and matter density. We therefore have to resort to an iterative procedure to derive the real-space matter and H I densities from the estimated redshift-space H I density. Two issues need to be stressed here: (i) the velocity field is not uniquely specified by the line of sight density field but is influenced by the unknown three-dimensional

matter distribution, and (ii) redshift distortions can and do result in multivalued zones where regions which do not overlap in real space are mapped on to the same redshift coordinate. In appendix A we describe a method that resembles Wiener filtering and which addresses the first point. It allows us to obtain the most probable velocity field which is consistent with the estimated line of sight DM density field and has the statistical properties of a gravitationally clustering Gaussian random field of a given power spectrum. In principle the power spectrum can also be determined from the recovered density field but for simplicity we have chosen to adopt a power spectrum a priori. The results do not change much for reasonable choices of the assumed power spectrum. Multivalued zones mainly occur in regions of high density where the inversion is difficult due to saturation effects in the spectrum. We did not try to correct for this effect, but discuss some of the biases introduced by peculiar velocities in Section 4.

The iterative scheme to correct for redshift distortions that we have adopted can be summarized as follows.

(i) Assume that the redshift-space and real-space H I density are equal and use (3) to compute the matter density field along the line of sight.

(ii) Estimate the peculiar velocity along the line of sight using the method described in the Appendix.

(iii) Use the estimated peculiar velocity field to correct for redshift distortions in the density field.

(iv) Use the corrected density field to obtain a better estimate for the peculiar velocity field.

(v) Repeat steps (ii)–(iv) until convergence is achieved.

The scheme proved to be efficient; it typically converges after a few iterations. Iterative schemes of this kind have been used for correcting redshift distortions in galaxy redshift surveys (e.g. Yahil et al. 1991).

4 TESTS WITH NUMERICAL SIMULATIONS

Ideally one would like to test the method with a large high-resolution hydrodynamical simulation of gas and dark matter. However, simulations of collisionless particles are still vastly superior in dynamical range and especially speed. We have therefore chosen to test our inversion algorithm with a high-resolution *N*-body simulation of pure collisionless dark matter particles where we used the relations in Section 2.1 to obtain the H I distribution from the smoothed dark matter density field. As demonstrated e.g. by Gnedin & Hui (1998) such a procedure results in a realistic H I distribution. There are nevertheless a number of effects that are neglected in this procedure, e.g. the smoothing of fluctuations on scales smaller than the Jeans length due to pressure effects, the collisional heating of the gas in shocks and the scatter in the temperature density relation due to reionization effects. We will address this point again at the end of Section 4.3.

The simulation used was run on the Cray T3E parallel supercomputer in Garching with a modified version (MacFarland et al. 1998) of Couchman's P^3M code (Couchman, Thomas & Pearce 1995). The initial conditions were generated from the power spectrum for a standard cold dark matter (CDM) universe with $\Omega = 1$ and $H_0 = 50 \, km \, s^{-1}$ cubic box of comoving length $60 \, h^{-1}$ Mpc. The simulation was normalized such that the linear rms density fluctuation in a sphere of radius $800 \, km \, s^{-1}$ was 0.5 at redshift $z = 0$. The softening parameter was 13.2 per cent of the mean particle separation and the mesh size was $N = 512$ in one dimension.

4.1 Inverting mock spectra

We have generated mock spectra from velocity and density fields along lines of sight randomly drawn from the simulation box. We used the output of the simulation at two redshifts, $z = 3$ and $z = 0$, to investigate the effect of varying the fluctuation amplitude of the density field. We emphasize here that the mock spectra generated from the simulation at $z = 0$ are not meant to resemble absorption spectra at $z = 0$. The density and velocity fields in the simulations show significant variations in structure and amplitude at the two redshifts and the $z = 0$ spectra are investigated with the same mean flux as the $z = 3$ spectra in order to test the inversion algorithm under varying conditions.

Reliable velocity and density fields cannot be derived from the simulation on scales much smaller than the mean particle separation. We therefore applied Gaussian smoothing of widths 0.3 and $0.6\,h^{-1}$ Mpc, respectively, to the outputs of the simulation at $z = 3$ and $z = 0$. Both smoothing scales correspond to the same physical scale ($60\,\mathrm{km\,s^{-1}}$). The rms values of the smoothed density fluctuations were 0.89 and 3.52, respectively. The density of neutral hydrogen was assumed to follow the relation (3) with $\alpha = 1.7$. The absorption optical depth was computed according to equations (2), (3), and (4) with $b_0 = 30\,\mathrm{km\,s^{-1}}$. For both redshifts the value of $\mathcal{A}$ was adjusted so that the mean normalized flux of a large ensemble of mock spectra was 0.69 (about the observed mean flux at $z = 3$ as determined by Rauch et al. 1997). The spectra were convolved with a Gaussian of $8\,\mathrm{km\,s^{-1}}$ FWHM to mimic instrumental broadening and photon/read-out noise was added with a total signal-to-noise ratio S/N=50 per $3\,\mathrm{km\,s^{-1}}$ pixel.

The solid curve in the top panel of Fig. 1 is a mock spectrum generated from the simulation output at $z = 3$. The dotted curve is the same spectrum with the peculiar velocities set to zero. The peculiar motions not only introduce a systematic shift but also a narrowing of the absorption features (see also Weinberg et al. 1998 for a discussion of this point). Furthermore, unsaturated lines become deeper due to the enhanced clustering in redshift space. The mean flux is also affected. For constant $\mathcal{A}$ it increased from 0.66 to 0.69 including peculiar motions while the rms flux fluctuations increased from 0.31 to 0.35.
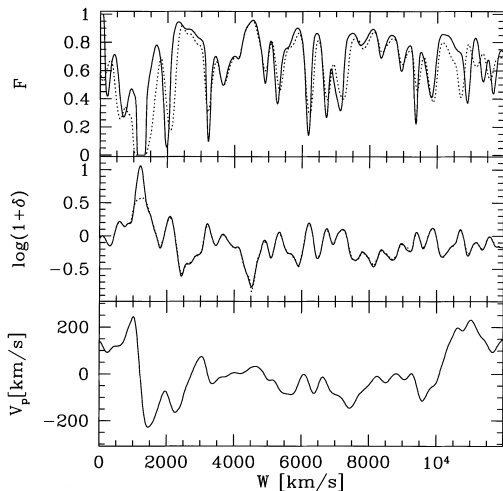

Figure 1. Normalized flux (top), density (middle) and velocity (bottom) along one line of sight through the simulation box at $z = 3$. The dotted curve in the top panel shows the flux with peculiar velocities set to zero and the dotted curve in the middle panel shows the corresponding recovered density field.
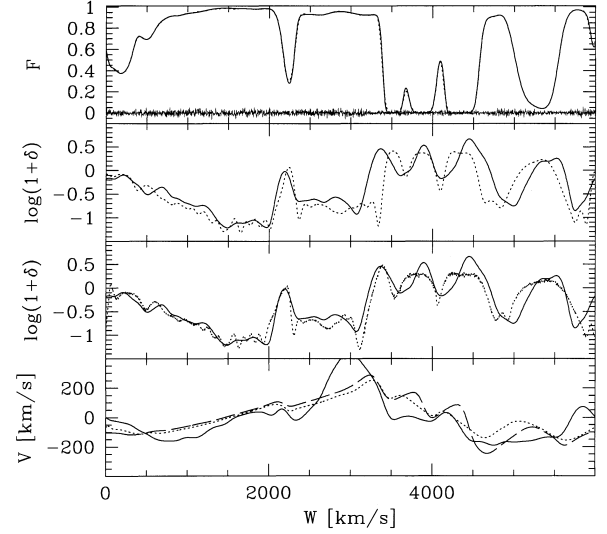

Figure 2. Top panel: normalized flux (solid curve) in a line of sight through the simulation at $z = 0$ and the flux corresponding to the density and peculiar velocity field recovered with a Lucy-type inversion algorithm (dotted curve). The noise added to the input spectrum is shown separately at the bottom of the panel. Second panel: true real-space density in the line of sight (solid curve) and recovered redshift-space matter density (dotted curve). Third panel: true real-space density (solid curve, same as in second panel) and real-space matter density recovered applying redshift corrections (dotted curve). Bottom panel: the true line-of-sight peculiar velocity field (solid curve), and the peculiar velocity estimated from the recovered density in redshift space (dotted curve). For comparison, also shown is the velocity recovered from the true real-space density (dashed line).

Fig. 2 shows some typical results of our inversion procedure. In the top panel a mock spectrum is plotted (solid line) together with the spectrum corresponding to a recovered density and peculiar velocity field (dotted line). The smoothness of the dotted curve demonstrates that our inversion procedure succeeds in mitigating the noise in the data (which for clarity is shown separately at the bottom of the panel).

Let us now have a look at the DM density field recovered from a mock spectrum generated with no peculiar velocities, as shown in the middle panel of Fig. 1. In this case the correspondence is excellent up to densities where the absorption features become heavily saturated (for absorption at an overdensity of about 3 at $z = 3$. Without peculiar velocities our inversion procedure works as well as we could reasonably expect. The obvious way to extend the inversion procedure to regions of higher density would be to use the higher order Lyman series lines, which have smaller effective absorption cross-sections and therefore saturate at increasingly higher densities (Cowie, private communication). However, as we discuss below, peculiar velocities significantly affect the quality of the recovered density field, especially in high-density regions. The incorporation of higher order Lyman series lines in the inversion procedure will therefore be difficult and we leave this to future work.

The second panel of Fig. 2 shows the DM density field recovered from a mock spectrum generated with peculiar velocities but before correcting for redshift distortions. As expected the correspondence between true and recovered density is degraded compared with the case where the spectrum was generated with the peculiar velocities set to zero. The main features of the density field are still recovered but there is a significant shift between true and recovered density in real space and there of course remains the systematic underestimate of the density in saturated regions.

The third panel shows the reconstructed density for our full iterative procedure including the correction of redshift distortions as described in Section 2.4. The shift between true and recovered density in real space is reduced and the overall correspondence has significantly improved. This indicates that we succeed in removing a major fraction of the redshift space distortion. The recovered and true line of sight velocity fields are shown in the bottom panel. In all cases shown we have used the correct value for $\mathcal{A}$, i.e. the value used to generate the mock spectrum. It should have become clear in this section that peculiar velocities significantly influence the quality of the recovery.

4.2 The density probability distribution

For a statistical analysis of the recovered density field we will use the probability distribution function in differential (PDF) and cumulative form (CPDF). Let us for the time being still assume that the true value of $\mathcal{A}$ is known. We will discuss how $\mathcal{A}$ can be estimated from the PDF of the recovered density field in the next section.

Fig. 3 shows the CPDF for the true and recovered DM density field for the simulations at $z = 3$ and $z = 0$. At low densities both curves correspond very well but the deviations become large at high densities due to saturation effects, which confirms the visual impression from Figs 1 and 2. We further quantify the differences between the CPDFs of true and recovered density field in terms of a number of moments of the density distribution in Table 1. The bias in the recovered density introduces large discrepancies between the estimated and true values of the moments. For example, the mean and rms values estimated at $z = 0$ are 0.614 and 0.608, instead of the true values of 1 and 3.52.

What we would like to have is a simple parametric form for the PDF of the DM density. This could then be used to correct for the biases in the recovered density field. It has been suggested (e.g. Kofman et. al. 1994; Coles & Jones 1991) that the PDF of the density field of a gravitationally clustering Gaussian random field in the mildly non-linear regime is well described by a lognormal PDF. In that case the quantity

$$\nu \equiv [\ln(1+\delta) - \mu_1]/\mu_2 \qquad (16)$$

has a normal (Gaussian) PDF, where μ_1 and μ_2 are the average and rms values of $\ln(1+\delta)$. In Fig. (4) we compare the PDF of the DM density in the simulations (filtered with Gaussian windows of widths $R_s = 1.2$ and 5 Mpc in comoving units for $z = 0$ and $z = 3$) with a lognormal distribution. For large smoothing scale
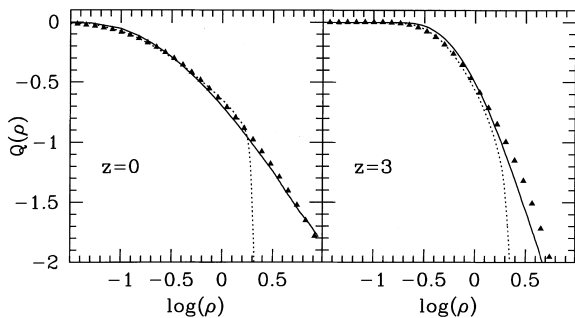

Figure 3. The cumulative probability distribution of the true density field (solid curve) and that of the density field recovered (dotted curve) from mock spectra generated from the simulation at $z = 0$ (left) and $z = 3$ (right). The triangles are the corresponding parametric form derived from an Edgeworth expansion of the PDF (23) as explained in the text.

Table 1. Moments of the density field. The symbols $\langle . \rangle$ and $|.|$ denote average and absolute values respectively. Also $x = \delta/\sigma$, where $\delta = \rho/<\rho> -1$. The *top* figures are for the true density field, the *middle* figures are for the recovered density field and the *bottom* figures are for the best-fitting Edgeworth expansion as described in the text. The moments are computed assuming the true value of $\mathcal{A}$.

	$\rho/\langle\rho\rangle$	σ	$S_3 = \langle x^3 \rangle/\sigma$	$S_{3r} = \langle x|x| \rangle$	$S_{4r} = \langle |x| \rangle$
True					
$z=3$:	1	0.89	4.85	0.66	0.61
$z=0$:	1	3.52	5.08	0.93	0.33
Rec					
$z=3$:	0.80	0.51	2.25	0.36	0.79
$z=0$:	0.61	0.61	1.97	0.40	0.83
Fit					
$z=3$:	1	1.12	4.27	0.69	0.59
$z=0$:	1	2.97	8.75	0.85	0.37

the density field is indeed adequately described by a lognormal distribution. However, for decreasing smoothing scale, the true PDF becomes more and more skewed. The skewness of ν, which we define as $\mu_3 = <\nu^3/\mu_2>$, is one way of quantifying the deviation from a lognormal distribution and is also indicated on the plot.

To obtain an improved fit to the PDFs of our simulation we use the first term of an Edgeworth expansion (Colombi 1994; Juszkiewicz et. al. 1995),

$$P(\nu) = G(\nu)\left[1 + \frac{1}{3!}\mu_3\mu_2(\nu^3 - 3\nu)\right], \qquad (17)$$

where $G(\nu)$ is a Gaussian with zero mean and unit variance. By imposing the condition $<\delta> = 0$ we find the following relation between the three parameters:

$$\mu_1 + \frac{\mu_2^2}{2} + \ln\left(1 + \frac{\mu_2^4\mu_3}{3!}\right) = 0. \qquad (18)$$

The functional form (17) therefore has two free parameters (one more than a lognormal PDF). In the remainder of this section we arbitrarily choose to work with μ_2 and μ_3. We also tried to reduce (17) to a one-parameter family. We failed, however, at establishing a tight relation between the two parameters for the PDFs of our simulations, which holds at all redshifts (see also Colombi 1994).

In practice we again find it more convenient to work with the CPDF rather than the PDF itself,

$$Q(\rho) \equiv \int_\rho^\infty P(\nu)\, d\nu$$

$$= \frac{1}{2}\text{erfc}\left(\frac{\nu}{\sqrt{2}}\right) + \frac{\mu_2\mu_3}{3!\sqrt{2\pi}}(\nu^2 - 1)\exp\left(-\frac{\nu^2}{2}\right). \qquad (19)$$

We fit the functional form (19) for the CPDF with μ_1 expressed in terms of μ_2 and μ_3 (see 18) to $Q_0(\rho)$ (the CPDF computed directly from the recovered density) by minimizing the quantity

$$R = \int_0^{\rho_{max}} d\rho \left[Q(\mu_2, \mu_3; \rho) - Q_0(\rho)\right]^2 \qquad (20)$$

with respect to μ_2 and μ_3. The cutoff ρ_{max} is introduced in order to give no weight to high-density regions where the discrepancy between the recovered and true densities is severe. Typically $\rho_{max} \approx 2$. Figs 3 and 4 show these fits as triangles. The quality of the fit is generally satisfying. At the bottom of Table 1 we list the corresponding moments. The mean is correctly recovered by
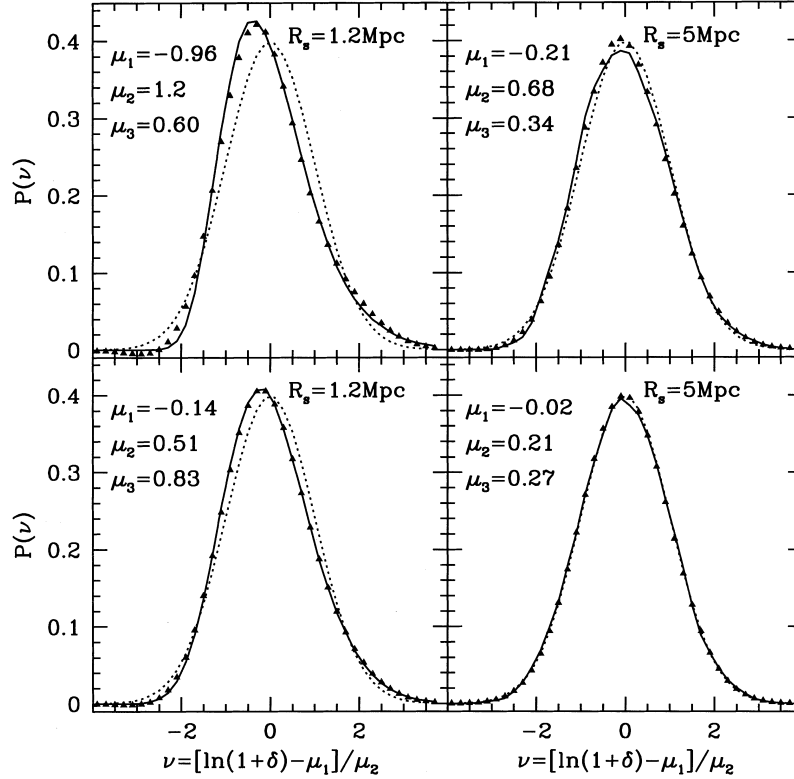
Figure 4. The probability distribution of DM density in the simulation at $z = 0$ (top panels) and $z = 3$ (bottom panels) in terms of $\nu = [\ln(1 + \delta) - \mu_1]/\mu_2$ (the density field was smoothed with Gaussian windows of width 1.2 and 5 Mpc in comoving units, respectively). The moments $\mu_1 = <x>$, $\mu_2 = <(x - \mu_1)^2>^{1/2}$ and $\mu_3 = <(x - \mu_1)^3>/\mu_2^2$, where $x = \ln(1 + \delta)$, are indicated on the plot. The dotted curves are the best-fitting log-normal distribution, while the triangles show the best-fitting Edgeworth expansion of the PDF as in equation (17).

construction but the other moments are also in significantly better agreement with the moments of the true density field than those calculated directly from the recovered density field. The rms value at $z = 0$ is now 2.97, compared with 0.62 from direct calculation and 3.52 for the true density field. When we fitted a lognormal distribution to the CPDF we found similar values for the moments to those for the Edgeworth expansion.

4.3 Determining $\mathcal{A}$ ($\propto \Omega_{bar}^2/J$)

So far we have assumed that we know the value of $\mathcal{A}$. As is clear from equations (3) and (6), assuming a wrong value of $\mathcal{A}$ leads to an estimate of $\rho_{DM}/\bar{\rho}_{DM}$ that is wrong by a constant factor $(\mathcal{A}/\mathcal{A}_{true})^{1/\alpha}$ independent of the density. It is therefore possible to estimate directly $\mathcal{A}$ by fitting one of the parametric forms discussed in the last section to the CPDF of the recovered density with $\mathcal{A}$ being left as an additional free parameter. It turned out that the Edgeworth expansion is less suitable for this procedure as is already has two free parameters (it then sometimes becomes difficult to get a unique fit). Fig. 5 shows the residuals of such a fit for the lognormal distribution as a function of $\mathcal{A}$ at the two redshifts. The residuals change significantly with varying $\mathcal{A}$ and show pronounced minima but the values of $\mathcal{A}$ at the minima are biased low by about 25 and 15 per cent, respectively. The error of $\mathcal{A}$ determined in this way is probably of the same order as that when $\mathcal{A}$ is estimated by adjusting the mean flux level in mock spectra generated from numerical simulations. We also show the case for a recovery from mock spectra with the peculiar velocities set to zero (dotted curves). The minima shift significantly. This demonstrates that the bias in $\mathcal{A}$ is
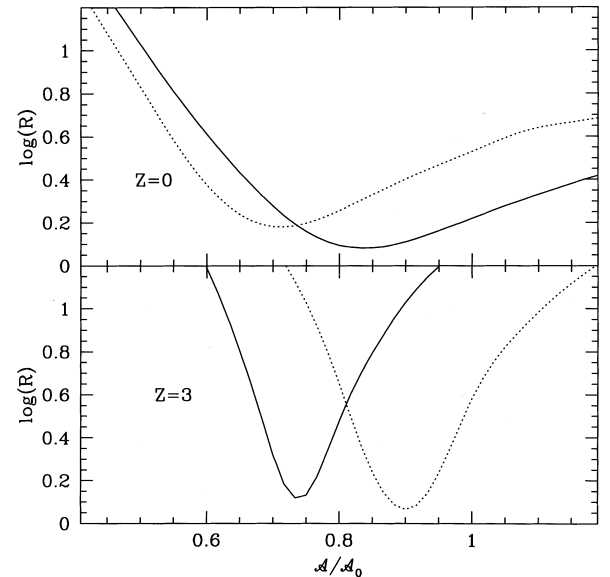

Figure 5. Residuals of the fit of a log-normal distribution to the distribution function of the recovered density as a function of the ratio of assumed to true value of $\mathcal{A}$. Solid and dotted lines refer to mock spectra generated with and without peculiar velocities, respectively.
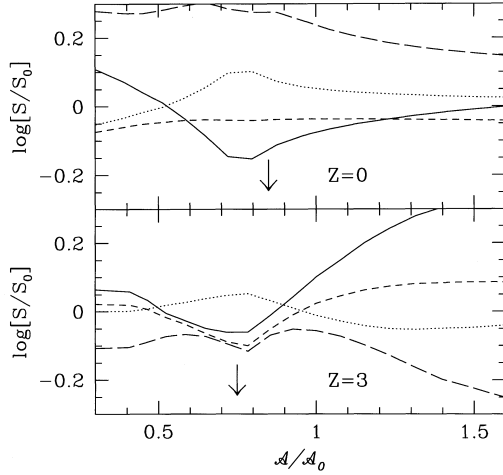
Figure 6. Moments of the density probability distribution determined by fitting an Edgeworth expansion as a function of the assumed value of $\mathcal{A}$. All values are divided by their true values. Solid, long-dashed, short-dashed and dotted lines correspond to σ, S_3, S_{3r} and S_{4r} as defined in Table 1. The arrows indicate the position of the minima in Fig. 5 for the case with peculiar velocities.

not only due to the systematic underestimate of the density in high-density regions and the inaccuracy of the lognormal distribution but also due to the effect of peculiar velocities. Peculiar velocities introduce a dependence of the bias on the amount of small-scale power in the density fluctuation spectrum, as is apparent from the comparison between the cases at $z = 3$ and $z = 0$, which differ by a factor of 4 in the rms density fluctuations. More numerical work is needed to quantify the effect of changing the cosmological model/power spectrum and the resolution, but it seems nevertheless feasible to correct for the bias in $\mathcal{A}$ to about 10 per cent accuracy.

The assumed value of $\mathcal{A}$ will obviously also affect the determination of the moments described in the last section. Fig. 6 shows the dependence of the estimated moments on the assumed value of $\mathcal{A}$. For a reasonable range of $\mathcal{A}$ the biases range from 20 to 50 per cent. If $\mathcal{A}$ is determined by fitting a lognormal CPDF and the moments are then estimated by fitting an Edgeworth expansion to the CPDF with this value of $\mathcal{A}$, the biases in the moments are generally smaller than 25 per cent.

As already mentioned, we would not expect the relations in Section 2.1 to hold perfectly. There could be considerable scatter. Furthermore we have to choose the parameters α and $\hat{T}$. We have therefore also run our inversion procedure with equations of state different from the one used to generate the mock spectra and changed the mean flux in order to test how sensitive the biases are to these parameters. The corresponding estimates of $\mathcal{A}$ and the moments are listed in Table 2. The biases seem to be robust, at least for the parameter space explored.

Table 2. Sensitivity of the estimated $\mathcal{A}$ and the moments to uncertainties in α. The bottom line shows the effect of varying the mean flux from 0.69 (as in Table 1) to 0.75 ($\alpha = 1.7$ was assumed). Results are for $z = 0$.

	$\mathcal{A}_{\min}$	μ	σ	S_3	S_{3r}	S_{4r}
$\alpha = 2.0$	0.72	1	2.00	6.35	0.82	0.46
$\alpha = 1.7$	0.84	1	2.97	8.75	0.85	0.37
$\alpha = 1.5$	1.08	1	3.70	15.64	0.82	0.31
$\langle F \rangle = 0.75$	1	0.92	2.89	8.89	0.85	0.37

5 DISCUSSION AND CONCLUSIONS

We have used numerical simulations to demonstrate that the line of sight matter density field can be recovered from QSO absorption spectra using an analytical model for the IGM and a Lucy-type iterative inversion algorithm. We have thereby estimated the line of sight peculiar velocity field from the recovered density field and have corrected for redshift distortions in an iterative procedure. The inversion works well for the most underdense regions up to the density where the absorption features saturate. For this occurs at an overdensity of a few but this limit can be pushed to higher densities by incorporating higher order lines of the Lyman series. However, an inversion will become increasingly more difficult at higher densities due to shell-crossing and shock heating of the gas.

We have then used the fact that the density probability distribution seems to have a universal shape to obtain reliable estimates of higher order moments of the PDF despite the missing high-density tail. We found that the Edgeworth expansion is an excellent approximation to the PDF of the logarithm of the density, consistent with the results of Colombi (1994). By fitting the first two terms of such an Edgeworth expansion we estimated a number of moments of the PDF with an accuracy of about 25 per cent. Estimates of similar accuracy were obtained for the parameter combination $(\Omega_{\text{bar}} h^2)^2 J^{-1} H(z)^{-1}$ by fitting to a lognormal distribution. There are a number of physical effects, such as the pressure of the gas and shock heating, that we neglected and which could decrease this accuracy. We achieved, however, similar accuracy if we recovered the density with an equation of state significantly different from that used to calculate the spectrum, indicating that this is not the case.

In principle it should also be possible to estimate the correlation function and the power spectrum directly from the recovered density field. These will, however, be affected by peculiar motions, the inability to recover high-density regions and the bias in the determination of $\mathcal{A}$. We have not yet investigated in detail how large the resulting errors are. It might well be that a comparison of observed flux power spectra with those of mock spectra as suggested by Croft et al. (1998) is the most favourable way to deal with these problems.

Observational information on the PDF of the matter density comes so far mainly from the analysis of galaxy surveys and is restricted to small redshifts. These constraints could be checked and extended to a much wider redshift range by applying our inversion technique to a moderate number of QSO absorption spectra of the kind that are now routinely taken with 10-m class telescopes. This should also check and tighten existing constraints on the evolution of the UV background. By combining our inversion technique with a redshift survey along QSO sightlines, the clustering of galaxies and matter can be related without referring to a particular cosmological model. Such a determination of the bias between galaxy and matter clustering is especially worthwhile. Most currently favoured models agree with the recently determined clustering strength of high-redshift galaxies but differ strongly in their predictions for the yet unknown bias relation (Adelberger et al. 1998). Three-dimensional information on the density and peculiar velocity field can be gained by applying the method to two or more spatially close lines of sight. An intermediate-resolution spectral survey of quasars down to about 22nd magnitude in a field of size a couple of square degree has e.g. been suggested as a possible project for the Very Large Telescope (VLT) (Petitjean 1997). Such a survey will result in about 100 lines of sight in a region about 30 Mpc across. With the method proposed in the Appendix, such a region could become a unique laboratory for the study of how galaxy formation

is related to the distribution and dynamics of the underlying matter field.

REFERENCES

Adelberger K. L., Steidel C. S., Giavalisco M., Dickinson M. E., Pettini M., Kellog M., 1998, ApJ, 505, 18
Bahcall J. N., Salpeter E. E., 1965, ApJ, 142, 1677
Bechtold J., Crotts A. P. S., Duncan R. S., Fang Y., 1994, ApJ, 437, L83
Bi H. G., 1993, ApJ, 405, 479
Bi H. G., Davidsen A. F., 1997, ApJ, 479, 523
Bi H. G., Börner G., Chu Y., 1992, A&A, 266, 1
Bond J. R., Szalay A. S., Silk J., 1988, ApJ, 324, 627
Cen R., Miralda-Escudé J., Ostriker J. P., Rauch M., 1994, ApJ, 437, L9
Coles P., Jones B. J. T., 1991, MNRAS, 248, 1
Colombi S., 1994, ApJ, 435, 536
Cooke A. J., Espey B., Carswell R. F., 1997, MNRAS, 284, 552
Couchman H. M. P., Thomas P. A., Pearce F. R., 1995, ApJ, 452, 797
Croft R. A. C., Weinberg D. H., Katz N., Hernquist L., 1998, ApJ, 495, 44
Dinshaw N., Impey C. D., Foltz C. B., Weymann R. J., Chaffee F. H., 1994, ApJ, 437, L87
Fisher K. B., Lahav O., Hoffman Y., Lynden-Bell D., Zaroubi S., 1995, MNRAS, 272, 885
Gnedin N., Hui L., 1998, MNRAS, 296, 44
Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633
Haehnelt M., Steinmetz M., 1998, MNRAS, 298, L21
Hernquist L., Katz N., Weinberg D. H., Miralda-Escudé J., 1996, ApJ, 457, L51
Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27
Hoffman Y., Ribak E., 1991, ApJ, 380, L5
Jenkins A. et al., 1998, ApJ, in press (astro-ph/9709010).
Kofman L., Bertschinger E., Gelb J., Nusser A., Dekel A., 1994, ApJ, 420, 44
Lucy L. B., 1974, AJ, 79, 745
MacFarland T. J., Couchman H. M. P., Pearce F. R., Pichlmeier J., 1998, preprint (astro-ph/9805096)
McGill C., 1990, MNRAS, 242, 544
Miralda-Escudé J., Cen R., Ostriker J. P., Rauch M., 1996, ApJ, 471, 582
Nusser A., Dekel A., Bertschinger E., Blumenthal G., 1991, ApJ, 379, 6
Petitjean P., 1997, in Bergeron J., ed., ESO Workshop, The Early Universe with the VLT. Springer p. 266
Petitjean P., Mücket J. P., Kates R. E., 1995, A&A, 295, L9
Rauch M., 1998, ARAA, in press
Rauch M. et al., 1997, ApJ, 489, 7
Sigad Y., Eldar A., Dekel A., Strauss M. A., Yahil A., 1998, ApJ, 495, 516
Wechsler R. H., Gross M. A. K., Primack J. R., Blumenthal G. R., Dekel A., 1998, ApJ, 506, 19
Weinberg D.H., 1992, MNRAS, 254, 315
Weinberg D.H., Miralda-Escudé J., Hernquist L., Katz N., 1997, ApJ, 490, 564
Weinberg D.H., Miralda-Escudé J., Hernquist L., Katz N., Miralda-Escudé J., 1998, in Petitjean P., Charlot S., eds, Structure and Evolution of the IGM from QSO absorption lines. Editions Frontieres p. 133
Wiener N., 1949, in Extrapolation and Smoothing of Stationary Time Series. Wiley, New York
Yahil A., Strauss M. A., Davis M., Huchra J. P., 1991, ApJ, 372, 380
Zaroubi S., Hoffman Y., Fisher K. B., Lahav O., 1995, ApJ, 449, 446
Zhang Y., Anninos P., Norman M. L., 1995, ApJ, 453, L57

APPENDIX A: PECULIAR VELOCITIES AND THREE-DIMENSIONAL DENSITIES FROM LINE-OF-SIGHT DENSITY FIELDS

A1 From line-of-sight density to line-of-sight peculiar velocity

In this section we describe the method that we used to estimate the line of sight peculiar velocity component from the (estimated) *real*-space density along the line of sight. We first have to assume a non-linear relation between the three-dimensional velocity and density fields. Various such relations have been suggested; here we work with the relation found empirically by Nusser et. al. (1991). For brevity of notation we present the method for $\Omega = 1$ only. Let $\theta = -\text{div } \boldsymbol{v}$, where $\boldsymbol{v}$ is the physical peculiar velocity and the divergence is with respect to the physical coordinate r in units of kms^{-1}. The approximation suggested by Nusser et al. (1991) is given by

$$\theta = \frac{\delta}{1 + 0.18\delta} + \text{constant.} \tag{A1}$$

The relation is local and provides θ in terms of the density contrast along the line of sight The constant factor is introduced to ensure $\langle \theta \rangle = 0$. Since θ involves derivatives of velocity components perpendicular to the line of sight, the line of sight peculiar velocity component cannot be uniquely determined from θ. Therefore, the best we can do is provide an estimate for the line of sight peculiar velocity based on some statistical assumptions. Neither of the fields θ or δ is Gaussian. However, for an approximate treatment we assume here that θ is Gaussian. Note that by (A1) this assumption does not imply Gaussian density fluctuations. At the end of the next subsection we will briefly describe how estimates of the peculiar velocity can be obtained without assuming that the field θ is Gaussian. Let v be the component of $\boldsymbol{v}$ along the line of sight. Subsequently we work with the (one-dimensional) Fourier tranforms, $\tilde{\theta}(q)$ and $\tilde{v}(q)$, of the line of sight fields v and θ. According to Bi (1993) $\tilde{\theta}$ and $\tilde{v}$ are related by

$$\tilde{\theta}(q) = \tilde{u}(q) + \tilde{w}(q), \tag{A2}$$

$$v(q) = iq\alpha(q)\tilde{w}(q), \tag{A3}$$

where $\tilde{w}(q)$ and $\tilde{u}(q)$ are two uncorrelated Gaussian fields with power spectra P_w and P_u. These power spectra can be written, in terms of the three-dimensional matter power spectrum P,

$$P_w(q) = 2\pi\alpha^{-1} \int_q^\infty P(k)k^{-1}\, dk, \tag{A4}$$

and

$$P_u(q) = 2\pi \int_q^\infty P(k)k\, dk - P_w(q), \tag{A5}$$

with

$$\alpha(q) = \frac{\int_q^\infty P(k)k^{-3}\,dk}{\int_q^\infty P(k)k^{-1}\,dk}. \tag{A6}$$

The relations (A2) and (A3) can be used to estimate $\tilde{v}(q)$ from $\tilde{\theta}(q)$. For simplicity, we work with $\tilde{w}$ instead of $\tilde{v}$ and switch back to $\tilde{v}$ at the end of the calculation. According to Bayes' theorem the conditional probability $P_r[w|\tilde{\theta}]$ is given by

$$P_r[w|\tilde{\theta}] = \frac{P(w)}{P(\tilde{\theta})}P(\tilde{\theta}|w). \tag{A7}$$

Using the relations (A2) and (A3) we find

$$P_r(\tilde{\theta}|w) \propto \exp\left[-\frac{(\tilde{\theta} - w)^2}{2P_u}\right], \tag{A8}$$

and

$$P_r(\tilde{w}) \propto \exp\left[-\frac{w^2}{2P_w}\right]. \tag{A9}$$

Therefore

$$P_r(\tilde{w}|\tilde{\theta}) \propto \exp\left[-\frac{(\tilde{\theta}-\tilde{w})^2}{2P_u} - \frac{\tilde{w}^2}{2P_w}\right]. \tag{A10}$$

This is a Gaussian with mean

$$<\tilde{w}|\tilde{\theta}> = \left(1+\frac{P_u}{P_w}\right)^{-1}\tilde{\theta}, \tag{A11}$$

and variance (power spectrum)

$$P_{w|\tilde{\theta}} = \frac{P_w{}^2}{P_u + P_w}. \tag{A12}$$

Thus, given $\tilde{\theta}(q)$, one may write

$$\tilde{w}(q) = \left(1+\frac{P_u}{P_w}\right)^{-1}\tilde{\theta} + n(q), \tag{A13}$$

where $n(q)$ is a random Gaussian variable with power spectrum $P_{\tilde{w}|\tilde{\theta}}$, which is uncorrelated with $\tilde{\theta}(q)$. The form of (A11) is reminiscent of Wiener filtering of noisy data (Wiener 1949, see also Zaroubi et. al. 1995 and Fisher et. al. 1995 for applications to cosmology). To see this simply identify P_u and P_w as, respectively, the signal and noise power spectra. According to (A2), the power spectrum of the unconditional $\tilde{\theta}$ is $P_\theta = P_w + P_u$. This clearly differs from the power spectrum of $<\tilde{w}|\tilde{\theta}>$ given in (A11). Following Sigad et. al. (1998), here we replace the filter $(1 + P_u/P_w)^{-1}$ in (A11) by $(1 + P_u/P_w)^{-1/2}$ in order to preserve the power spectrum of the unconditional field. Alternatively we could have complemented $<\tilde{w}|\tilde{\theta}>$ with the random field $n(q)$ in order to preserve the power.

A2 Reconstruction of three-dimensional fields

For simplicity we discuss the case of a single line of sight. The generalization to multiple parallel lines of sight is obvious. We arbitrarily choose the line of sight to be along the x axis. Once the density δ^{los} in the line of sight is given, we can easily compute the corresponding Fourier transform $\tilde{\delta}^{los}(q)$ defined by

$$\tilde{\delta}^{los}(q) = \frac{1}{(2\pi)^{1/2}} \int \delta^{los}(x) \exp(iqx)\, dx. \tag{A14}$$

We write the three-dimensional density field $\delta(r)$ in terms of its Fourier transform,

$$\delta(r) = \frac{1}{(2\pi)^{3/2}} \int \tilde{\delta}(k) \exp(-ik\cdot r)\, d^3k. \tag{A15}$$

Define $l(x) = \delta(r = x\hat{x})$ where $\hat{x}$ is a unit vector along the x-axis. Hence, by combining (A14) and (A15) we obtain

$$\tilde{\delta}^{los}(q) = \frac{1}{(2\pi)} \int \tilde{\delta}(k_\parallel, k_\perp) \exp(iqx - ik_\parallel x)\, dk_\parallel\, d^2k_\perp\, dx$$

$$= \int \tilde{\delta}(q, k_\perp)\, d^2k_\perp, \tag{A16}$$

where $k_\parallel$ and $k_\perp$ are the components parallel and perpendicular to the line of sight. The problem of reconstructing the three-dimensional density field reduces to estimating $\tilde{\delta}(k)$ given the coefficients $\tilde{\delta}^{los}(q)$ and the relation (A16). Note that $\tilde{\delta}^{los}$ can be computed from δ^{los}. The method of Hoffman and Ribak (1991) (hereafter HR) can

be used to obtain such an estimate. First, we derive an expression for the mean value $\tilde{\delta}^{MV}(k) = <\tilde{\delta}(k)|\{\tilde{\delta}^{los}(q)\}>$ of $\tilde{\delta}(k)$ corresponding to a particular k given the set of coefficients $\{\tilde{\delta}^{los}(q)\}$. According to HR, we have

$$\tilde{\delta}^{MV}(k) = \int M(k, q') N^{-1}(q, q') \tilde{\delta}^{los}(q)\, dq\, dq', \tag{A17}$$

where $M(k, q') = <\tilde{\delta}(k)\tilde{\delta}^{los}(q)>$ is the covariance matrix of $\tilde{\delta}$ and $\tilde{\delta}^{los}$, and $N^{-1}(q, q')$ is the *inverse* of the covariance matrix $N(q, q') = <\tilde{\delta}^{los}(q)\tilde{\delta}^{los}(q')>$; both matrices are of infinite dimension. Using the homogeneity condition, $<\tilde{\delta}^{los}(k)\tilde{\delta}^{los}(k')> = \delta^D(k - k')P(k)$ and the relation (A16) it can be seen that N and M are given by

$$N(q, q') = \delta^D(q - q')P^{los}(q) \tag{A18}$$

and

$$M(k, q) = \delta^D(q - k_\parallel)P\left(\sqrt{q^2 + k_\perp^2}\right), \tag{A19}$$

where δ^D is the Dirac δ-function and

$$P^{los} = 2\pi \int_q^\infty P(k)k\, dk \tag{A20}$$

is the one-dimensional power spectrum of the density field along the line of sight. The mean is given by

$$\tilde{\delta}^{MV}(k) = \frac{P(k)}{P^{los}(k_\parallel)}\, \tilde{\delta}^{los}(k_\parallel). \tag{A21}$$

Once the mean values are given, a constrained random realization $\tilde{\Delta}^C$ can be generated from an unconstrained random realization $\tilde{\Delta}$ using (HR),

$$\tilde{\Delta}^C(k) = \tilde{\Delta}(k) + \frac{P(k)}{P^{los}(k_\parallel)}\left[\tilde{\delta}^{los}(k_\parallel) - \tilde{\Delta}^{los}(k_\parallel)\right], \tag{A22}$$

where $\tilde{\Delta}^{los}(k_\parallel)$ are the one-dimensional Fourier coefficients of the unconstrained density field in the line of sight.

The treatment so far has relied on the HR method which assumes Gaussian fields. This treatment will not be satisfactory for a three-dimensional reconstruction in the non-linear regime. The unconstrained density field as given by (A22) is not guaranteed to have positive values everywhere. The scheme can be extended to the non-linear regime by two modifications: (i) use $\ln(1 + \delta)$ instead of δ, and (ii) extract an unconstrained random field $\Delta(r)$ or its Fourier transform $\tilde{\Delta}(k)$ from a fully non-linear N-body simulation. The simulations can be used to compute the one- and three-dimensional power spectra of $\ln(1 + \delta)$ that are required to generate the constrained realization. Note that these power spectra can in principle be estimated from the density field along the line of sight.

A three-dimensional constrained realization obtained with this scheme adapted to the non-linear regime can then be used to estimate the peculiar velocity without assuming θ to be Gaussian. θ can be computed at any point in space from the three-dimensional constrained realization of the density field using the velocity–density relation (A1) and assuming a potential flow. The peculiar velocity is then obtained by solving a Poisson-like equation to recover θ. The implementation of this alternative scheme for recovering the peculiar velocity field and a comparisons with the scheme described in the last section is left to future work.

This paper has been typeset from a T_EX/L^AT_EX file prepared by the author.