Mon. Not. R. Astron. Soc. **312**, 257–284 (2000)

Uncorrelated modes of the non-linear power spectrum

A. J. S. Hamilton★
JILA and Department of Astrophysical & Planetary Sciences, Box 440, University of Colorado, Boulder, CO 80309, USA

Accepted 1999 August 23. Received 1999 May 26

ABSTRACT

Non-linear evolution causes the galaxy power spectrum to become broadly correlated over different wavenumbers. It is shown that pre-whitening the power spectrum – transforming the power spectrum in such a way that the noise covariance becomes proportional to the unit matrix – greatly narrows the covariance of power. The eigenfunctions of the covariance of the pre-whitened non-linear power spectrum provide a set of almost uncorrelated non-linear modes somewhat analogous to the Fourier modes of the power spectrum itself in the linear, Gaussian regime. These almost uncorrelated modes make it possible to construct a near-minimum variance estimator and Fisher matrix of the pre-whitened non-linear power spectrum analogous to the Feldman–Kaiser–Peacock (FKP) estimator of the linear power spectrum. The paper concludes with summary recipes, in gourmet, fine and fastfood versions, of how to measure the pre-whitened non-linear power spectrum from a galaxy survey in the FKP approximation. An appendix presents FFTLog, a code for taking the fast Fourier or Hankel transform of a periodic sequence of logarithmically spaced points, which proves useful in some of the manipulations.

Key words: cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

Most of the information about cosmological parameters bottled inside current[1] and forthcoming galaxy surveys, notably the Two-Degree Field Survey (2dF) (Colless 1999; Folkes et al. 1999) and the Sloan Digital Sky Survey (SDSS) (Gunn & Weinberg 1995; Margon 1999), lies in the non-linear regime. Even in the linear regime, non-linearities perturb.

At large, linear scales, the power spectrum – the covariance of the density field, expressed in the Fourier representation – is the pre-eminent measure of large-scale structure. It is a generic, although by no means universal, prediction of inflation (Turner 1997) that linear density fluctuations should be Gaussian. More generally, primordial fluctuations should be Gaussian whenever they result from superpositions of many independent processes, thanks to the central limit theorem. Observations of large-scale structure are consistent with linear density fluctuations being Gaussian (Bouchet et al. 1993; Juszkiewicz, Bouchet, & Colombi 1993; Gaztañaga 1994; Gaztañaga & Frieman 1994; Nusser, Dekel & Yahil 1995; Stirling & Peacock 1996; Colley 1997; Chiu, Ostriker & Strauss 1998; Frieman & Gaztañaga 1999) although the evidence is not definitive (White 1999). If linear density fluctuations are Gaussian, then the three-point and higher irreducible moments are zero, so that the covariance of the density field contains complete information about the statistical properties of the field, hence all information about cosmological parameters. Compared with other measures of covariance such as the correlation function, the power spectrum has the additional advantage that estimates of power at different wavenumbers are uncorrelated, for Gaussian fluctuations. This asset of the power spectrum is intimately related to the assumption that the field is statistically translation-invariant, and to the fact that Fourier modes are eigenfunctions of the translation operator.

At smaller, non-linear scales, the power spectrum loses some of its glow. Non-linear evolution drives the density field away from Gaussianity, coupling Fourier modes, feeding higher order moments, and causing power at different wavenumbers to become correlated. The broad extent of the correlation of the non-linear power spectrum has been emphasized by Meiksin & White (1999) and Scoccimarro, Zaldarriaga & Hui (1999), and is illustrated in Fig. 2 of the present paper (later).

★ E-mail: Andrew.Hamilton@Colorado.EDU; http://casa.colorado.edu/~ajsh

[1] For a review of redshift surveys of galaxies, see Strauss (1999) and references therein. Recent surveys include: Updated Zwicky Catalog (UZC) (Falco et al. 1999); *IRAS* Point Source Catalogue Redshift Survey (PSCz) (Sutherland et al. 1999); Redshift Survey of Zwicky Catalog Galaxies in a $2^h \times 15°$ Region around 3C 273 (Grogin, Geller & Huchra 1998); Durham/UKST (Ratcliffe et al. 1998); Southern Sky Redshift Survey (SSRS) (da Costa et al. 1998); ESO Slice Project (ESP) (Vettolani et al. 1998); Muenster Redshift Project (MRSP) (Schuecker et al. 1998); CNOC2 Field Galaxy Redshift Survey (Carlberg et al. 1999); Century Survey (Geller et al. 1997); Norris Survey of the Corona Borealis Supercluster (Small, Sargent & Hamilton 1997); Stromlo–APM (Loveday et al. 1996); Las Campanas Redshift Survey (LCRS) (Shectman et al. 1996); Hawaii Deep Fields (Cowie et al. 1996); Canada–France Redshift Survey (CFRS) (Lilly et al. 1995).

The purpose of the present paper is to show how to unfold the non-linear power spectrum into a set of nearly uncorrelated modes, somewhat analogous to the Fourier modes of the power spectrum itself in the linear, Gaussian regime. The present paper is a natural successor to Hamilton (1997a,b, hereafter Papers I and II), which showed how to derive the minimum variance estimator and Fisher matrix of the power spectrum of a galaxy survey in the Feldman, Kaiser & Peacock (1994, hereafter FKP) approximation, for Gaussian fluctuations. Section 5.2 of Paper I posed, but was unable to solve, the non-Gaussian problem solved in the present paper. A following paper (Hamilton & Tegmark 2000, hereafter Paper IV), describes how to complete the processing of the power spectrum into fully decorrelated band-powers.

It turns out that a key to solving the non-Gaussian problem is to 'pre-whiten' the power spectrum – to transform the non-linear power spectrum in such a way that the (two-point) shot-noise contribution to the covariance matrix is proportional to the unit matrix. The properties of the pre-whitened non-linear power spectrum appear empirically to be sweeter than might reasonably have been expected.

This paper is devoted entirely to the problem of non-linearity. It ignores the equally important problem of redshift distortions (Hamilton 1998), and the problematic question of light-to-mass bias (Coles 1993; Fry & Gaztañaga 1993; Mo, Jing & White 1997; Mann, Peacock & Heavens 1998; Tegmark & Peebles 1998; Moscardini et al. 1998; Scherrer & Weinberg 1998; Dekel & Lahav 1999; Colín et al. 1999; Cen & Ostriker 1999; Narayanan, Berlind & Weinberg 1999; Blanton et al. 1999; Benson et al. 1999; Bernardeau & Schaeffer 1999; Coles, Melott & Munshi 1999).It further assumes that uncertainties arising either from the selection function (Binggeli, Sandage & Tammann 1988; Willmer 1997; Tresse 1999) or from evolution in the cosmological volume element or the galaxy population are negligible.

Several authors have recently published estimates of how well measurements of the power spectrum from future galaxy surveys will constrain cosmological parameters (Tegmark 1997b; Goldberg & Strauss 1998; Hu, Eisenstein & Tegmark 1998; Eisenstein, Hu & Tegmark 1998, 1999). The procedures described in the present paper should assist this enterprise.

The aims of the present paper are complementary to those of Bond, Jaffe & Knox (1998b). The question that Bond et al. considered was: if the power spectrum (of the cosmic microwave background, specifically) is quadratically compressed (Tegmark 1997a; Tegmark, Taylor & Heavens 1997; Tegmark et al. 1998) into a set of band-powers, then what is the best way to use those band-powers in maximum likelihood estimation of parameters? For example, one general procedure is to use not the band-powers themselves, but rather functions of the band-powers arranged such that their variances remain constant as the prior power is varied. Bond et al. argued that the likelihood function is then more nearly Gaussian. The purpose of this paper and Paper IV is rather to arrive at the point where one has decorrelated band-powers to work with in the first place.

The plan of this paper is as follows. Section 2 sets up the notation and defines reference material needed in subsequent sections. Section 3 goes through the difficulties one meets in attempting to measure the non-linear power spectrum in minimum variance fashion, and describes how to overcome them. Section 4 reveals the unexpectedly nice properties of the pre-whitened covariance of the power spectrum, key to the whole enterprise of this paper. Section 5 defines the pre-whitened power spectrum. Sections 6 and 7 show how the approximations motivated in previous sections lead to a practical way to evaluate the Fisher matrix of the pre-whitened non-linear power, and to measure the pre-whitened non-linear power spectrum from a galaxy survey. Section 8 discusses how to evaluate the Fisher matrix and non-linear power spectrum using the FKP approximation alone, without any additional approximation. Section 9 summarizes the results of previous sections into recipes, in gourmet, fine and fastfood versions, for measuring non-linear power, the end product being a set of uncorrelated pre-whitened non-linear band-powers, with error bars, over some prescribed grid of wavenumbers. Section 10 summarizes the conclusions. Appendix B gives details of FFTLOG, a code for taking the fast Fourier or Hankel transform of a periodic sequence of logarithmically spaced points.

2 PRELIMINARIES

This section contains reference material needed in subsequent sections. The reader interested in new results may like to skip to the next section, Section 3, referring back to the present section as needed.

2.1 Data and parameters

'He will, of course, use maximum likelihood because his textbooks have told him that' – E. T. Jaynes (1996, p. 624).

According to Bayes' theorem, the probability distribution of parameters θ_α given data δ_i is, up to a normalization factor, the product of the prior probability with the likelihood function $\mathcal{L}(\delta_i|\theta_\alpha)$. The data δ_i in a galaxy survey can be taken to be overdensities $\delta(\boldsymbol{r})$ at positions $\boldsymbol{r}$ in the survey

$$\delta(\boldsymbol{r}) \equiv \frac{n(\boldsymbol{r}) - \bar{n}(\boldsymbol{r})}{\bar{n}(\boldsymbol{r})}, \tag{1}$$

where $n(\boldsymbol{r})$ is the observed number density of galaxies, and $\bar{n}(\boldsymbol{r})$ is the selection function. The parameters θ_α are, for the present purpose, some parametrization of the galaxy power spectrum; the focus of this paper is on the case in which the parameters are the power spectrum ξ_α itself.

This paper conforms to the common convention used by cosmologists to relate the power spectrum $\xi(k)$ in Fourier space to the correlation function $\xi(r)$ in real space, notwithstanding the extraneous factors of 2π that result:

$$\xi(k) = \int \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}} \xi(r)\, \mathrm{d}^3 r = \int_0^\infty j_0(kr)\xi(r) 4\pi r^2\, \mathrm{d}r, \tag{2}$$

$$\xi(r) = \int \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}} \xi(k) \frac{\mathrm{d}^3 k}{(2\pi)^3} = \int_0^\infty j_0(kr)\xi(k) \frac{4\pi k^2\, \mathrm{d}k}{(2\pi)^3}, \tag{3}$$

where $j_0(x) = \sin x/x$ is a spherical Bessel function.

2.2 Hilbert space

As in Paper I, it is convenient to adopt a notation in which Latin indices i, j, …, refer to three-dimensional positions, while Greek indices α, β, …, run over the space of parameters, and more specifically over the one-dimensional space of wavenumbers or pair separations.

For generality, brevity and ease of manipulation, it is convenient to treat quantities such as the data vector δ_i, or the power spectrum ξ_α, as vectors in a Hilbert space (for a didactic exposition, see Hamilton 1998, section 3.3). Such vectors have a meaning

independent of the particular basis, i.e. complete set of linearly independent functions, with respect to which they might be expressed. For example, the data vector has components $\delta_{\boldsymbol{r}}[=\delta(\boldsymbol{r})]$ when expressed in real space, or components $\delta_{\boldsymbol{k}}[=\delta(\boldsymbol{k})=\int \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}\delta(\boldsymbol{r})\,\mathrm{d}^3r]$ when expressed in Fourier space, but from a Hilbert-space point of view these are the same vector, and in this paper they are both denoted by the same symbol δ_i.

Similarly the power spectrum ξ_α has components $\xi_k[=\xi(k)]$ when expressed in Fourier space, or $\xi_r[=\xi(r)]$ when expressed in real space, but again from a Hilbert-space point of view these are the same vector, and in this paper they are both denoted by the same symbol ξ_α.

Latin indices $i, j, \ldots$, on vectors and matrices run over the three-dimensional space of positions $\boldsymbol{r}$, or more generally over any three-dimensional basis of the Hilbert space. Unless stated otherwise, repeated pairs of indices signify the inner product in Hilbert space, as in

$$a^i b_i = \int a^*(\boldsymbol{r})b(\boldsymbol{r})\,\mathrm{d}^3r = \int a^*(\boldsymbol{k})b(\boldsymbol{k})\,\mathrm{d}^3k/(2\pi)^3. \tag{4}$$

By definition, the inner product is a scalar, the same quantity independent of the choice of basis. The raised index a^i denotes the Hermitian conjugate (if the basis is orthonormal) of the vector a_i. One of the indices in an inner product is always raised, the other lowered. In this paper, all vectors in the Hilbert space are real-valued when expressed in real space, so that $a^*(\boldsymbol{r}) = a(\boldsymbol{r})$ and $a^*(\boldsymbol{k}) = a(-\boldsymbol{k})$.

Adhering to the raised/lowered index convention serves as a useful reminder that one of the pair of vectors in an inner product is a Hermitian conjugate (if the basis is orthonormal). In Fourier space, for example, this means using $-\boldsymbol{k}$ for one index (raised) and $+\boldsymbol{k}$ for the other index (lowered) of an inner product.

Greek indices $\alpha, \beta, \ldots$, run over the space of one-dimensional pair separations r, or wavenumbers k, or more generally over any one-dimensional basis in the associated Hilbert space. Again, unless stated otherwise, repeated indices signify the inner product

$$a^\alpha b_\alpha = \int a^*(r)b(r)4\pi r^2\,\mathrm{d}r = \int a^*(k)b(k)4\pi k^2\,\mathrm{d}k/(2\pi)^3 \tag{5}$$

which is again a scalar, the same quantity independent of the choice of basis. Again, in this paper all vectors in the Hilbert space are real-valued in real space, so $a^*(r) = a(r)$ and $a^*(k) = a(k)$. Although there is no distinction in this case between vectors with raised and lowered indices in either real or Fourier space, adhering to the raised/lowered index convention again serves as a useful reminder.

The unit matrix $\mathbf{1}_\alpha^\beta$ in any representation is defined such that its inner product with any vector a_β leaves the vector unchanged,

$$\mathbf{1}_\alpha^\beta a_\beta = a_\beta \mathbf{1}_\alpha^\beta = a_\alpha. \tag{6}$$

In the continuous real representation, the unit matrix is

$$\mathbf{1}_{r_\alpha}^{r_\beta} = \delta_{3\mathrm{D}}(r_\alpha - r_\beta), \tag{7}$$

where $\delta_{3\mathrm{D}}(r_\alpha - r_\beta)$ denotes the three-dimensional Dirac delta-function, defined such that

$$\int \delta_{3\mathrm{D}}(r_\alpha - r_\beta)4\pi r_\alpha^2\,\mathrm{d}r_\alpha = 1. \tag{8}$$

In the continuous Fourier representation, the unit matrix is

$$\mathbf{1}_{k_\alpha}^{k_\beta} = (2\pi)^3\delta_{3\mathrm{D}}(k_\alpha - k_\beta), \tag{9}$$

again a three-dimensional Dirac delta-function.

2.3 Discretization of matrices

Many of the operations in this paper involve manipulations of matrices in the one-dimensional space of separations. Continuous matrices must be discretized to manipulate them numerically. Discretization should be done in such a way as to preserve the inner product (5), so that integration over the volume element, $4\pi r^2\,\mathrm{d}r$ in real space, or $4\pi k^2\,\mathrm{d}k/(2\pi)^3$ in Fourier space, translates into summation in the corresponding discrete space. This ensures that matrix operations such multiplication, diagonalization and inversion can be done in the usual fashion.

Most of the manipulations in this paper are done in Fourier space on a logarithmically spaced grid of wavenumbers k_α. In this case, a continuous vector $a(k_\alpha)$ is discretized by multiplying it by $[4\pi k_\alpha^3\Delta\ln k/(2\pi)^3]^{1/2}$:

$$a(k_\alpha) \rightarrow \mathsf{a}_{k_\alpha} = a(k_\alpha)[4\pi k_\alpha^3\Delta\ln k/(2\pi)^3]^{1/2}; \tag{10}$$

and a continuous matrix $A(k_\alpha, k_\beta)$ is discretized by multiplying it by $4\pi(k_\alpha k_\beta)^{3/2}\Delta\ln k/(2\pi)^3$:

$$A(k_\alpha, k_\beta) \rightarrow \mathbf{A}_{k_\alpha k_\beta} = A(k_\alpha, k_\beta)4\pi(k_\alpha k_\beta)^{3/2}\Delta\ln k/(2\pi)^3. \tag{11}$$

The unit matrix $(2\pi)^3\delta_{3\mathrm{D}}(k_\alpha - k_\beta)$ in the continuous Fourier representation translates to the unit matrix $\mathbf{1}_{\alpha\beta}$ in the discrete case:

$$(2\pi)^3\delta_{3\mathrm{D}}(k_\alpha - k_\beta) \rightarrow \mathbf{1}_{\alpha\beta}. \tag{12}$$

Similarly, a continuous vector $a(r_\alpha)$ in real space is discretized on to a logarithmically spaced grid of separations r_α by multiplying the vector by $[4\pi r_\alpha^3\Delta\ln r]^{1/2}$:

$$a(r_\alpha) \rightarrow \mathsf{a}_{r_\alpha} = a(r_\alpha)[4\pi r_\alpha^3\Delta\ln r]^{1/2}; \tag{13}$$

and a continuous matrix $A(r_\alpha,r_\beta)$ is discretized by multiplying it by $4\pi(r_\alpha r_\beta)^{3/2}\Delta\ln r$:

$$A(r_\alpha, r_\beta) \rightarrow \mathbf{A}_{r_\alpha r_\beta} = A(r_\alpha, r_\beta)4\pi(r_\alpha r_\beta)^{3/2}\Delta\ln r. \tag{14}$$

The unit matrix $\delta_{3\mathrm{D}}(r_\alpha - r_\beta)$ in the continuous real representation translates to the unit matrix $\mathbf{1}_{\alpha\beta}$ in the discrete case:

$$\delta_{3\mathrm{D}}(r_\alpha - r_\beta) \rightarrow \mathbf{1}_{\alpha\beta}. \tag{15}$$

The transformation between Fourier and real space for logarithmically spaced wavenumbers k_α and separations r_α may be accomplished with FFTLog (Appendix B).

2.4 Gaussian density field

If the density distribution $\delta(\boldsymbol{r})$ were Gaussian – which is *not* true in the present case – then one would have the luxury of being able to write down an explicit Gaussian likelihood function

$$\mathcal{L} \propto \frac{1}{|C|^{1/2}}\exp\left(-\frac{1}{2}\delta_i C^{-1ij}\delta_j\right), \tag{16}$$

where $|C|$ and C^{-1} are the determinant and inverse of the covariance matrix C of overdensities

$$C_{ij} \equiv \langle\delta_i\delta_j\rangle. \tag{17}$$

Angle-brackets here and throughout this paper signify averages over possible data sets δ_i predicted by the likelihood function

$$\langle t\rangle \equiv \int t\mathcal{L}\,(\delta_i|\theta_\alpha)\,\mathrm{d}[\delta_i]. \tag{18}$$

Maximum likelihood (ML) estimates $\hat{\theta}_\alpha$ of the parameters θ_α (the hat distinguishing the estimate $\hat{\theta}_\alpha$ from the true value θ_α) are given by the vanishing of the vector of partial derivatives of the log-likelihood function

$$\frac{\partial \ln \mathcal{L}}{\partial \theta_\alpha} = \frac{1}{2}\frac{\partial C_{ij}}{\partial \theta_\alpha} C^{-1ik} C^{-1jl} (\delta_k \delta_l - C_{kl}), \tag{19}$$

$$\left.\frac{\partial \ln \mathcal{L}}{\partial \theta_\alpha}\right|_{\theta_\alpha = \hat{\theta}_\alpha} = 0. \tag{20}$$

The covariance $\langle \Delta\hat{\theta}_\alpha \Delta\hat{\theta}_\beta \rangle$ of the estimated parameters is given approximately by the inverse of the Fisher information matrix $F^{\alpha\beta}$, defined to be minus the expectation value of the matrix of second partial derivatives of the log-likelihood function:

$$F^{\alpha\beta} \equiv -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial \theta_\alpha \partial \theta_\beta} \right\rangle = \frac{1}{2}\frac{\partial C_{ij}}{\partial \theta_\alpha} C^{-1ik} C^{-1jl} \frac{\partial C_{kl}}{\partial \theta_\beta}, \tag{21}$$

$$\langle \Delta\hat{\theta}_\alpha \Delta\hat{\theta}_\beta \rangle \approx F^{-1}_{\alpha\beta}. \tag{22}$$

The approximation (22) is exact if the estimated parameters $\hat{\theta}_\alpha$ are Gaussianly distributed about their expectation values. The central limit theorem asserts that the parameters become Gaussianly distributed in the asymptotic limit of a large amount of data.

It is commonly assumed, and the same assumption is adopted here, that the dominant source of variance in a galaxy survey is a combination of cosmic (sample) variance and shot-noise arising from the discrete sampling of galaxies. If the sampling of galaxies is random – a Poisson process – then the covariance C_{ij} is a sum of the cosmic covariance ξ_{ij} with Poisson sampling noise N_{ij}:

$$C_{ij} = \xi_{ij} + N_{ij}. \tag{23}$$

In the real representation, the cosmic covariance ξ_{ij} is the correlation function

$$\xi_{ij} = \xi(|\boldsymbol{r}_i - \boldsymbol{r}_j|), \tag{24}$$

and the noise matrix N_{ij} is the diagonal matrix

$$N_{ij} = \frac{\delta_{3D}(\boldsymbol{r}_i - \boldsymbol{r}_j)}{\bar{n}(\boldsymbol{r}_i)}, \tag{25}$$

with $\delta_{3D}(\boldsymbol{r}_i - \boldsymbol{r}_j)$ a three-dimensional Dirac delta-function. In the Fourier representation the cosmic covariance ξ_{ij} is the diagonal matrix

$$\xi_{ij} = (2\pi)^3 \delta_{3D}(\boldsymbol{k}_i + \boldsymbol{k}_j) \xi(k_i), \tag{26}$$

the eigenvalues $\xi(k_i)$ of which constitute the power spectrum.

The focus of this paper is on the case in which the parameters θ_α are the power spectrum ξ_α itself [in this paper the cosmic covariance function ξ_α, expressed in an arbitrary representation, will often be referred to as the 'power spectrum', even though this name is commonly reserved for the covariance $\xi(k)$ expressed in Fourier space; no confusion should result]. In this case the covariance C_{ij} is a linear function of the parameters ξ_α:

$$C_{ij} = D^\alpha_{ij} \xi_\alpha + N_{ij}, \tag{27}$$

where in real space $\xi_\alpha = \xi(r_\alpha)$ is the correlation function, and

$$D^\alpha_{ij} = \delta_{3D}(|\boldsymbol{r}_i - \boldsymbol{r}_j| - r_\alpha) \tag{28}$$

is a three-dimensional Dirac delta-function, equation (15), while in Fourier space $\xi_\alpha = \xi(k_\alpha)$ is the thing commonly called the power spectrum, and

$$D^\alpha_{ij} = (2\pi)^6 \delta_{3D}(\boldsymbol{k}_i + \boldsymbol{k}_j) \delta_{3D}(k_i - k_\alpha). \tag{29}$$

It follows from equations (19) and (20) that the ML estimator $\hat{\xi}_\alpha$ of the power spectrum, for Gaussian fluctuations, is that solution of

$$\hat{\xi}_\alpha = \frac{1}{2} F^{-1}_{\alpha\beta} D^\beta_{ij} C^{-1ik} C^{-1jl} (\delta_k \delta_l - N_{kl}) \tag{30}$$

for which the estimate is equal to the prior, $\hat{\xi}_\alpha = \xi_\alpha$. The variance of the ML estimator is

$$\langle \Delta\hat{\xi}_\alpha \Delta\hat{\xi}_\beta \rangle \approx F^{-1}_{\alpha\beta} \tag{31}$$

and the Fisher matrix is

$$F^{\alpha\beta} = \frac{1}{2} D^\alpha_{ij} C^{-1ik} C^{-1jl} D^\beta_{kl}. \tag{32}$$

If the prior power ξ_α is regarded as fixed, then equation (30) yields an estimated power $\hat{\xi}_\alpha$ that is quadratic in overdensities δ_i. If this estimated power is folded back into the prior, then equation (30) with the revised prior yields another estimate of power. Iterated to convergence, the result is the ML estimator of the power. It is to be noted that, even without iteration, equation (30) yields a measurement of power that (as long as the prior is at least roughly correct) should already be a good approximation, since 'if the prior matters, then you are not learning much from the data', to quote one of the refrains from the 1997 Aspen workshop on Precision Measurement of Large Scale Structure.

The question of how to apply quadratic estimators [such as given by equation (30)] to measure the power spectrum is addressed by Tegmark et al. (1998) for galaxies, and by Tegmark (1997a), Tegmark et al. (1997) and Bond, Jaffe & Knox (1998a,b) for the cosmic microwave background.

2.5 Non-Gaussian density field

Ultimately, one might look forward to a wondrous N-body machine able to compute the probability distribution of linear initial conditions given noisy and incomplete data from a survey (Narayanan & Weinberg 1998; Monaco & Efstathiou 1999; and references therein).

In the meantime, it is far from clear what to write down as a likelihood function for the non-linear density field (Dodelson, Hui & Jaffe 1999). Certainly it would be a bad idea to use a Gaussian likelihood function for a non-Gaussian density field, since that would lead to a serious underestimate of the true uncertainty in the measured non-linear power spectrum.

An alternative procedure is to seek a minimum variance unbiased estimator of power. Now the power spectrum is by definition a covariance of overdensities, and, by the presumption of Poisson sampling, any a priori weighted sum of quantities quadratic in observed overdensities (with self-terms excluded, to eliminate shot-noise) provides an unbiased estimate of the power spectrum linearly windowed in some fashion. It was shown in section 2.3 of Paper I that, amongst estimators quadratic in observed overdensities δ_i, the unbiased estimator $\hat{\xi}_\alpha$ of the power spectrum having minimum variance is

$$\hat{\xi}_\alpha = F^{-1}_{\alpha\beta} D^\beta_{ij} \mathfrak{C}^{-1ijkl} (\delta_k \delta_l - \hat{N}_{kl}) \tag{33}$$

with variance

$$\langle \Delta\hat{\xi}_\alpha \Delta\hat{\xi}_\beta \rangle = F^{-1}_{\alpha\beta}, \tag{34}$$

where $F^{\alpha\beta}$ is the Fisher matrix

$$F^{\alpha\beta} = D^\alpha_{ij} \mathfrak{C}^{-1ijkl} D^\beta_{kl}, \tag{35}$$

$\mathfrak{C}_{ijkl}$ is the covariance of shot-noise-subtracted products of overdensities

$$\mathfrak{C}_{ijkl} \equiv \langle(\delta_i\delta_j - \hat{N}_{ij} - \xi_{ij})(\delta_k\delta_l - \hat{N}_{kl} - \xi_{kl})\rangle \tag{36}$$

and $\mathfrak{C}^{-1ijkl}$ is its inverse, meaning $\mathfrak{C}_{ijmn}\mathfrak{C}^{-1mnkl} = \mathrm{Sym}_{(kl)}\mathbf{1}_i^k\mathbf{1}_j^l$. The symbol $\mathrm{Sym}_{(ij)}$ signifies symmetrization over its underscripts, as in

$$\underset{(ij)}{\mathrm{Sym}}\, A_{ij} \equiv \frac{1}{2}(A_{ij} + A_{ji}). \tag{37}$$

The quantity $\hat{N}_{kl}$ in equations (33) and (36) is the 'actual' shot-noise, the contribution to $\delta_k\delta_l$ from self-pairs of galaxies, pairs consisting of a galaxy and itself. The actual shot-noise $\hat{N}_{kl}$ in a survey is to be distinguished from its expectation value $N_{kl} \equiv \langle\hat{N}_{kl}\rangle$. If the expected shot-noise N_{kl} is used in equation (33) in place of the actual shot-noise, then additional terms [given in equation (8) of Paper I] appear in the covariance matrix $\mathfrak{C}_{ijkl}$, increasing the variance of the estimator. Why does the ML estimator $\hat{\xi}_\alpha$ in the Gaussian case, equation (30), involve the expected shot-noise N_{kl} rather than the actual shot-noise $\hat{N}_{kl}$? Because a discretely sampled Gaussian field is not really Gaussian, except in the limit where a cubic wavelength contains many galaxies, so the assumption of a Gaussian likelihood function is not strictly correct. In fact it is plain that the Gaussian ML estimator $\hat{\xi}_\alpha$ would also be improved if the actual shot-noise $\hat{N}_{kl}$ were used in place of the expected shot-noise N_{kl} in equation (30), since using the actual shot-noise exploits additional information about the character of the Poisson sampling that is discarded by the Gaussian likelihood. However, as discussed by Tegmark et al. (1998, appendix A), the gain from subtracting the actual versus the expected shot-noise is in practice small at linear scales, where a cubic wavelength is likely to contain many galaxies.

In the same Poisson sampling approximation as equation (23), the covariance $\mathfrak{C}_{ijkl}$ of shot-noise-subtracted products of overdensities, equation (36), is, in the real representation with no implicit summation,

$$\begin{aligned}\mathfrak{C}_{ijkl} = {} & \xi_{ik}\xi_{jl} + \xi_{il}\xi_{jk} + \eta_{ijkl} \\ & + [N_{ik}(\xi_{jl} + \zeta_{ijl}) + (i \leftrightarrow j, k \leftrightarrow l)](4 \text{ terms}) \\ & + (N_{ik}N_{jl} + N_{il}N_{jk})(1 + \xi_{ij}),\end{aligned} \tag{38}$$

in which the top line is the four-point, the middle the three-point, and the bottom line the two-point contribution to the covariance, as illustrated in Fig. 1. For Gaussian density fluctuations equation (38) reduces to

$$\mathfrak{C}_{ijkl} = 2\,\underset{(kl)}{\mathrm{Sym}}\, C_{ik}C_{jl} \tag{39}$$

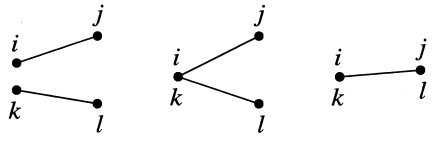

Figure 1. Schematic illustration of the four-point, three-point and two-point contributions to the covariance $\mathfrak{C}_{ijkl}$ of pairs ij with other pairs kl. The three-point and two-point contributions are shot-noise contributions in which one or both galaxies of the pair ij are the same as one or both of the pair kl.

with inverse

$$\mathfrak{C}^{-1ijkl} = \frac{1}{2}\,\underset{(kl)}{\mathrm{Sym}}\, C^{-1ik}C^{-1jl}. \tag{40}$$

It follows from equation (40) that for Gaussian fluctuations the minimum variance estimator of the power spectrum, equation (33), is the same as the ML estimator, equation (30), if the estimate is folded back into the prior and iterated to convergence (modulo the comments about shot-noise in the previous paragraph).

3 PROBLEMS

3.1 FKP approximation

Calculating the minimum variance estimate $\hat{\xi}_\alpha$ of the power spectrum, equation (33), involves the formidable problem of inverting the pair covariance $\mathfrak{C}_{ijkl}$, a rank 4 matrix of three-dimensional quantities. Whereas for Gaussian fluctuations the rank 4 matrix $\mathfrak{C}_{ijkl}$ factorizes into a product of rank 2 matrices, equation (39), for non-Gaussian fluctuations it does not factorize. Again, whereas for Gaussian fluctuations it may be possible, at least at the largest scales, to pixellize a survey into large enough pixels that brute force numerical inversion is feasible, for non-Gaussian fluctuations brute force inversion is quite impossible.

A natural way to simplify the problem is to adopt the Feldman, Kaiser & Peacock (1994, FKP) approximation, where the selection function $\bar{n}(\boldsymbol{r})$ is taken to be locally constant. The FKP approximation is expected to be valid at wavelengths much smaller than the characteristic size of the survey. Section 5 of Paper I terms this the 'classical' approximation, since it is valid to the extent that the position and wavelength of a density mode can be measured simultaneously. While the FKP approximation is liable to break down at larger scales, particularly for pencil beam or slice surveys, it should be a good approximation at smaller, non-linear scales, especially in surveys with broad contiguous sky coverage.

Even if the selection function $\bar{n}$ is taken to be constant, the general problem of inverting the rank 4 matrix $\mathfrak{C}_{ijkl}$ remains intractable. Notice, however, that $\mathfrak{C}^{-1ijkl}$ appears multiplied in both equations (33) and (35) by the matrix D_{ij}^α. Now D_{ij}^α has translation and rotation symmetry, and in the FKP approximation the matrix $\mathfrak{C}_{ijkl}$ also has translation and rotation symmetry, the selection function $\bar{n}$ being constant. Indeed, inspection of equation (38) reveals that the matrix $\mathfrak{C}_{ijkl}$ remains translation- and rotation-invariant even if the selection functions $\bar{n}_i$ and $\bar{n}_j$ at positions i and j are two different constants. It follows that the combination $\mathfrak{C}_{ijkl}D_\alpha^{kl}$ is likewise translation- and rotation-symmetric, which implies that it can be expressed in the form

$$\mathfrak{C}_{ijkl}D_\alpha^{kl} = \mathfrak{C}_{\alpha\beta}(\bar{n}_i, \bar{n}_j)D_{ij}^\beta \tag{41}$$

for some matrix $\mathfrak{C}_{\alpha\beta}$, which can be termed the 'reduced' covariance matrix. Equation (41) is the FKP approximation, expressed in concise mathematical form; additional details of the justification of this equation are provided in Appendix A. The reduced matrix is written in equation (41) as $\mathfrak{C}_{\alpha\beta}(\bar{n}_i, \bar{n}_j)$ to emphasize the fact that it is a function of the selection functions $\bar{n}_i$ and $\bar{n}_j$ at positions i and j; note that no implicit summation over i or j is intended on the right-hand side of equation (41). Inspection of equation (38) for $\mathfrak{C}_{ijkl}$ shows that the reduced covariance $\mathfrak{C}_{\alpha\beta}(\bar{n}_i, \bar{n}_j)$ takes the form

$$\mathfrak{C}_{\alpha\beta}(\bar{n}_i, \bar{n}_j) = 2[K_{\alpha\beta} + (\bar{n}_i^{-1} + \bar{n}_j^{-1})J_{\alpha\beta} + \bar{n}_i^{-1}\bar{n}_j^{-1}H_{\alpha\beta}], \tag{42}$$

a linear combination of four-point, three-point and two-point contributions $K_{\alpha\beta}$, $J_{\alpha\beta}$ and $H_{\alpha\beta}$. Multiplying equation (41) by $\mathfrak{C}^{-1\gamma\alpha}\mathfrak{C}^{-1mnij}$ shows that the inverse of $\mathfrak{C}_{ijkl}$ is similarly related to the inverse of the reduced matrix $\mathfrak{C}_{\alpha\beta}$:

$$\mathfrak{C}^{-1ijkl}D^{\alpha}_{kl} = \mathfrak{C}^{-1\alpha\beta}(\bar{n}_i,\bar{n}_j)D^{ij}_{\beta}. \tag{43}$$

Physically, to the extent that the selection functions $\bar{n}_i$ and $\bar{n}_j$ at positions i and j are constants, the minimum variance pair-weighting attached to a pair ij should be a function only of the separation α of the pair, not of their position or orientation. Just as $\mathfrak{C}_{ijkl}$ is the covariance between a pair ij and another pair kl, so the reduced covariance matrix $\mathfrak{C}_{\alpha\beta}$ is the covariance between a pair separated by α and another pair separated by β.

In the FKP approximation given by equation (43), the minimum variance estimate (33) of the power spectrum is

$$\hat{\xi}_{\alpha} = F^{-1}_{\alpha\beta}\mathfrak{C}^{-1\beta\gamma}D^{ij}_{\gamma}(\delta_i\delta_j - \hat{N}_{ij}) \tag{44}$$

and the associated Fisher matrix (35) is

$$F^{\alpha\beta} = \mathfrak{C}^{-1\alpha\gamma}D^{ij}_{\gamma}D^{\beta}_{ij}. \tag{45}$$

Notice that the approximate Fisher matrix given by this equation (45) is not symmetric, whereas the original Fisher matrix, equation (35), was symmetric. The asymmetry results from the asymmetry of the FKP approximation, equation (41). The approximate expression (45) would be symmetric if the FKP approximation were exact, and in practice it should be nearly symmetric; if not, it is a signal that the FKP approximation is breaking down.

To ensure symmetry of the Fisher matrix, one might be inclined at this point to symmetrize equation (45), since after all an equally good approximation to the Fisher matrix would be the same expression (45) with the indices swapped on the right-hand side, $\alpha \leftrightarrow \beta$. However, it is desirable that the FKP estimator $\hat{\xi}_{\alpha}$, equation (44), should be unbiased, meaning that

$$\langle\hat{\xi}_{\alpha}\rangle = \xi_{\alpha}. \tag{46}$$

Averaging equation (44) gives, since $\langle\delta_i\delta_j - \hat{N}_{ij}\rangle = D^{\alpha}_{ij}\xi_{\alpha}$ according to equation (27),

$$\langle\hat{\xi}_{\alpha}\rangle = F^{-1}_{\alpha\beta}\mathfrak{C}^{-1\beta\gamma}D^{ij}_{\gamma}D^{\epsilon}_{ij}\xi_{\epsilon}, \tag{47}$$

which shows that the FKP estimator $\hat{\xi}_{\alpha}$ is unbiased only if the Fisher matrix in equation (44) is interpreted as satisfying the asymmetric expression (45). A detailed discussion of this issue is deferred to Section 7. Here it suffices to remark that, to the extent that the FKP approximation is valid, the variance of the FKP estimator $\hat{\xi}_{\alpha}$ is equal to the inverse of the symmetrized Fisher matrix given by equation (45):

$$\langle\Delta\hat{\xi}_{\alpha}\Delta\hat{\xi}_{\beta}\rangle = F^{-1}_{(\alpha\beta)}, \tag{48}$$

where $F^{(\alpha\beta)} \equiv \mathrm{Sym}_{(\alpha\beta)}F^{\alpha\beta}$ denotes the symmetrized Fisher matrix, and $F^{-1}_{(\alpha\beta)}$ its inverse.

3.2 Hierarchical model

The pair covariance matrix $\mathfrak{C}_{ijkl}$, equation (38), hence also the reduced covariance matrix $\mathfrak{C}_{\alpha\beta}$, equation (41), involves the three-point and four-point correlation functions ζ_{ijk} and η_{ijkl}. The problem here is that these correlation functions are not known precisely.

Available observational and N-body evidence [see, for example, the summaries by Scoccimarro & Frieman (1999) and Hui & Gaztañaga (1999)] is consistent with a hierarchical model in which the three-point and four-point functions are, in the real representation with no implicit summation,

$$\zeta_{ijk} = Q(\xi_{ij}\xi_{jk} + \xi_{jk}\xi_{ki} + \xi_{ki}\xi_{ij}), \tag{49}$$

$$\eta_{ijkl} = R_a[\xi_{ij}\xi_{jk}\xi_{kl} + \text{cyclic (12 snake terms)}] + R_b[\xi_{ij}\xi_{ik}\xi_{il} + \text{cyclic (4 star terms)}], \tag{50}$$

with approximately constant hierarchical amplitudes Q, R_a and R_b. On the other hand, it is clear that the hierarchical amplitudes do vary at some level, as a function of both scale and configuration shape.

In the translinear regime, perturbation theory predicts that the hierarchical amplitudes should vary (somewhat) with both scale and configuration, for density fluctuations growing by gravity from Gaussian initial conditions (Fry 1984; Scoccimarro et al. 1998).

In the deeply non-linear regime, predictions for the behaviour of the hierarchical amplitudes are more empirical. Scoccimarro & Frieman (1999) have recently suggested an ansatz, which they dub hyperextended perturbation theory (HEPT), that the hierarchical amplitudes in the highly non-linear regime go over to the values predicted by perturbation theory for configurations collinear in Fourier space. For power-law power spectra $\xi(k) \propto k^n$, HEPT predicts a three-point amplitude

$$Q = \frac{4 - 2^n}{1 + 2\ 2^n}, \tag{51}$$

and four-point amplitudes $R_a = R_b = Q_4$ with

$$Q_4 = \frac{54 - 27\ 2^n + 2\ 3^n + 6^n}{2(1 + 6\ 2^n + 3\ 3^n + 6\ 6^n)}. \tag{52}$$

For simplicity, the present paper adopts the hierarchical model, with constant hierarchical amplitudes set equal to the HEPT values (51) and (52). For reasons to be discussed shortly (namely that the Schwarz inequality is violated otherwise), most of the calculations shown take

$$R_a = -R_b = Q_4, \tag{53}$$

although where possible results are also shown for

$$R_a = R_b = Q_4. \tag{54}$$

In addition to power-law power spectra, the present paper shows results for the power spectrum derived from observations by Peacock (1997), and for an observationally concordant ΛCDM model from the fitting formulae of Eisenstein & Hu (1998), non-linearly evolved according to the procedure of Peacock & Dodds (1996). In these cases the adopted amplitudes are those corresponding to $n = -1.2$, i.e. a correlation function with slope $\gamma = n + 3 = 1.8$, for which $Q = 1.906$ and $Q_4 = 4.195$.

In the hierarchical model with constant hierarchical amplitudes, the four-point, three-point and two-point contributions to the reduced covariance matrix $\mathfrak{C}_{\alpha\beta}$, equation (42), are, in the Fourier representation with no implicit summation,

$$K(k_\alpha,k_\beta) = (2\pi)^3\delta_{3\mathrm{D}}(k_\alpha - k_\beta)\xi(k_\alpha)^2 + R_a[\xi(k_\alpha) + \xi(k_\beta)]^2A(k_\alpha,k_\beta) + R_b\xi(k_\alpha)\xi(k_\beta)[\xi(k_\alpha) + \xi(k_\beta)], \tag{55}$$

$$J(k_\alpha, k_\beta) = (2\pi)^3 \delta_{3D}(k_\alpha - k_\beta)\xi(k_\alpha) + Q[\xi(k_\alpha) + \xi(k_\beta)]A(k_\alpha, k_\beta) + Q\xi(k_\alpha)\xi(k_\beta), \quad (56)$$

$$I(k_\alpha, k_\beta) = (2\pi)^3 \delta_{3D}(k_\alpha - k_\beta) + A(k_\alpha, k_\beta), \quad (57)$$

where in the real-space representation the matrix $A_{\alpha\beta}$ is the diagonal matrix

$$A(r_\alpha, r_\beta) = \delta_{3D}(r_\alpha - r_\beta)\xi(r_\alpha), \quad (58)$$

while in the Fourier representation $A_{\alpha\beta}$ is

$$A(k_\alpha, k_\beta) = \frac{1}{2k_\alpha k_\beta} \int_{|k_\alpha - k_\beta|}^{k_\alpha + k_\beta} \xi(k) k \, \mathrm{d}k. \quad (59)$$

Convergence of $A(k_\alpha, k_\beta)$ at $k_\alpha = k_\beta$ requires that $\xi(k) \sim k^n$ with $n > -2$ at small wavenumber k. Convergence of $\int \xi(r) 4\pi r^2 \, \mathrm{d}r$ at small r requires that $\xi(r) \sim r^{-\gamma}$ with $\gamma < 3$ at small separation r. Thus for power-law power spectra $\xi(k) \propto k^n$ (this is the evolved, non-linear power spectrum, not the original, linear power spectrum), equivalent to power-law correlation functions $\xi(r) \propto r^{-\gamma}$ with $\gamma = n + 3$, the hierarchical model is consistent only for

$$-2 < n < 0 \quad \text{or equivalently} \quad 1 < \gamma < 3. \quad (60)$$

It is straightforward to determine that, for power-law power spectra $\xi(k) \propto k^n$ in the hierarchical limit (where the Gaussian contribution becomes negligible), the correlation coefficient of the four-point contribution $K_{\alpha\beta}$ to the reduced covariance $\mathfrak{C}_{\alpha\beta}$ is, for $k_\alpha \gg k_\beta$,

$$\frac{K_{\alpha\beta}}{(K_{\alpha\alpha} K_{\beta\beta})^{1/2}} \rightarrow \frac{(R_a + R_b)}{2\left(\frac{2^{n+2}}{n+2} R_a + R_b\right)} \left(\frac{k_\beta}{k_\alpha}\right)^{n/2}, \quad (61)$$

which diverges as $k_\alpha / k_\beta \rightarrow \infty$ (for $-2 < n < 0$) unless $R_b = -R_a$. Thus the Schwarz inequality, which requires that the absolute value of the correlation coefficient be less than or equal to unity, is violated unless $R_b = -R_a$. This problem has been remarked upon and discussed by Scoccimarro et al. (1999, section 3.3). Scoccimarro et al. show from N-body simulations that the traditional relation $R_a \approx R_b$ holds approximately for $k_\alpha \approx k_\beta$, but that indeed $R_a + R_b$ decreases systematically as k_α and k_β become more and more separated. Scoccimarro et al. conclude that the simple hierarchical model with constant amplitudes is not a good description of the four-point function in the highly non-linear regime.

For simplicity, the present paper adopts the hierarchical model with constant amplitudes, and either $R_b = -R_a$ or $R_b = R_a$. Ultimately, the latter choice leads to unphysically huge variances, plainly a consequence of the violation of the Schwarz inequality. Thus the canonical models in this paper have $R_b = -R_a$. However, where possible, intermediate results are also shown for $R_b = R_a$.

3.3 Pre-whitening

The minimum variance estimator $\hat{\bar{\xi}}_\alpha$ and associated Fisher matrix $F^{\alpha\beta}$, equations (44) and (45), involve six-dimensional integrals of $\mathfrak{C}^{-1\alpha\beta}(\bar{n}_i, \bar{n}_j)$ over all pairs ij of volume elements in a survey. This is actually quite a feasible numerical problem. The reduced covariance matrix $\mathfrak{C}_{\alpha\beta}(\bar{n}_i, \bar{n}_j)$ is a rank 2 matrix of one-dimensional quantities, so is straightforward to invert numerically for any particular values of the selection functions $\bar{n}_i$ and $\bar{n}_j$. If, as is typical, the selection function separates into the product of an angular mask and a radial selection function, then the angular integrals can be done analytically (Hamilton 1993), leaving a double integral of $\mathfrak{C}^{-1\alpha\beta}(\bar{n}_i, \bar{n}_j)$ over the radial directions, which is doable. This direct procedure is discussed further in Section 8, and forms the basis of the gourmet recipe summarized in Section 9.1. Still, the integration is burdensome, and it is enlightening to explore whether further simplification is possible.

Ideally what one would like is that there would exist a representation in which $\mathfrak{C}_{\alpha\beta}(\bar{n}_i, \bar{n}_j)$ were simultaneously diagonal for arbitrary values of the selection function $\bar{n}$. Precisely this situation obtains in the case of Gaussian fluctuations, for which the reduced covariance matrix $\mathfrak{C}_{\alpha\beta}$ is diagonal in Fourier space:

$$\mathfrak{C}_{\alpha\beta}(\bar{n}_i, \bar{n}_j) = 2(2\pi)^3 \delta_{3D}(k_\alpha - k_\beta)[\xi(k_\alpha) + \bar{n}_i^{-1}][\xi(k_\alpha) + \bar{n}_j^{-1}], \quad (62)$$

regardless of the values $\bar{n}_i$ and $\bar{n}_j$ of the selection function.

For non-Gaussian fluctuations, the reduced covariance $\mathfrak{C}_{\alpha\beta}(\bar{n}_i, \bar{n}_j)$ is a linear combination of four-point, three-point and two-point matrices $K_{\alpha\beta}$, $J_{\alpha\beta}$ and $H_{\alpha\beta}$, according to equation (42). Finding a representation in which $\mathfrak{C}_{\alpha\beta}(\bar{n}_i, \bar{n}_j)$ is diagonal for any $\bar{n}_i$ and $\bar{n}_j$ thus means diagonalizing the three matrices K, J and H simultaneously. This is of course generically impossible.

However, it is possible to diagonalize two (K and H) of the three matrices simultaneously by the trick of pre-whitening, and to cross one's fingers on the third matrix (J). The term pre-whitening refers to the operation of multiplying a signal by a function in such a way that the noise becomes white, or constant (Blackman & Tukey 1959, section 11). Pre-whitening is commonly used in the construction of Karhunen–Loève modes (signal-to-noise ratio eigenmodes), in order to allow a signal and its noise to be diagonalized simultaneously (Vogeley & Szalay 1996; Tegmark et al. 1997, 1998).

Define the pre-whitened reduced covariance $\mathfrak{B}_{\alpha\beta}$ to be

$$\mathfrak{B} \equiv H^{-1/2} \mathfrak{C} H^{-1/2}, \quad (63)$$

and similarly define the pre-whitened three-point and four-point matrices $M_{\alpha\beta}$ and $L_{\alpha\beta}$ to be

$$M \equiv H^{-1/2} K H^{-1/2}, \quad (64)$$

$$L \equiv H^{-1/2} J H^{-1/2}. \quad (65)$$

By construction, the pre-whitened two-point matrix is the unit matrix, $H^{-1/2} H H^{-1/2} = \mathbf{1}$. In terms of the pre-whitened four-point and three-point matrices $M_{\alpha\beta}$ and $L_{\alpha\beta}$, the pre-whitened reduced covariance $\mathfrak{B}_{\alpha\beta}(\bar{n}_i, \bar{n}_j)$ is (compare equation 42)

$$\mathfrak{B}_{\alpha\beta}(\bar{n}_i, \bar{n}_j) = 2[M_{\alpha\beta} + (\bar{n}_i^{-1} + \bar{n}_j^{-1}) L_{\alpha\beta} + \bar{n}_i^{-1} \bar{n}_j^{-1} \mathbf{1}_{\alpha\beta}]. \quad (66)$$

The properties of the pre-whitened four-point and four-point matrices M and L are examined in Section 4.

3.4 FFTLog

Several of the manipulations described in this paper involve transforming between real and Fourier space. Ideally, one would like to be able to cover several orders of magnitude in separation or wavenumber. The SDSS, for example, should be able to probe scales from 10^{-2} to $10^3 h^{-1}$ Mpc, a range of 10^5. If the Fourier transforms were done using standard fast Fourier transform (FFT) techniques, which require lineary spaced points, covering such a range would require 10^5 points. The trouble with this is that one

would then have to manipulate $10^5 \times 10^5$ matrices. Clearly this is a problem of the shoe not fitting the foot – that is, a linear spacing of points is not well suited to the case at hand: while the difference between separations of 0.01 and $0.02\,h^{-1}$ Mpc may be significant, the difference between 1000.01 and $1000.02\,h^{-1}$ Mpc is practically irrelevant.

The problem may be solved by using an FFT method originally proposed by Talman (1978) that works for logarithmically spaced points, and which I have implemented in a code FFTLog. FFTLog is analogous to the normal FFT in that it gives the exact Fourier transform of a discrete sequence that is uniformly spaced and periodic in logarithmic space. More generally, FFTLog yields fast Hankel (=Fourier–Bessel) transforms of arbitrary order, including both integral and 1/2-integral orders. FFTLog, like the normal FFT, suffers from the usual problems of ringing (response to sudden steps) and aliasing (periodic folding of frequencies), but, under appropriate circumstances and with suitable precautions, discussed in Appendix B, it yields reliable Fourier transforms covering ranges of many orders of magnitude with modest numbers of points.

Appendix B gives further details of FFTLog. The code may be downloaded from http://casa.colorado.edu/~ajsh/FFTLog/.

4 PRE-WHITENED FOUR-POINT AND THREE-POINT COVARIANCE MATRICES

4.1 Computation

Before showing pictures, it is helpful to comment on the numerical computation of the four-point and three-point covariance matrices $K_{\alpha\beta}$ and $J_{\alpha\beta}$ and their pre-whitened counterparts $M_{\alpha\beta}$ and $L_{\alpha\beta}$.

Equations (55) and (56) give expressions for the four-point and three-point matrices $K(k_\alpha, k_\beta)$ and $J(k_\alpha, k_\beta)$ in Fourier space, for the hierarchical model with constant hierarchical amplitudes. These are discretized as described in Section 2.3. An issue here is the calculation of the subsidiary matrix $A(k_\alpha, k_\beta)$. This matrix $A_{\alpha\beta}$ is diagonal in real space with diagonal entries $\bar{\xi}(r_\alpha)$, equation (58), so one way to calculate $A(k_\alpha, k_\beta)$ is to start with the diagonal matrix $A(r_\alpha, r_\beta)$ in real space, and then Fourier transform it into Fourier space. Unfortunately, the resulting Fourier transformed matrix $A(k_\alpha, k_\beta)$ shows evident signs of ringing and aliasing, which is true whether the wavenumbers k_α are linearly spaced (FFT) or logarithmically spaced (FFTLog). Part of the difficulty is that the diagonal matrix $A(r_\alpha, r_\beta)$ is liable to vary by several orders of magnitude along the diagonal; since the FFT (or FFTLog) assumes that the matrix is periodic, the matrix appears to have a sharp step at its boundary. These problems can be reduced by padding the matrix, and in the case of FFTLog by biasing the matrix with a suitable power law (see Appendix B). Still, artefacts from the FFT remain a concern.

A more robust procedure, the one used in this paper, is to avoid FFTs altogether, and to calculate the matrix $A(k_\alpha, k_\beta)$ directly from its Fourier expression (59).

A similar issue arises when pre-whitening the four-point and three-point matrices K and J. The pre-whitening matrix $H^{-1/2} = (\mathbf{1} + A)^{-1/2}$ is again diagonal in real space, with diagonal entries $[1 + \bar{\xi}(r)]^{-1/2}$. Thus one way to pre-whiten K (say) is to start with $K(k_\alpha, k_\beta)$ in Fourier space, Fourier transform it into real space, pre-whiten $M(r_\alpha, r_\beta) = [1 + \bar{\xi}(r_\alpha)]^{-1/2} K(r_\alpha, r_\beta)[1 + \bar{\xi}(r_\beta)]^{-1/2}$, and then Fourier transform back into Fourier space. Once again the resulting matrix $M(k_\alpha, k_\beta)$ shows signs of ringing and aliasing.

Again, a more robust procedure, the one used in this paper, is to avoid FFTs, and to calculate the pre-whitening matrix $H^{-1/2} = (\mathbf{1} + A)^{-1/2}$ directly in Fourier space. Specifically, take the Fourier expression (59) for $A(k_\alpha, k_\beta)$, add the unit matrix $\mathbf{1}$ to form H, and evaluate the inverse positive square root $H^{-1/2}$ via an intermediate diagonalization. This yields the pre-whitening matrix $H^{-1/2}$ in Fourier space, which can be used directly to pre-whiten the four-point and three-point covariance matrices K and J in Fourier space. This manner of constructing $H^{-1/2}$ guarantees that the pre-whitened two-point covariance matrix $H^{-1/2} H H^{-1/2}$ is numerically equal to the unit matrix $\mathbf{1}$, as it should be. Although this procedure is slower than using FFTs, it yields results that are robust with respect to range, resolution and linear or logarithmic binning, and consistent with the results from FFTs if due care is taken with the latter.

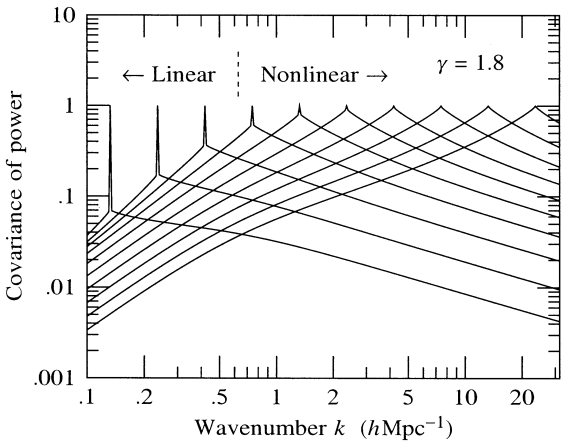

Figure 2. Correlation coefficient $K(k_\alpha, k_\beta)/[K(k_\alpha, k_\alpha)K(k_\beta, k_\beta)]^{1/2}$ of the four-point contribution $K(k_\alpha, k_\beta)$ to the covariance of the power (i.e. the covariance without shot-noise) in the case of a power-law power spectrum with correlation function $\bar{\xi}(r) = (r/5\,h^{-1}\,\mathrm{Mpc})^{-1.8}$. Each line is the correlation coefficient for a fixed k_β, and each line peaks at $k_\alpha = k_\beta$, whereat the value is unity. The hierarchical amplitudes are $R_a = -R_b = 4.195$. The resolution is 128 points per decade, $\Delta \log k = 1/128$.

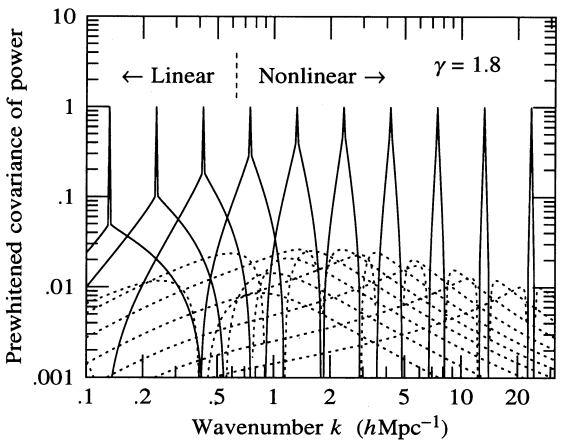

Figure 3. Correlation coefficient $M(k_\alpha, k_\beta)/[M(k_\alpha, k_\alpha)M(k_\beta, k_\beta)]^{1/2}$ of the four-point contribution $M(k_\alpha, k_\beta)$ to the pre-whitened covariance of a power-law power spectrum with correlation function $\bar{\xi}(r) = (r/5\,h^{-1}\,\mathrm{Mpc})^{-1.8}$. Lines are dotted where the correlation coefficient is negative. This is the same as Fig. 2, except that the covariance is pre-whitened.

4.2 Pre-whitened four-point covariance matrix

Fig. 2 shows the correlation coefficient $K_{\alpha\beta}/(K_{\alpha\alpha}K_{\beta\beta})^{1/2}$ (no implicit summation) of the four-point contribution $K_{\alpha\beta}$ to the (unpre-whitened) reduced covariance matrix $\mathfrak{C}_{\alpha\beta}$, equation (42), for the case of a power-law power spectrum having correlation function $\xi(r) = (r/5\,h^{-1}\,\mathrm{Mpc})^{-1.8}$. Physically, the quantity plotted is the (correlation coefficient of the) covariance of estimates of power in the case of a perfect survey with no shot-noise, $\bar{n} \to \infty$.

The correlation coefficient offers a good way to visualize the covariance, since a value of $(-)1$ means that two quantities are perfectly (anti-)correlated, and the Schwarz inequality requires that the absolute value of the correlation coefficient always be less than or equal to unity.

The Gaussian spikes evident in the curves on the leftward, linear, side of Fig. 2 reflect the fact that the covariance of power becomes diagonal in the linear, Gaussian regime. In the non-linear regime, the hierarchical contribution to the covariance dominates, and the covariance of power becomes quite broad, a point previously made by Meiksin & White (1999) and Scoccimarro et al. (1999).

It should be borne in mind that the shape of the correlation coefficient shown in Fig. 2 depends on the resolution in wavenumber k, a point emphasized by Scoccimarro et al. (1999). In Fig. 2 the points are logarithmically spaced with 128 points per decade, so $\Delta \log k = 1/128$. However, the correlation coefficient varies in an unsurprising way: as the resolution increases, the Gaussian spikes gets spikier, tending in principle to a Dirac delta-function in the limit of infinite resolution.

Fig. 3 shows the correlation coefficient $M_{\alpha\beta}/(M_{\alpha\alpha}M_{\beta\beta})^{1/2}$ of the four-point contribution $M_{\alpha\beta}$ to the pre-whitened reduced covariance $\mathfrak{B}_{\alpha\beta}$, equations (63) and (66), again for the case of a power-law power spectrum having correlation function $\xi(r) = (r/5\,h^{-1}\,\mathrm{Mpc})^{-1.8}$. The only difference between this figure and Fig. 2 is that the covariance is now pre-whitened.

The pre-whitened covariance M plotted in Fig. 3 appears to be remarkably narrow, certainly substantially narrower than the covariance shown in Fig. 2. The Gaussian spikes again show up in the linear regime, and again the hierarchical contribution to the pre-whitened covariance dominates in the non-linear regime. The hierarchical contribution appears empirically to have a constant width of $\Delta k \approx \pi/r_0 \approx 1\,h\,\mathrm{Mpc}^{-1}$, where $r_0 = 5\,h^{-1}\,\mathrm{Mpc}$ is the correlation length. Thus the pre-whitened covariance appears to become relatively narrower at large wavenumber k.

Fig. 4 shows the correlation coefficients of the covariance of the power, both straight K and pre-whitened M, for several other power spectra. In each case the covariance of power with the power at $k = 1\,h\,\mathrm{Mpc}^{-1}$ is plotted, which is essentially the 'worst case', where the pre-whitened covariance M is relatively broadest.

The solid lines in Fig. 4 are for four-point hierarchical amplitudes $R_b = -R_a$, while the dashed lines are for $R_b = R_a$. As discussed in Section 3.2, the hierarchical model violates the Schwarz inequality at $k_\alpha \gg k_\beta$ (or $k_\alpha \ll k_\beta$) unless $R_b = -R_a$.

Fig. 4 illustrates that the pattern encountered in Figs 2 and 3 is remarkably robust over different power spectra. That is, while the covariance of the power is itself broad, in all cases the covariance of the pre-whitened power is substantially narrower, at least for $R_b = -R_a$ (solid lines). Note that the power-law power spectra illustrated in Fig. 4 cover essentially the full range of indices, $1 < \gamma < 3$, allowed by the hierarchical model.

The situation for $R_b = R_a$ is muddier. Although the core of the pre-whitened covariance is for the most part reasonably narrow also in this case, the off-diagonal covariances at $k_\alpha \gg k_\beta$ (or $k_\alpha \ll k_\beta$) are starting to become worrying large in several cases. Some of this behaviour is undoubtedly inherited from the unphysical (Schwarz-inequality-violating) behaviour of the ordinary covariance, and is surely not realistic. Here I leave the problem with the comment that further investigation is clearly required, along the lines being pioneered by Scoccimarro et al. (1999).

4.3 Pre-whitened three-point covariance matrix

As discussed in Section 3.3, it would be ideal if the pre-whitened three-point contribution $L_{\alpha\beta}$ to the covariance of power were diagonal in the same representation as the four-point contribution $M_{\alpha\beta}$.

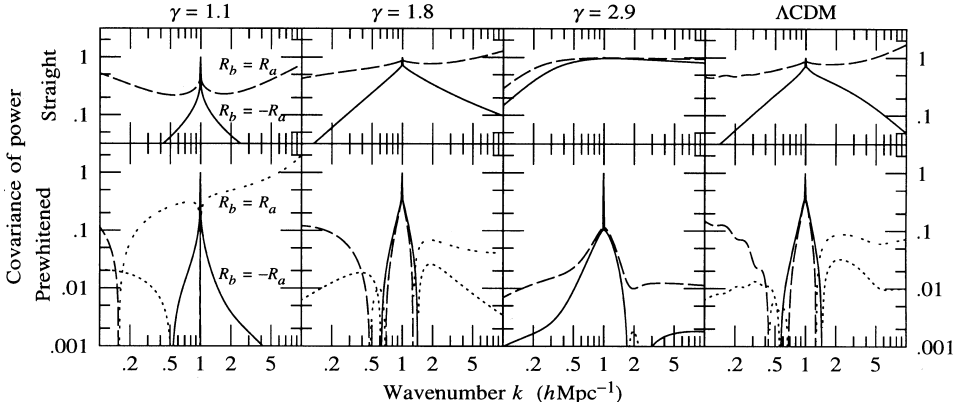

Figure 4. Correlation coefficients (top) $K(k_\alpha, k_\beta)/[K(k_\alpha, k_\alpha)K(k_\beta, k_\beta)]^{1/2}$ of the covariance, and (bottom) $M(k_\alpha, k_\beta)/[M(k_\alpha, k_\alpha)M(k_\beta, k_\beta)]^{1/2}$ of the pre-whitened covariance, of the power spectrum in four different models of the power spectrum. Each curve is the correlation coefficient at fixed $k_\beta = 1\,h\,\mathrm{Mpc}^{-1}$, plotted as a function of k_α. The three sets of panels starting from the left are for power-law power spectra with correlation functions $\xi(r) = (r/5\,h^{-1}\,\mathrm{Mpc})^{-\gamma}$ with indices $\gamma = 1.1$, 1.8 and 2.9, while the rightmost panel is for the ΛCDM power spectrum of Eisenstein & Hu (1998) with $\Omega_\Lambda = 0.7$, $\Omega_\mathrm{m} = 0.3$, $\Omega_\mathrm{b}h^2 = 0.02$ and $h = 0.65$, non-linearly evolved by the procedure of Peacock & Dodds (1996). The two lines on each graph are for four-point hierarchical amplitudes (solid) $R_b = -R_a$, and (long-dashed) $R_b = R_a$. Lines are dotted where the correlation coefficient is negative. The Schwarz inequality, which requires that the correlation coefficient be ≤ 1, is violated by the hierarchical model with $R_b = R_a$ at values $k \ll k'$ and $k \gg k'$. The resolution is 128 points per decade, the same as in Figs 2 and 3.

Fig. 5 shows the correlation coefficient of the pre-whitened three-point covariance $L_{\alpha\beta}$ in the representation of eigenfunctions of the pre-whitened four-point covariance $M_{\alpha\beta}$, for the case of $\xi(r) = (r/5\,h^{-1}\,\mathrm{Mpc})^{-1.8}$. The horizontal axis here is a nominal wavenumber k_α labelling each eigenfunction $\phi_\alpha(k)$ of the pre-whitened four-point covariance $M_{\alpha\beta}$. In practice, the eigenfunctions are simply ordered by eigenvalue, which in most cases (see below) yields a satisfactory ordering by wavenumber, in the sense that the corresponding eigenfunctions $\phi_\alpha(k)$ have their largest components around $k \approx k_\alpha$.

At first sight, the correlation coefficient plotted in Fig. 5 looks astonishingly diagonal at all wavenumbers, for both $R_b = -R_a$ and $R_b = R_a$. However, as Scoccimarro et al. (1999) emphasize, off-diagonal elements, although they may be small, are many. The resolution in Fig. 5 is 128 points per decade, and the off-diagonal elements in the case $R_b = -R_a$ are down at the level of $\sim 1/100$, which means that cumulative off-diagonal covariance over a decade of wavenumber would be comparable to the diagonal variance. Curiously, the off-diagonal elements are somewhat smaller for $R_b = R_a$ than for $R_b = -R_a$.

In Section 6 and thereafter the approximation will be made, equation (80), that the three-point matrix L is indeed diagonal in the representation of four-point eigenfunctions. If L is not precisely diagonal, then the 'minimum variance' pair-weighting that emerges from assuming diagonality will not be precisely minimum variance. However, a linear error in the pair-weighting will raise the variance quadratically from its minimum, so the pair-weighting should be close to minimum variance as long as L is not too far from being diagonal. In any case, as discussed in Section 7.1, the estimate of power remains unbiased whatever approximations are made.

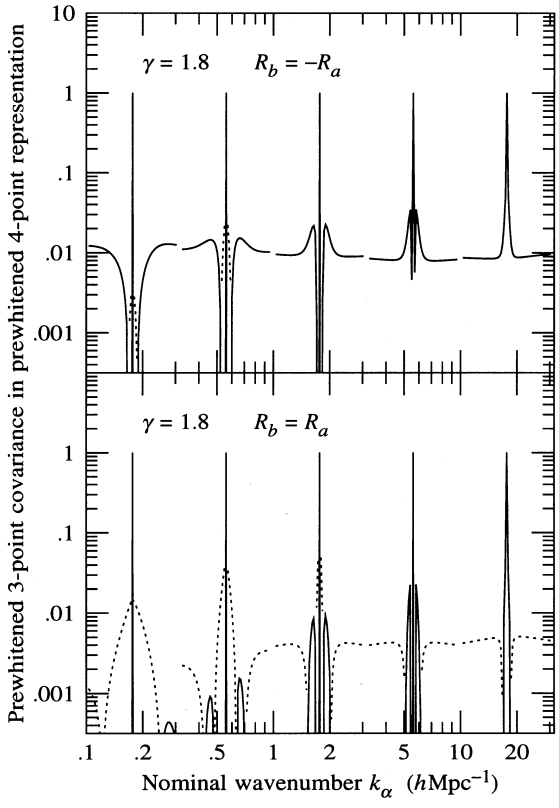

Figure 5. Correlation coefficient $L_{\alpha\beta}/(L_{\alpha\alpha}L_{\beta\beta})^{1/2}$ of the pre-whitened three-point covariance $L_{\alpha\beta}$ in the representation of eigenfunctions ϕ_α of the pre-whitened four-point covariance $M_{\alpha\beta}$, for a power-law power spectrum with correlation function $\xi(r) = (r/5\,h^{-1}\,\mathrm{Mpc})^{-1.8}$. Each line is the correlation coefficient at a fixed nominal wavenumber k_β, plotted against the nominal wavenumber k_α, which labels the four-point eigenfunctions ϕ_α ordered by eigenvalue. Each line peaks at $k_\alpha = k_\beta$, whereat the correlation coefficient is unity. The upper panel is for four-point hierarchical amplitudes $R_b = -R_a$; the lower panel is for $R_b = R_a$. Lines are dotted where the correlation coefficient is negative. The resolution is $\Delta \log k = 1/128$.

Fig. 6 shows the correlation coefficient of the pre-whitened three-point covariance $L_{\alpha\beta}$ in the representation of eigenfunctions of the pre-whitened four-point covariance $M_{\alpha\beta}$ for a number of different power spectra, at a representative nominal wavenumber $k_\alpha = 1\,h\,\mathrm{Mpc}^{-1}$. The figure illustrates that this correlation coefficient remains remarkably diagonal for all power spectra. Again, the range of power-law power spectra shown covers essentially the full range $1 < \gamma < 3$ allowed by the hierarchical model.

In the case $\gamma = 2.9$, the off-diagonal elements of the correlation coefficient shown in Fig. 6 appear to bounce around, even though taken as a whole the correlation coefficient appears more diagonal in this case than in any other. The apparent noise is caused by a near-degeneracy of eigenvalues. Such degeneracy is not too surprising, since in the limit $\gamma \to 3$ the pre-whitened three-point and four-point matrices $L_{\alpha\beta}$ and $M_{\alpha\beta}$ are both expected to become proportional to the unit matrix. Numerically, for both three-point and four-point matrices, there is a degeneracy of eigenvalues between eigenfunctions at small and large wavenumbers (in the sense that eigenfunctions with nearly the same eigenvalue may have their largest components at either small or large wavenumber): the eigenvalues are larger at small and large wavenumbers, and go through a minimum at intermediate wavenumber. The degeneracy causes mixing of the eigenfunctions at small and large wavenumbers, making the correspondence between eigenvalue and nominal wavenumber ambiguous, and resulting in the oscillations in the off-diagonal components apparent in Fig. 6.

4.4 Four-point and three-point eigenvalues

Denote the eigenvalues of the four-point and three-point pre-whitened covariance matrices M and L by

$$M\phi_\alpha = \mu_\alpha^2 \phi_\alpha, \tag{67}$$

$$L\varphi_\alpha = \lambda_\alpha \varphi_\alpha \tag{68}$$

(no implicit summation on the right-hand side), so that for Gaussian fluctuations the eigenvalues μ_α and λ_α would be $\mu_\alpha = \lambda_\alpha = \xi(k_\alpha)$.

Fig. 7 shows the ratio $\mu_\alpha/\xi(k_\alpha)$ of the four-point eigenvalues μ_α to the non-linear power spectrum $\xi(k_\alpha)$, plotted as a function of the nominal wavenumber k_α, which labels the eigenfunctions ϕ_α ordered by eigenvalue, for a power-law power spectrum with correlation function $\xi(r) = (r/5\,h^{-1}\,\mathrm{Mpc})^{-1.8}$. The eigenvalue is comparable to the power spectrum at all wavenumbers, $\mu_\alpha \sim \xi(k_\alpha)$. In the Gaussian, small-k_α regime the eigenvalue is equal to the power spectrum, $\mu_\alpha = \xi(k_\alpha)$, as expected, while in the hierarchical, large-k_α regime the eigenvalue tends asymptotically to close to $2R_a^{1/2}$ times the power spectrum, $\mu_\alpha \approx 2R_a^{1/2}\xi(k_\alpha)$. Similar behaviour is found for other power spectra (not plotted), and for the three-point eigenvalue λ_α, which in the hierarchical regime tends asymptotically to $\lambda_\alpha \approx 2Q\xi(k_\alpha)$.

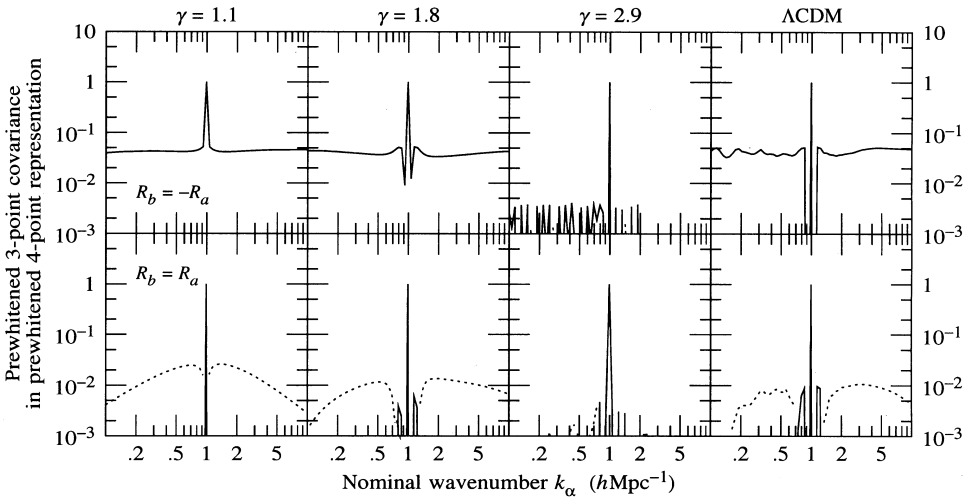

Figure 6. Correlation coefficient $L_{\alpha\beta}/(L_{\alpha\alpha}L_{\beta\beta})^{1/2}$ of the pre-whitened three-point covariance $L_{\alpha\beta}$ in the representation of eigenfunctions ϕ_α of the pre-whitened four-point covariance $M_{\alpha\beta}$, at a representative nominal wavenumber $k_\beta = 1\,h\,\mathrm{Mpc}^{-1}$. The horizontal axis is the nominal wavenumber k_α, which labels the four-point eigenfunctions ϕ_α ordered by eigenvalue. The three sets of panels starting from the left are for power-law power spectra with correlation functions $\xi(r) = (r/5\,h^{-1}\,\mathrm{Mpc})^{-\gamma}$ with indices $\gamma = 1.1$, 1.8 and 2.9, while the rightmost panel is for the ΛCDM power spectrum of Eisenstein & Hu (1998) with $\Omega_\Lambda = 0.7$, $\Omega_\mathrm{m} = 0.3$, $\Omega_\mathrm{b}h^2 = 0.02$ and $h = 0.65$. Upper panels are for four-point hierarchical amplitudes $R_b = -R_a$, lower panels for $R_b = R_a$. Lines are dotted where the correlation coefficient is negative. The resolution is $\Delta \log k = 1/32$, four times coarser than that of Fig. 5. What appears to be noise in the curve for $\gamma = 2.9$ results from a degeneracy of eigenvalues that mixes the correspondence between eigenfunctions ϕ_α and nominal wavenumbers k_α.

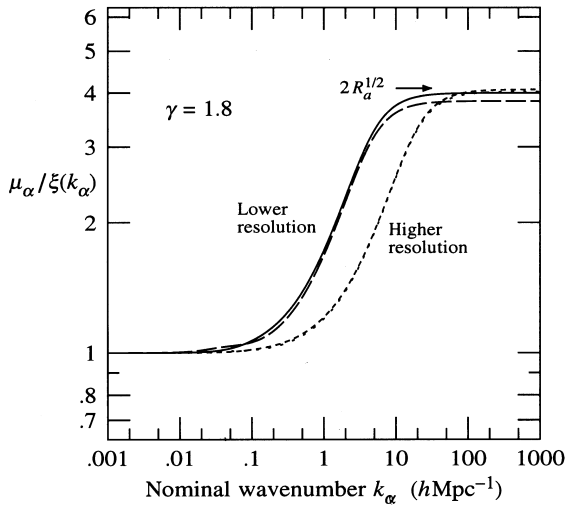

Figure 7. Ratio $\mu_\alpha/\xi(k_\alpha)$ of the eigenvalue μ_α of the four-point pre-whitened covariance matrix M to the non-linear power spectrum $\xi(k_\alpha)$, as a function of the nominal wavenumber k_α, which labels the four-point eigenfunctions ϕ_α ordered by eigenvalue, for a power-law power spectrum with correlation function $\xi(r) = (r/5\,h^{-1}\,\mathrm{Mpc})^{-1.8}$. The relation between eigenvalue and nominal wavenumber varies with resolution. The low-resolution case has $\Delta \log k = 1/32$; the solid line is for $R_b = -R_a$, the long-dashed line for $R_b = R_a$. The high-resolution case has $\Delta \log k = 1/128$; here the (dashed) $R_b = -R_a$ and (dotted) $R_b = R_a$ curves lie practically on top of each other.

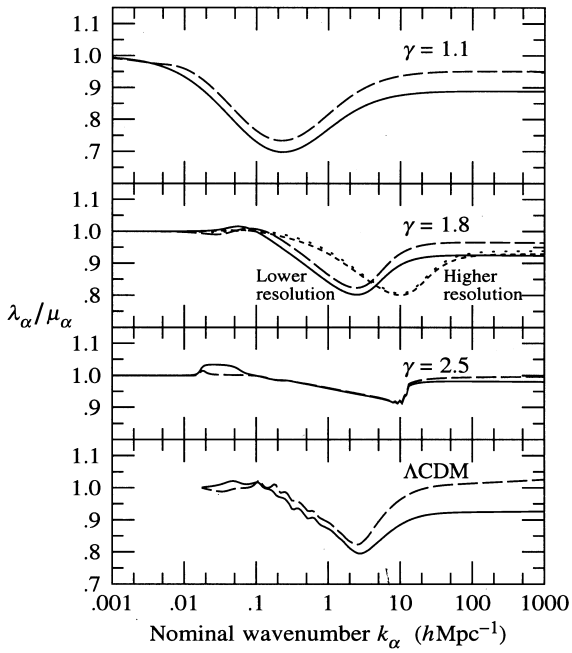

Figure 8. Ratio $\lambda_\alpha/\mu_\alpha$ of the eigenvalues of the three-point and four-point pre-whitened covariance matrices, for various power spectra. The two lines in each case are for four-point hierarchical amplitudes (solid) $R_b = -R_a$ and (long-dashed) $R_b = R_a$. The horizontal axis is the nominal wavenumber k_α, which labels the three-point and four-point eigenfunctions φ_α and ϕ_α ordered by eigenvalue. The relation between eigenvalue and nominal wavenumber varies with resolution. The resolution is $\Delta \log k = 1/32$, except for a high-resolution case shown for $\gamma = 1.8$, where $\Delta \log k = 1/128$.

Fig. 8 shows the ratio $\lambda_\alpha/\mu_\alpha$ of three-point to four-point eigenvalues, as a function of the nominal wavenumber k_α, for several power spectra. Remarkably, the ratio $\lambda_\alpha/\mu_\alpha$ of eigenvalues is quite close to unity at all wavenumbers and for all power spectra. The case $\gamma = 2.9$ is not plotted, in part because of the same problem of mixing of eigenfunctions shown in Fig. 6. In any case, for $\gamma = 2.9$ the ratio $\lambda_\alpha/\mu_\alpha$ differs from unity by less than 1 per cent at all wavenumbers. Analytically, the ratio is expected to equal one in the limit $\gamma \to 3$.

In the ΛCDM model, the eigenfunctions ϕ_α (and φ_α) mix where the eigenvalues μ_α (and λ_α) are degenerate, which happens because the ΛCDM power spectrum goes through a maximum at $k \approx 0.017\,h\,\mathrm{Mpc}^{-1}$. For the purpose of plotting the ratio $\lambda_\alpha/\mu_\alpha$ for the ΛCDM model in the bottom panel of Fig. 8,

this mixing was avoided by the device of truncating the matrices $M_{\alpha\beta}$ and $L_{\alpha\beta}$ at a wavenumber close to the peak. Mixing causes no problems for the evaluation of the minimum variance estimator and Fisher matrix of the pre-whitened power spectrum in Sections 6 and 7 (so there is no need to truncate the matrices in general), but mixing does muddy the physical interpretation of the eigenfunctions.

Curiously, the ratios $\mu_\alpha/\xi(k_\alpha)$ and $\lambda_\alpha/\xi(k_\alpha)$, regarded as functions of the nominal wavenumber k_α, vary with the resolution $\Delta\log k$ of the matrix, as illustrated in Figs 7 and 8 for the case $\gamma = 1.8$. In the Gaussian limit of small k_α, the ratios do not change with resolution, but in the hierarchical limit of large k_α the ratios seems to shift (to the right in the figures, as the resolution increases) in such a way that the ratios are functions of the product $k_\alpha\Delta\log k$. At intermediate k_α, the shift is intermediate. Now the wavenumber k_α is only a nominal wavenumber, a labelling of the eigenfunctions ordered by eigenvalue, and it is only in the Gaussian regime that the eigenmodes are Fourier modes and the correspondence between nominal and true wavenumber is precise. Still, the shift seems surprising: for example, in the limit of infinite resolution $\Delta\log k \to 0$, the ratio $\mu_\alpha/\xi(k_\alpha)$ plotted in Fig. 7 would shift to the right so far that $\mu_\alpha/\xi(k_\alpha)$ would equal 1 at all finite wavenumbers. Similarly, the ratio $\lambda_\alpha/\mu_\alpha$ plotted in Fig. 8 would shift to the right so far that $\lambda_\alpha/\mu_\alpha$ would equal 1 at all finite wavenumbers. Numerically, to the limit to which I have tested it ($\Delta\log k = 1/1024$), this is indeed what seems to happen: both $\mu_\alpha/\xi(k_\alpha)$ and $\lambda_\alpha/\xi(k_\alpha)$, hence also their ratio $\lambda_\alpha/\mu_\alpha$, shift to the right together as the resolution increases, for all power spectra.

This does not appear to be a numerical error, because 'observable' quantities computed via the eigenfunctions ϕ_α and their eigenvalues μ_α, such as the error bars attached to the pre-whitened power spectrum $\hat{X}(k)$ in Fourier space (Section 7), appear robust against changes in resolution.

Examination of the eigenfunctions of the four-point and three-point matrices M and L reveals at least part of the reason why their eigenvalues seem to shift as the resolution increases. Fig. 9 shows

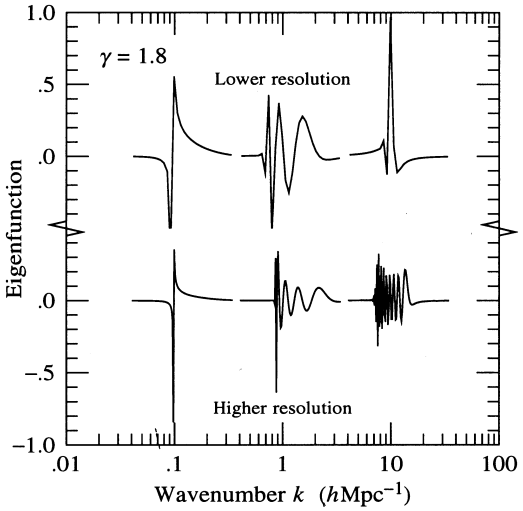

Figure 9. Sample of (discretized) eigenfunctions $\phi_\alpha(k)[4\pi k^3\Delta\ln k/(2\pi)^3]^{1/2}$ of the pre-whitened four-point covariance $M_{\alpha\beta}$ for a power-law power spectrum with correlation function $\xi(r) = (r/5\,h^{-1}\,\mathrm{Mpc})^{-1.8}$, and $R_b = -R_a$, at resolutions of (top) $\Delta\log k = 1/32$ and (bottom) $\Delta\log k = 1/128$. The eigenfunctions have nominal wavenumbers k_α of 0.1, 1 and $10\,h\,\mathrm{Mpc}^{-1}$. In the hierarchical regime, the eigenfunctions grow wigglier as the resolution increases.

a sampling of eigenfunctions ϕ_α of the four-point matrix M for the case $\xi(r) = (r/5\,h^{-1}\,\mathrm{Mpc})^{-1.8}$, at two different resolutions, $\Delta\log k = 1/32$ and 1/128. Whereas in the Gaussian, small-k_α regime the eigenfunctions go over to delta-functions in Fourier space, in the hierarchical, large-k_α regime the eigenfunctions grow ever wigglier as the resolution increases. What seems to happen is that, as the resolution increases, eigenfunctions at neighbouring nominal wavenumbers k_α strive to remain orthogonal to each other, which they accomplish by becoming wigglier and wigglier. To the limit to which I have tested it numerically, there seems to be no end to the wiggliness. Given that there is no asymptotic limit to which the eigenfunctions appear to tend, perhaps it is not surprising that their eigenvalues should shift systematically too. However, it would be nice to have a better understanding of what is going on.

5 PRE-WHITENED POWER SPECTRUM

5.1 Definition

Given the nice properties of the pre-whitened covariance of power established in the previous section, Section 4, it makes sense to define a pre-whitened power spectrum X_α, and a corresponding estimator $\hat{X}_\alpha$ thereof, with the property that the covariance of the pre-whitened power equals the pre-whitened covariance of power.

Define, therefore, the pre-whitened power spectrum X_α by, in the real-space representation,

$$X(r) \equiv \frac{2\xi(r)}{1 + [1 + \xi(r)]^{1/2}}. \tag{69}$$

The expression (69) is equivalent to $X(r) \equiv 2[1 + \xi(r)]^{1/2} - 2$, but the former expression (69) is numerically more stable to evaluate when $\xi(r)$ is small. Similarly, define an estimator $\hat{X}_\alpha$ of the pre-whitened power in terms of the minimum variance estimator $\hat{\xi}_\alpha$, equation (33), of the power spectrum by, again in the real-space representation,

$$\hat{X}(r) \equiv \frac{2\hat{\xi}(r)}{1 + [1 + \hat{\xi}(r)]^{1/2}}, \tag{70}$$

which by construction has the property that for small $\Delta\hat{X}(r)$, as should be true in the limit of a large amount of data [the following equation is essentially the derivative of equation (70)],

$$\Delta\hat{X}(r) = \frac{\Delta\hat{\xi}(r)}{[1 + \hat{\xi}(r)]^{1/2}}. \tag{71}$$

The covariance of the estimate $\hat{X}_\alpha$ of the pre-whitened power spectrum is given by

$$\langle\Delta\hat{X}_\alpha\Delta\hat{X}_\beta\rangle = (H^{-1/2})^\gamma_\alpha\langle\Delta\hat{\xi}_\gamma\Delta\hat{\xi}_\delta\rangle(H^{-1/2})^\delta_\beta = E^{-1}_{\alpha\beta}, \tag{72}$$

where the Fisher matrix $E^{\alpha\beta}$ of the pre-whitened power equals the pre-whitened Fisher matrix of the power, equation (35),

$$E^{\alpha\beta} = (H^{1/2})^\alpha_\gamma F^{\gamma\delta}(H^{1/2})^\beta_\delta. \tag{73}$$

In Section 7 it will be found convenient to deal with another pre-whitened estimator $\hat{Y}_\alpha$ defined by

$$\hat{Y}_\alpha \equiv (H^{-1/2})^\beta_\alpha\hat{\xi}_\beta. \tag{74}$$

The pre-whitened estimator $\hat{Y}_\alpha$ has the same covariance as $\hat{X}_\alpha$:

$$\langle\Delta\hat{Y}_\alpha\Delta\hat{Y}_\beta\rangle = \langle\Delta\hat{X}_\alpha\Delta\hat{X}_\beta\rangle = E^{-1}_{\alpha\beta}. \tag{75}$$

So why not define $\hat{Y}_\alpha$ to be the pre-whitened power? The problem with the estimator $\hat{Y}_\alpha$ is that it depends explicitly on the prior power spectrum ξ_α. That is, $\hat{Y}_\alpha$ in real space is

$$\hat{Y}(r) = \frac{\hat{\xi}(r)}{[1+\xi(r)]^{1/2}}, \tag{76}$$

which involves an estimated quantity $\hat{\xi}(r)$ in the numerator and the prior quantity $\xi(r)$ in the denominator. Imagine plotting $\hat{Y}_\alpha$ on a graph. Of what is this quantity supposed to be an estimate? Obviously $\hat{Y}_\alpha$ is an estimate of $Y_\alpha \equiv \langle \hat{Y}_\alpha \rangle = (H^{-1/2})^\beta_\alpha \xi_\beta$. If one wanted to attach error bars to the estimate, then to be fair one should include the full covariance of the quantity being estimated, including the covariance that arises from the denominator $[1+\xi(r)]^{1/2}$ in equation (76), not just the covariance $\langle \Delta\hat{Y}_\alpha \Delta\hat{Y}_\beta \rangle$ with the denominator held fixed. Indeed, if one goes through the usual ML cycle of permitting the data to inform the prior, so that the estimated $\hat{\xi}(r)$ is inserted into the denominator of equation (76), then it becomes abundantly evident that it would be correct to include covariance arising from the denominator.

To avoid confusion, it should be understood that the quantities $\hat{Y}_\alpha$ are of course perfectly fine for carrying out ML estimation of parameters. In ML estimation, 'error bars are attached to the model, not to the data', to quote another of the refrains from the 1997 Aspen workshop on Precision Measurement of Large Scale Structure. Whereas in ML parameter estimation with $\hat{X}_\alpha$ one might form a likelihood from the 'data' quantities $\Delta\hat{X}(r) = 2\hat{\xi}(r)/\{1+[1+\hat{\xi}(r)]^{1/2}\} - 2\xi(r)/\{1+[1+\xi(r)]^{1/2}\}$, in ML parameter estimation with $\hat{Y}_\alpha$ one would instead form a likelihood from the 'data' quantities $\Delta\hat{Y}(r) = [\hat{\xi}(r) - \xi(r)]/[1+\xi(r)]^{1/2}$.

For the purpose of plotting quantities on a graph, however, plainly it is the pre-whitened power spectrum $\hat{X}_\alpha$ defined by equation (70) that should be plotted, not $\hat{Y}_\alpha$.

5.2 Picture

Fig. 10 shows pre-whitened non-linear power spectra $X(k)$, along with linear and non-linear power spectra $\xi_{\rm L}(k)$ and $\xi(k)$, for the observationally derived power spectrum of Peacock (1997) with $\Omega_{\rm m} = 0.3$, and for a ΛCDM model of Eisenstein & Hu (1998) with observationally concordant parameters as indicated on the graph.

The non-linear power spectra $\xi(k)$ were constructed from the linear power spectra $\xi_{\rm L}(k)$ according to the formula of Peacock & Dodds (1996). Amongst other things, the Peacock & Dodds formula depends on the logarithmic slope of the linear power spectrum. Now the Eisenstein & Hu power spectrum contains baryonic wiggles, causing the slope to oscillate substantially, whereas what Peacock & Dodds had in mind was a rough average slope. For the slope of the ΛCDM model in the Peacock & Dodds formula, I therefore used the slope of the 'no-wiggle' power spectrum provided by Eisenstein & Hu as a smooth fit through the baryonic wiggles. The alternative of using the wiggly slope has the additional demerit that it amplifies baryonic wiggles in the non-linear regime, which is opposite to the suppression of baryonic wiggles in the non-linear regime observed in *N*-body simulations by Meiksin, White & Peacock (1999).

The pre-whitened power spectra $X(k)$ shown in Fig. 10 were computed by transforming the non-linear power spectrum $\xi(k)$ into real space using FFTLog (see Appendix B, Fig. B1), constructing the pre-whitened power $X(r)$ from $\xi(r)$ according to equation (69), and Fourier transforming back.

The pre-whitened power spectra shown in Fig. 10 appear to be interestingly close to the linear power spectra, $X(k) \approx \xi_{\rm L}(k)$, another one-eyebrow-raising property of the pre-whitened power spectrum. Surely this is just coincidence, however, since for a primordial power spectrum $\xi(k) \propto k^n$ the pre-whitened correlation in the highly non-linear regime should go as $X(r) \approx 2\xi(r)^{1/2} \propto r^{-3(n+3)/2(n+5)}$ assuming stable clustering (Peebles 1980, equation 73.12), whereas the linear power spectrum would go as $r^{-(n+3)}$, the power-law exponents of which agree only in the limiting case $n \to -3$. Still, the coincidence is curious.

Fig. 10 points up one defect of the pre-whitened power spectrum, which is that, surprisingly enough, it does not reproduce the linear power spectrum at the very largest scales (small k). Indeed, the pre-whitened power goes negative in the Peacock (1997) case at $k \approx 0.0023\,h\,{\rm Mpc}^{-1}$, and in the ΛCDM case at $k \approx 0.000\,21\,h\,{\rm Mpc}^{-1}$. This turns out to be a generic feature of the pre-whitened power spectrum if the true power spectrum goes to zero at zero wavenumber, as is true for Harrison–Zel'dovich

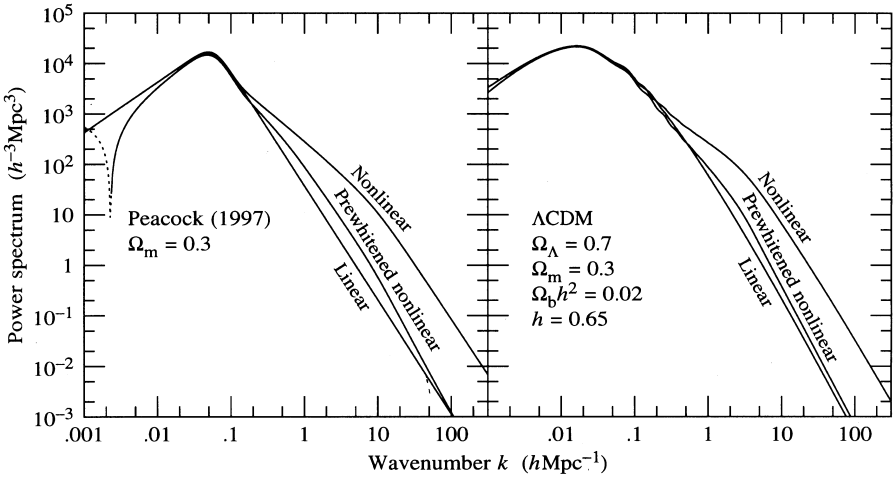

Figure 10. Linear power spectrum $\xi_{\rm L}(k)$, non-linear power spectrum $\xi(k)$, and pre-whitened non-linear power spectrum $X(k)$ for (left) the $\Omega_{\rm m} = 0.3$ power spectrum derived from observations by Peacock (1997), and (right) the *COBE*-normalized ΛCDM power spectrum from the fitting formulae of Eisenstein & Hu (1998), with parameters as listed on the graph. The non-linear power spectra were computed from the linear power spectra according to the formula of Peacock & Dodds (1996). The ΛCDM power spectrum is the one used in Figs 4, 6, 8, 11 and B1.

models, $\xi(k) \propto k$ as $k \to 0$. For, if it is true that the power spectrum $\xi(k)$ goes to zero at zero wavenumber k,

$$\lim_{k\to 0} \xi(k) = \int_0^\infty \xi(r) 4\pi r^2 \, \mathrm{d}r = 0, \tag{77}$$

then it follows that the pre-whitened power must go to a negative constant at zero wavenumber,

$$\lim_{k\to 0} X(k) = \int_0^\infty \frac{2\xi(r)}{1 + [1+\xi(r)]^{1/2}} 4\pi r^2 \, \mathrm{d}r < 0, \tag{78}$$

since the factor $2/\{1 + [1 + \xi(r)]^{1/2}\}$ in the integrand is less than one for all positive $\xi(r)$, and greater than one for all negative $\xi(r)$. It is not clear what to do about this, if indeed anything needs to be done. Adding a constant to $X(k)$ and $\hat{X}(k)$ (which would leave $\Delta\hat{X}$, hence the covariance $\langle \Delta\hat{X}\Delta\hat{X}\rangle$, unchanged) would spoil the nice behaviour of the pre-whitened power in the non-linear regime.

6 FISHER MATRIX OF PRE-WHITENED NON-LINEAR POWER IN A SURVEY

It was found in Section 4 that the pre-whitened reduced covariance $\mathfrak{B}$ of power appears to have some unexpectedly pleasant properties: first, the pre-whitened covariance is surprisingly narrow in Fourier space; secondly, the four-point and three-point contributions M and L, equation (66), to the pre-whitened reduced covariance $\mathfrak{B}$ are almost simultaneously diagonal (the two-point contribution is by construction the unit matrix, so is automatically diagonal in any representation); thirdly, the four-point and three-point eigenvalues μ_α and λ_α, as defined by equations (67) and (68), are approximately equal; fourthly, all these results hold for all power spectra tested.

It should be emphasized that the pleasant properties of the pre-whitened power are not perfect, and that they are premised on the validity of the hierarchical model with constant hierarchical amplitudes, which as discussed in Section 4.2 is certainly wrong at some level.

These pretty properties lead to an approximate expression, equation (83), for the Fisher matrix of the pre-whitened non-linear power spectrum of a galaxy survey, which looks the same as the FKP approximation to the Fisher matrix of the power in the linear, Gaussian case, with the difference that the eigenmodes of the pre-whitened covariance M of the non-linear power take the place of the Fourier modes in the linear case.

6.1 Fisher matrix

To the extent that the pre-whitened four-point and three-point matrices M and L are simultaneously diagonal, the pre-whitened reduced covariance matrix $\mathfrak{B}_{\alpha\beta}(\bar{n}_i, \bar{n}_j)$ is diagonal in the representation of eigenfunctions ϕ_α of M and L, with

$$\mathfrak{B}_{\alpha\beta}(\bar{n}_i, \bar{n}_j) \approx 2\mathbf{1}_{\alpha\beta}[\mu_\alpha^2 + (\bar{n}_i^{-1} + \bar{n}_j^{-1})\lambda_\alpha + \bar{n}_i^{-1}\bar{n}_j^{-1}]. \tag{79}$$

To the further extent that $\lambda_\alpha \approx \mu_\alpha$, the pre-whitened covariance matrix $\mathfrak{B}_{\alpha\beta}(\bar{n}_i, \bar{n}_j)$ is just

$$\mathfrak{B}_{\alpha\beta}(\bar{n}_i, \bar{n}_j) \approx 2\mathbf{1}_{\alpha\beta}(\mu_\alpha + \bar{n}_i^{-1})(\mu_\alpha + \bar{n}_j^{-1}). \tag{80}$$

The Fisher matrix $F^{\alpha\beta}$ of the power spectrum is given in the FKP approximation by equation (45). In terms of the pre-whitened reduced covariance $\mathfrak{B}_{\alpha\beta}$, the Fisher matrix $F^{\alpha\beta}$ is

$$F^{\alpha\beta} = (H^{-1/2})^\alpha_\delta \mathfrak{B}^{-1\delta\epsilon} (H^{-1/2})^\gamma_\epsilon D^{ij}_\gamma D^\beta_{ij}. \tag{81}$$

Now $(H^{-1/2})^\gamma_\epsilon$ commutes with $D^{ij}_\gamma D^\beta_{ij}$, since both are simultaneously diagonal in real space. It follows that the Fisher matrix $E \equiv H^{1/2} F H^{1/2}$ of the pre-whitened power, equation (73), is, in the FKP approximation,

$$E^{\alpha\beta} = \mathfrak{B}^{-1\alpha\gamma} D^{ij}_\gamma D^\beta_{ij}. \tag{82}$$

Like $F^{\alpha\beta}$, the pre-whitened Fisher matrix $E^{\alpha\beta}$ is asymmetric, inheriting its asymmetry from the FKP approximation, equation (41).

To the extent that the approximation (80) to $\mathfrak{B}$ is true, it follows from equation (82) that the Fisher matrix $E_{\alpha\beta}$ of pre-whitened power in the FKP approximation is, in the representation of eigenfunctions ϕ_α of the pre-whitened covariance,

$$E_{\alpha\beta} = \int_0^\infty \phi_\alpha(r)\phi_\beta(r) R(r;\mu_\alpha) 4\pi r^2 \, \mathrm{d}r, \tag{83}$$

where $R(r;\mu_\alpha)$ are FKP-weighted pair integrals (commonly denoted $\langle RR\rangle$ in the literature, for random–random),

$$R(r;\mu_\alpha) = \int \frac{\delta_{3\mathrm{D}}(r_{ij} - r)}{2[\mu_\alpha + \bar{n}(\boldsymbol{r}_i)^{-1}][\mu_\alpha + \bar{n}(\boldsymbol{r}_j)^{-1}]} \mathrm{d}^3 r_i \, \mathrm{d}^3 r_j, \tag{84}$$

the integration being taken over all pairs of volume elements ij separated by $r_{ij} \equiv |\boldsymbol{r}_i - \boldsymbol{r}_j| = r$ in the survey.

The FKP approximation to the Fisher matrix $E_{\alpha\beta}$ of pre-whitened power, equation (83), takes the same form as the FKP approximation to the Fisher matrix of the power spectrum for Gaussian fluctuations derived in section 5 of Paper I and computed in section 3 of Paper II. The difference is that the eigenfunctions $\phi_\alpha(r)$ and their eigenvalues μ_α here take the place of the Fourier eigenfunctions $j_0(k_\alpha r)$ and their eigenvalues $\xi(k_\alpha)$ in the Gaussian case.

6.2 Numerics

Equation (83) for $E_{\alpha\beta}$ involves the eigenfunctions $\phi_\alpha(r)$ of the pre-whitened four-point matrix M in real space, whereas in Section 4.1 it was suggested that the most robust way to compute M is in Fourier space. The problem is that FFTing the matrix M from Fourier into real space is liable to introduce ringing and aliasing, which one would like to avoid.

A more robust procedure is not to FFT M into real space, but rather to FFT the pair integrals $R(r;\mu_\alpha)$ into Fourier space; this is the same procedure as adopted in section 3 of Paper II (except that R here is 1/2 that of Paper II). If μ_α is treated, temporarily, as a constant, then equation (83) can be transformed into real space to yield the diagonal matrix

$$E(r, r'; \mu_\alpha) = \delta_{3\mathrm{D}}(r - r') R(r;\mu_\alpha). \tag{85}$$

Beware of equation (85)! It does not signify that the Fisher matrix is diagonal in real space, because the constant μ_α is different for each row of the Fisher matrix $E_{\alpha\beta}$. The Fourier transform of $E(r, r';\mu_\alpha)$ is $E(k, k';\mu_\alpha) = \int j_0(kr) j_0(k'r) R(r;\mu_\alpha) 4\pi r^2 \, \mathrm{d}r$, which simplifies to

$$E(k, k';\mu_\alpha) = \frac{\pi}{kk'}[\tilde{R}(k - k';\mu_\alpha) - \tilde{R}(k + k';\mu_\alpha)] \tag{86}$$

where $\tilde{R}(k;\mu_\alpha)$ is the one-dimensional cosine transform of $R(r;\mu_\alpha)$:

$$\tilde{R}(k;\mu_\alpha) \equiv 2\int_0^\infty \cos(kr) R(r;\mu_\alpha) \, \mathrm{d}r. \tag{87}$$

Transforming $E(k,k';\mu_\alpha)$ into ϕ_α-space gives

$$E_{\alpha\beta}=\int\phi_\alpha(k)\phi_\beta(k')E(k,k';\mu_\alpha)\frac{4\pi k^2\,\mathrm{d}k\,4\pi k'^2\,\mathrm{d}k'}{(2\pi)^6}. \tag{88}$$

The cosine transform $\tilde{R}(k;\mu_\alpha)$, equation (87), can be done with either FFT or FFTLOG; both work well. To ensure that $\tilde{R}(k;\mu_\alpha)$ remains accurate at large (and small) wavenumbers k, it helps to extrapolate $R(r;\mu_\alpha)$ to small (and large) separations r before transforming. The transformation into ϕ_α-space, equation (88), is done by discrete summations.

Evaluating the Fisher matrix $E_{\alpha\beta}$ with equations (86)–(88) successfully eliminates ringing and aliasing, but it introduces another problem. The problem is that equation (86) is liable to overestimate the value of $E(k,k';\mu_\alpha)$ along the diagonal $k=k'$ if the gridding in k-space is too coarse to resolve the diagonal properly, as typically occurs at moderate and large k with logarithmic gridding. What is important is that the integral of $E(k,k';\mu_\alpha)$ over the diagonal be correct. Integrating $E(k,k';\mu_\alpha)$ over k' yields

$$\int_0^\infty E(k,k';\mu_\alpha)k'\,\mathrm{d}k'=\frac{2\pi}{k}\int_0^k\tilde{R}(k';\mu_\alpha)\,\mathrm{d}k'. \tag{89}$$

The integral on the right can be done conveniently and reliably by sine transforming (with FFT or FFTLOG) the pair integral

$$\int_0^k\tilde{R}(k';\mu_\alpha)\,\mathrm{d}k'=2\int_0^\infty\sin(kr)\frac{R(r;\mu_\alpha)}{r}\mathrm{d}r. \tag{90}$$

Discretized (Section 2.3) on a logarithmic grid of wavenumbers k, the continuous matrix $E(k,k';\mu_\alpha)$ becomes $\mathbf{E}_{kk'}(\mu_\alpha)=E(k,k';\mu_\alpha)\,4\pi(kk')^{3/2}\Delta\ln k/(2\pi)^3$, and equation (89) becomes

$$\sum_{k'}(k'/k)^{1/2}\mathbf{E}_{kk'}(\mu_\alpha)=\frac{1}{\pi}\int_0^k\tilde{R}(k';\mu_\alpha)\,\mathrm{d}k'. \tag{91}$$

Numerically, if the left-hand side of equation (91), with $\mathbf{E}_{kk'}(\mu_\alpha)$ discretized from equation (86), exceeds the right-hand side of equation (91), evaluated by equation (90), then the value of the diagonal element $\mathbf{E}_{kk}(\mu_\alpha)$ should be reduced so that the sum is satisfied. Ultimately, this procedure yields error bars on decorrelated band-powers (Paper IV) that are robust with respect to range, resolution and linear or logarithmic binning.

6.3 Coarse gridding

Typically the pair integral $R(r;\mu_\alpha)$ is broad in real space, so its cosine transform $\tilde{R}(k;\mu_\alpha)$ is a narrow window about $k\approx0$ with a width comparable to the inverse scalelength of the survey. It follows that the matrix $E(k,k';\mu_\alpha)$ given by equation (86) is likewise narrow in k-space, with a width comparable to the inverse scalelength of the survey. Moreover, the sum in equation (91) approximates $R(0;\mu_\alpha)$ at wavenumbers exceeding the inverse scalelength of the survey, which is to say at all except the largest accessible wavelengths:

$$\sum_{k'}(k'/k)^{1/2}\mathbf{E}_{kk'}(\mu_\alpha)\approx R(0;\mu_\alpha)\quad\text{for}\quad k\gg\text{scale}^{-1}, \tag{92}$$

where $R(0;\mu_\alpha)$ is the pair integral at zero separation

$$R(0;\mu_\alpha)=\int\frac{\mathrm{d}^3r}{2[\mu_\alpha+\bar{n}(\boldsymbol{r})^{-1}]^2}. \tag{93}$$

Thus, if the matrix $\mathbf{E}_{kk'}(\mu_\alpha)$ is discretized on a grid that is coarse compared with the inverse scalelength of the survey, then it is approximately proportional to the unit matrix:

$$\mathbf{E}_{kk'}(\mu_\alpha)\approx\mathbf{1}_{kk'}R(0;\mu_\alpha). \tag{94}$$

The resulting discrete Fisher matrix $\mathbf{E}_{\alpha\beta}$, equation (88), is diagonal in the ϕ_α-representation

$$\mathbf{E}_{\alpha\beta}\approx\mathbf{1}_{\alpha\beta}R(0;\mu_\alpha). \tag{95}$$

The result (95) is analogous to that obtained by FKP for Gaussian fluctuations.

Of course, if this diagonal Fisher matrix, equation (95), is transformed back into Fourier space, then it is no longer diagonal. That is, equation (95) asserts that the Fisher matrix of the pre-whitened non-linear power spectrum is approximately diagonal in ϕ_α-space, not in Fourier space.

7 ESTIMATE OF PRE-WHITENED NON-LINEAR POWER IN A SURVEY

7.1 Unbiased estimate

'In the case of a Gaussian distribution… rather than removing the bias we should approximately double it, in order to minimize the mean square sampling error' – E. T. Jaynes (1996, sentence containing equation 17-13).

It is convenient to start out by considering the pre-whitened estimator $\hat{Y}_\alpha$ defined by equation (74). The minimum variance estimator $\hat{\xi}_\alpha$ of the power spectrum in the FKP approximation is given by equation (44). Translating this equation into pre-whitened quantities, one concludes that the minimum variance pre-whitened estimator $\hat{Y}_\alpha$ in the FKP approximation is, in terms of the pre-whitened reduced covariance $\mathfrak{B}$, equation (63), and its associated Fisher matrix E, equation (73),

$$\hat{Y}_\alpha=E^{-1}_{\alpha\beta}\mathfrak{B}^{-1\beta\gamma}(H^{-1/2})^{\epsilon}_{\gamma}D^{ij}_{\epsilon}(\delta_i\delta_j-\hat{N}_{ij}). \tag{96}$$

The estimator $\hat{Y}_\alpha$ is minimum variance if and only if $\hat{\xi}_\alpha$ is minimum variance, since $\hat{Y}_\alpha\equiv(H^{-1/2})^{\beta}_{\alpha}\hat{\xi}_\beta$ is a linear combination of $\hat{\xi}_\alpha$.

Now the estimator $\hat{Y}_\alpha$, equation (96), is intended to be an estimate of $Y_\alpha\equiv(H^{-1/2})^{\beta}_{\alpha}\xi_\beta$, but is that really true, given the various approximations? It will be true provided that the estimator is unbiased, meaning that the expectation value of the estimator is equal to the true value

$$\langle\hat{Y}_\alpha\rangle=Y_\alpha. \tag{97}$$

The expectation value of the estimator $\hat{Y}_\alpha$ given by equation (96) is, since $\langle\delta_i\delta_j-\hat{N}_{ij}\rangle=D^{\alpha}_{ij}\xi_\alpha$ according to equation (27),

$$\begin{aligned}\langle\hat{Y}_\alpha\rangle&=E^{-1}_{\alpha\beta}\mathfrak{B}^{-1\beta\gamma}(H^{-1/2})^{\epsilon}_{\gamma}D^{ij}_{\epsilon}D^{\zeta}_{ij}\xi_\zeta\\&=E^{-1}_{\alpha\beta}\mathfrak{B}^{-1\beta\gamma}D^{ij}_{\gamma}D^{\epsilon}_{ij}Y_\epsilon,\end{aligned} \tag{98}$$

where the second line follows because $(H^{-1/2})^{\epsilon}_{\gamma}$ commutes with $D^{ij}_{\epsilon}D^{\zeta}_{ij}$, both being diagonal in real space. It follows that the estimator $\hat{Y}_\alpha$ will be unbiased, $\langle\hat{Y}_\alpha\rangle=Y_\alpha$, provided that the Fisher matrix $E^{\alpha\beta}$ is taken to satisfy the asymmetric equation (82), not, for example, a symmetrized version of that equation.

An important point to recognize here is that an estimate $\hat{Y}_\alpha$ of the form (96) will be unbiased for *any* a priori choice of the matrix $\mathfrak{B}$, regardless of the choice of prior power $\xi(k)$, regardless of the hierarchical model, regardless of the FKP approximation, and regardless of the approximation (such as equation 80) to $\mathfrak{B}$, just so

long as the matrix E in the estimator is interpreted as satisfying the unsymmetrized equation (82). Ultimately this property of being unbiased is inherited from the basic prior assumption that galaxies constitute a random, Poisson sampling of an underlying statistically homogeneous, isotropic density field, so that the product of overdensities $\delta_i\delta_j$ at any pair of points ij separated by r_α provides an unbiased estimate of the correlation function $\xi(r_\alpha)$. Note that the presumption here is that the galaxies sampled are an unbiased tracer of the galaxy density itself, not necessarily of the mass density.

Interpreting the estimator $\hat{Y}_\alpha$, equation (96), as involving the asymmetric matrix $E^{\alpha\beta}$, equation (82), should be regarded not as changing the estimator to make it unbiased, but rather as interpreting the estimator correctly. If instead the estimator $\hat{Y}_\alpha$ were interpreted as involving the symmetrized Fisher matrix $E^{(\alpha\beta)} \equiv \mathrm{Sym}_{(\alpha\beta)}E^{\alpha\beta}$, for example, then the expectation value of the estimator would be $\langle\hat{Y}_\alpha\rangle = E^{-1}_{(\alpha\beta)}E^{\beta\gamma}Y_\gamma$, which is not the same as Y_α, although of course it should be almost the same to the extent that $E_{\alpha\beta}$ is almost symmetric.

It is convenient to introduce yet another estimator $\hat{Z}^\alpha$ related to the estimator $\hat{Y}_\alpha$ by

$$\hat{Y}_\alpha = E^{-1}_{\alpha\beta}\hat{Z}^\beta. \tag{99}$$

In the FKP approximation, the estimator $\hat{Z}^\alpha$ is

$$\hat{Z}^\alpha = \mathfrak{B}^{-1\alpha\beta}(H^{-1/2})^\gamma_\beta D^{ij}_\gamma(\delta_i\delta_j - \hat{N}_{ij}). \tag{100}$$

If the approximation (80) to the pre-whitened covariance $\mathfrak{B}$ is used in the estimate (100) of $\hat{Z}$, then, in the representation of eigenfunctions ϕ_α,

$$\hat{Z}_\alpha = \int_0^\infty \frac{\phi_\alpha(r)\hat{S}(r;\mu_\alpha)}{[1+\xi(r)]^{1/2}} 4\pi r^2\,\mathrm{d}r, \tag{101}$$

where $\hat{S}(r;\mu_\alpha)$ is the FKP-weighted integral over pairs of overdensities $\delta(\boldsymbol{r}_i)\delta(\boldsymbol{r}_j)$ at points ij separated by $r_{ij} \equiv |\boldsymbol{r}_i - \boldsymbol{r}_j| = r$ (commonly denoted $\langle DD\rangle - 2\langle DR\rangle + \langle RR\rangle$ in the literature: D for data, R for random):

$$\hat{S}(r;\mu_\alpha) = \int \frac{\delta_{3\mathrm{D}}(r_{ij}-r)\delta(\boldsymbol{r}_i)\delta(\boldsymbol{r}_j)}{2[\mu_\alpha + \bar{n}(\boldsymbol{r}_i)^{-1}][\mu_\alpha + \bar{n}(\boldsymbol{r}_j)^{-1}]}\mathrm{d}^3r_i\,\mathrm{d}^3r_j. \tag{102}$$

The shot-noise $\hat{N}_{ij}$ is excluded from equation (102) by excluding from the integration the contribution from self-pairs of galaxies, which of course have zero separation. The associated asymmetric Fisher matrix $E^{\alpha\beta}$ is given by equation (83).

Equation (102) is expressed as an integral over pairs of overdensities $\delta(\boldsymbol{r}_i)\delta(\boldsymbol{r}_j)$ in real space. One could just as well express $\hat{S}$ as an integral over pairs of overdensities $\delta(\boldsymbol{k}_i)\delta(\boldsymbol{k}_j)$ in Fourier space, or pairs of overdensities $\delta(k_i,\ell_i,m_i)\delta(k_j,\ell_j,m_j)$ in spherical harmonic space, if one found it more convenient.

7.2 Numerics

As in Section 6.2, to avoid potential problems of ringing and aliasing, it is probably better to evaluate the estimator $\hat{Z}_\alpha$, equation (101), by means of an expression that involves the eigenfunctions $\phi_\alpha(k)$ in Fourier space rather than the eigenfunctions $\phi_\alpha(r)$ in real space.

If μ_α is treated, temporarily, as a constant, then transforming equation (101) into real space yields

$$\hat{Z}(r;\mu_\alpha) = \frac{\hat{S}(r;\mu_\alpha)}{[1+\xi(r)]^{1/2}}. \tag{103}$$

The Fourier transform of this is

$$\hat{Z}(k;\mu_\alpha) = \int_0^\infty j_0(kr)\hat{Z}(r;\mu_\alpha)4\pi r^2\,\mathrm{d}r, \tag{104}$$

in terms of which the estimator $\hat{Z}_\alpha$, equation (101), is

$$\hat{Z}_\alpha = \int_0^\infty \phi_\alpha(k)\hat{Z}(k;\mu_\alpha)\frac{4\pi k^2\,\mathrm{d}k}{(2\pi)^3}. \tag{105}$$

The transformation into ϕ_α-space, equation (105), is done by discrete summation.

The advantage of equation (105) over the nominally equivalent equation (101) is that in equation (105) it is the data that are Fourier transformed, $\hat{Z}(r;\mu_\alpha) \rightarrow \hat{Z}(k;\mu_\alpha)$, equation (104), whereas in equation (101) it is the eigenfunctions of the matrix M that must be transformed, $\phi_\alpha(k) \rightarrow \phi_\alpha(r)$. While the two methods would yield identical results for $\hat{Z}_\alpha$ if the same unitary Fourier transform were applied in both cases, in reality it may be advantageous to have the freedom to Fourier transform the data the best way one can, without regard to the irrelevant question of how the eigenfunctions ϕ_α behave when Fourier transformed.

7.3 The covariance of $\hat{Z}^\alpha$

It will now be argued that the covariances of the estimators $\hat{Z}^\alpha$ and $\hat{Y}^\alpha$ are approximately equal to, respectively, the symmetrized Fisher matrix $E^{(\alpha\beta)}$ and its inverse, equations (109) and (115). It seems worthwhile to go through the arguments rather carefully. As a general rule, one should estimate error bars as accurately as possible; but if some approximation is necessary, then one would prefer to err on the conservative side of overestimating the true errors.

Equation (109) will now be derived, commentary on the derivation being deferred to the end. The covariance of the estimate $\hat{Z}^\alpha$ is, from equation (100),

$$\begin{aligned}\langle\Delta\hat{Z}^\alpha\Delta\hat{Z}^\beta\rangle = {}& \mathfrak{B}^{-1\alpha\gamma}(\bar{n}_i,\bar{n}_j)(H^{-1/2})^\epsilon_\gamma D^{ij}_\epsilon \mathfrak{C}^{\mathrm{true}}_{ijkl} D^{kl}_\zeta \\ & \times (H^{-1/2})^\zeta_\eta \mathfrak{B}^{-1\eta\beta}(\bar{n}_k,\bar{n}_l),\end{aligned} \tag{106}$$

in which $\mathfrak{B}$ is the approximate pre-whitened reduced covariance matrix (80) used to construct the estimate $\hat{Z}^\alpha$, equation (101), while $\mathfrak{C}^{\mathrm{true}}_{ijkl}$ is the true covariance matrix, equation (38). To the extent that the FKP approximation, equation (41), is valid for $\mathfrak{C}^{\mathrm{true}}_{ijkl}$, equation (106) reduces to

$$\begin{aligned}\langle\Delta\hat{Z}^\alpha\Delta\hat{Z}^\beta\rangle &= \mathfrak{B}^{-1\alpha\gamma}(\bar{n}_i,\bar{n}_j)(H^{-1/2})^\epsilon_\gamma D^{ij}_\epsilon D^\theta_{ij} \\ &\quad \times \mathfrak{C}^{\mathrm{FKP}}_{\theta\zeta}(\bar{n}_i,\bar{n}_j)(H^{-1/2})^\zeta_\eta \mathfrak{B}^{-1\eta\beta}(\bar{n}_i,\bar{n}_j) \\ &= \mathfrak{B}^{-1\alpha\gamma}(\bar{n}_i,\bar{n}_j)D^{ij}_\gamma D^\epsilon_{ij}\mathfrak{B}^{\mathrm{FKP}}_{\epsilon\zeta}(\bar{n}_i,\bar{n}_j) \\ &\quad \times \mathfrak{B}^{-1\zeta\beta}(\bar{n}_i,\bar{n}_j) \\ &= E^{\alpha\gamma}\mathfrak{B}^{\mathrm{FKP}}_{\gamma\epsilon}(\bar{n}_i,\bar{n}_j)\mathfrak{B}^{-1\epsilon\beta}(\bar{n}_i,\bar{n}_j),\end{aligned} \tag{107}$$

where $\mathfrak{C}^{\mathrm{FKP}}$ is the FKP covariance, equation (42), and $\mathfrak{B}^{\mathrm{FKP}} \equiv H^{-1/2}\mathfrak{C}^{\mathrm{FKP}}H^{-1/2}$ is its pre-whitened counterpart, equation (66). Note that going from equation (106) to the second expression in equation (107) includes, as part of the FKP approximation, the assumption that $\bar{n}_k$ and $\bar{n}_l$ in $\mathfrak{B}^{-1\eta\beta}(\bar{n}_k,\bar{n}_l)$ are approximately constant. The expressions on the right-hand side of equation (107) are not symmetric in $\alpha\beta$, because of the asymmetry of the FKP approximation (41). To the further extent that the pre-whitened

covariance $\mathfrak{B}^{\rm FKP}$ equals the approximation $\mathfrak{B}$, equation (80), the covariance $\langle\Delta\hat{Z}^\alpha\Delta\hat{Z}^\beta\rangle$ reduces to the asymmetric matrix E given by equation (83):

$$\langle\Delta\hat{Z}^\alpha\Delta\hat{Z}^\beta\rangle = E^{\alpha\beta}, \tag{108}$$

the asymmetry of the right-hand side being inherited from the FKP approximation. An equally good approximation to the covariance would be the same expression (108) with the indices swapped on the right-hand side, $\alpha \leftrightarrow \beta$. Thus it seems reasonable to conclude that the covariance $\langle\Delta\hat{Z}^\alpha\Delta\hat{Z}^\beta\rangle$ should be approximately equal to the the symmetrized Fisher matrix $E^{(\alpha\beta)}$:

$$\langle\Delta\hat{Z}^\alpha\Delta\hat{Z}^\beta\rangle = E^{(\alpha\beta)} \equiv \mathop{\rm Sym}_{(\alpha\beta)} E^{\alpha\beta}. \tag{109}$$

Several comments can be made about the accuracy of the approximations made in the above derivation.

First, one partial test of the validity of the FKP approximation is the degree of asymmetry of the asymmetric Fisher matrix $E^{\alpha\beta}$, equation (83). If the survey is broad in real space, which is the condition for the FKP approximation to hold, then the pair integral $R(r;\mu_\alpha)$ in the integrand on the right-hand side of equation (83) will be a slowly varying function of pair separation r, so that the matrix $E^{\alpha\beta}$ will be nearly diagonal, hence symmetric. The test is not definitive because $E^{\alpha\beta}$ would be symmetric in any case if $\mu_\alpha = \mu_\beta$. In practice, however, $\mu_\alpha \approx \xi(k_\alpha)$ in both linear and non-linear regimes (Fig. 7), and realistically the power spectrum $\xi(k)$ varies substantially, so the consistency test should be indicative.

Secondly, one of the weaknesses of the FKP approximation is that it fails to deal with sharp edges – as typically occur at the angular boundaries of a survey – correctly. The FKP approximation tends to overestimate the variance contributed by regions near boundaries, since it assumes that those regions are accompanied by more correlated neighbours than is actually the case. Thus, at least as regards edge effects, the FKP approximate covariance, equation (107), should tend to overestimate the exact covariance, equation (106), of the approximate estimate $\hat{Z}^\alpha$.

Thirdly, it is possible to check the accuracy of the approximation made in going from equation (107) to equation (108). The approximation involves setting $\mathfrak{B}^{\rm FKP}\mathfrak{B}^{-1} = \mathbf{1}$, whereas comparing equation (66) for $\mathfrak{B}^{\rm FKP}$ with the approximation (80) for $\mathfrak{B}$ shows that this quantity is in fact, in the representation of eigenfunctions ϕ_α of the four-point matrix M,

$$\mathfrak{B}^{\rm FKP}_{\alpha\gamma}(\bar{n}_i,\bar{n}_j)\mathfrak{B}^{-1\gamma\beta}(\bar{n}_i,\bar{n}_j) = \mathbf{1}^\beta_\alpha + \frac{(\bar{n}_i^{-1}+\bar{n}_j^{-1})}{(\mu_\alpha+\bar{n}_i^{-1})(\mu_\alpha+\bar{n}_j^{-1})}(L^\beta_\alpha - \mu_\alpha\mathbf{1}^\beta_\alpha). \tag{110}$$

The correction term on the right-hand side of equation (110) should be small to the extent that the three-point matrix $L_{\alpha\beta}$ is near-diagonal in this four-point representation, with eigenvalues $\lambda_\alpha \approx \mu_\alpha$, as was found to be the case in Section 4.

If desired, one could use the expression on the right-hand side of equation (110) to compute a more accurate approximation to the covariance of $\hat{Z}^\alpha$, based on equation (107) rather than on equation (108). However, if one were willing to go to the trouble of computing a correction from equation (110), then one would probably be willing to revert to equation (100), and to integrate $\mathfrak{B}^{-1\alpha\beta}(\bar{n}_i,\bar{n}_j)$ numerically over all pairs of volume elements ij in the survey, inverting $\mathfrak{B}$ numerically for each pair $\bar{n}_i$, $\bar{n}_j$ of values of the selection function. This latter procedure is in fact the gourmet recipe of Section 9.1.

Fourthly, the approximation $\lambda_\alpha \approx \mu_\alpha$ adopted in the approximation (80) to $\mathfrak{B}$ tends to overestimate the true eigenvalues λ_α of the three-point matrix L, according to Fig. 8. This should lead to a slight overestimate of the variance. In the realistic ΛCDM case, Fig. 8, the approximation $\lambda_\alpha \approx \mu_\alpha$ overestimates the true eigenvalues λ_α by at worst 20 per cent, at moderately non-linear wavenumbers k. This 20 per cent overestimate is diluted to at worst 10 per cent because the three-point variance contributes at most half of the combined two-point, three-point and four-point variance, where the selection function satisfies $\bar{n}^{-1} = \mu_\alpha$. The overestimate is further diluted because in practice the selection function varies, and is unlikely to sit everywhere near the worst value.

The conclusion is that the covariance of the approximate estimator $\hat{Z}_\alpha$, equation (101) or (105), should be given approximately (equation 109) by the symmetrized Fisher matrix $E^{(\alpha\beta)}$ of equation (83), and that if anything this covariance is likely to be on the conservative side of the true covariance.

7.4 The covariance of $\hat{Y}_\alpha$

From the expression (109) for the covariance of $\hat{Z}^\alpha$, one might conclude (falsely) that the covariance of $\hat{Y}_\alpha$, equation (99), is

$$\langle\Delta\hat{Y}_\alpha\Delta\hat{Y}_\beta\rangle = E^{-1}_{\alpha\gamma}E^{(\gamma\delta)}E^{-1}_{\beta\delta}. \tag{111}$$

A more direct derivation of the covariance of $\hat{Y}_\alpha$, along the lines of equations (106)–(109), leads to the same (false) conclusion. The analogue of equation (108) is

$$\langle\Delta\hat{Y}_\alpha\Delta\hat{Y}_\beta\rangle = E^{-1}_{\beta\alpha} \tag{112}$$

with the asymmetric matrix E on the right-hand side. At this point one might be inclined to symmetrize this equation (112), as was done for $\langle\Delta\hat{Z}^\alpha\Delta\hat{Z}^\beta\rangle$ in equation (109), writing

$$\langle\Delta\hat{Y}_\alpha\Delta\hat{Y}_\beta\rangle = \mathop{\rm Sym}_{(\alpha\beta)} E^{-1}_{\alpha\beta}. \tag{113}$$

The symmetrized inverse $\mathrm{Sym}_{(\alpha\beta)}E^{-1}_{\alpha\beta}$ of the asymmetric Fisher matrix is to be distinguished from the inverse $E^{-1}_{(\alpha\beta)}$ of the symmetrized Fisher matrix. However, it is not hard to show that

$$\mathop{\rm Sym}_{(\alpha\beta)} E^{-1}_{\alpha\beta} = E^{-1}_{\alpha\gamma}E^{(\gamma\delta)}E^{-1}_{\beta\delta}. \tag{114}$$

Thus equations (111) and (113) are identical. However, both equations are wrong.

The problem is that, while the Fisher matrix E remains well-behaved in the presence of loud noise, with near-zero eigenvalues, its inverse E^{-1} becomes almost singular. Consider the example of some noisy mode, for which the eigenvalue of the Fisher matrix is almost zero. It may well happen that the asymmetric Fisher matrix $E^{\alpha\beta}$ is numerically non-singular, but that, because of approximations or numerics, the computed eigenvalue of the symmetrized Fisher matrix $E^{(\alpha\beta)}$ is exactly zero. Equation (111) would then say that the variance of the noisy mode is zero [for if the determinant of the symmetrized Fisher matrix is zero, $|E^{(\alpha\beta)}| = 0$, while the determinant of the asymmetric Fisher matrix is finite, $|E^{\alpha\beta}| \neq 0$, then the determinant of the variance in equation (111) is zero]. This is plainly absurd.

It is safer to take the covariance of $\hat{Y}_\alpha$ to be approximately equal to the inverse of the symmetrized Fisher matrix $E^{(\alpha\beta)}$,

$$\langle\Delta\hat{Y}_\alpha\Delta\hat{Y}_\beta\rangle = E^{-1}_{(\alpha\beta)}. \tag{115}$$

Here a noisy mode will always reveal itself by its small eigenvalue.

7.5 Convert to $\hat{X}_\alpha$

For the purpose of constructing uncorrelated quantities to be plotted on a graph, it is desirable to compute the pre-whitened power spectrum $\hat{X}_\alpha$.

To compute $\hat{X}_\alpha$, start from the estimate $\hat{Z}_\alpha$ given by equation (105), transform this into $\hat{Y}_\alpha = E^{-1}_{\alpha\beta}\hat{Z}^\beta$, equation (99), thence into the power spectrum $\hat{\xi}_\alpha = (H^{-1/2})^\beta_\alpha \hat{Y}_\beta$, equation (74), and thence into the pre-whitened power spectrum $\hat{X}_\alpha$, equation (70).

The covariance of the pre-whitened power $\hat{X}_\alpha$ is, by construction, the same as that of $\hat{Y}_\alpha$, equation (115),

$$\langle \Delta\hat{X}_\alpha \Delta\hat{X}_\beta \rangle = E^{-1}_{(\alpha\beta)}, \tag{116}$$

the inverse of the symmetrized Fisher matrix of the pre-whitened power.

The estimator $\hat{X}_\alpha$ of pre-whitened power, equation (70), is a non-linear transformation of the estimator $\hat{\xi}_\alpha$ of power, and is therefore biased if $\hat{\xi}_\alpha$ is unbiased. However, the estimator $\hat{X}_\alpha$ is unbiased in the asymptotic limit of a large quantity of data.

7.6 Decorrelate

One final step remains, which is to process the measured pre-whitened power spectrum $\hat{X}_\alpha$ into a set of decorrelated band-powers. How to accomplish such decorrelation is described in Paper IV.

One possibility would be to decorrelate the power spectrum $\hat{\xi}(k)$ itself. This is a bad idea, because the power spectrum is highly correlated in the non-linear regime, so the decorrelation matrices would be broad, with large negative off-diagonal entries, making it impossible to interpret the decorrelated band-powers as representing the power spectrum over some narrow band.

Another possibility would be to decorrelate the pre-whitened power $\hat{X}_\alpha$ not in Fourier space but rather in the representation of eigenfunctions ϕ_α of the pre-whitened four-point matrix M. Again this seems not so good an idea, in the first place because the physical meaning of this representation is obscure, and in the second place because the eigenfunctions can mix where their eigenvalues μ_α are degenerate. Since $\mu_\alpha \approx \xi(k_\alpha)$, such mixing in practice occurs between wavenumbers k_α where the power $\xi(k_\alpha)$ is the same, which happens to either side of the peak in the power spectrum, Fig. 10. Perhaps in the future a better understanding of the eigenfunctions ϕ_α will emerge, amongst other things allowing mixing to avoided, but in the meantime these problems remain.

The natural solution is to decorrelate the pre-whitened power $\hat{X}(k)$ in Fourier space. As seen in Section 4, the covariance of the pre-whitened power is encouragingly narrow in Fourier space, narrow enough that the decorrelation matrices will be narrow, so that the decorrelated band-powers can be interpreted as estimates of the pre-whitened power over narrow intervals of wavenumber k. In contrast to the pre-whitened power $\hat{X}_\alpha$ in the ϕ_α-representation, the pre-whitened power $\hat{X}(k)$ in Fourier space has a clear interpretation, and there is no problem arising from mixing of eigenfunctions.

8 THE FULL FKP

Sections 6 and 7 invoked not only the FKP approximation, but also the simplifying approximation (80) to the pre-whitened reduced covariance $\mathfrak{B}$. How much more work would it take to compute the minimum variance estimator and Fisher matrix of non-linear power making only the FKP approximation and no other approximation? The question is of both didactic and practical interest.

8.1 Fisher matrix

The FKP approximation to the Fisher matrix of the power spectrum, equation (45), looks simplest expressed in real space:

$$F(r_\alpha, r_\beta) = \int \mathfrak{C}^{-1}(r_\alpha, r_\beta; \bar{n}_i, \bar{n}_j)\delta_{3D}(r_{ij} - r_\beta)\,\mathrm{d}^3 r_i\,\mathrm{d}^3 r_j. \tag{117}$$

The corresponding expression for the FKP approximation to the Fisher matrix of the pre-whitened power spectrum, equation (82), is

$$E(r_\alpha, r_\beta) = \int \mathfrak{B}^{-1}(r_\alpha, r_\beta; \bar{n}_i, \bar{n}_j)\delta_{3D}(r_{ij} - r_\beta)\,\mathrm{d}^3 r_i\,\mathrm{d}^3 r_j. \tag{118}$$

These are five-dimensional (thanks to the Dirac delta-function) integrals over pairs of volume elements ij separated by $r_{ij} \equiv |\boldsymbol{r}_i - \boldsymbol{r}_j| = r_\beta$ in the survey. The integrals are actually quite doable. If, as is typical, the selection function $\bar{n}(\boldsymbol{r})$ separates into the product of an angular mask and a radial selection function, then the three-dimensional angular integrals can be done analytically (Hamilton 1993), leaving a double integral of $\mathfrak{C}^{-1}(r_\alpha, r_\beta; \bar{n}_i, \bar{n}_j)$ or $\mathfrak{B}^{-1}(r_\alpha, r_\beta; \bar{n}_i, \bar{n}_j)$ over the radial directions. The matrices $\mathfrak{C}(r_\alpha, r_\beta; \bar{n}_i, \bar{n}_j)$ or $\mathfrak{B}(r_\alpha, r_\beta; \bar{n}_i, \bar{n}_j)$, discretized (Section 2.3) over a grid of separations r_α and r_β, can be inverted numerically for each pair of values of the selection functions $\bar{n}_i$ and $\bar{n}_j$.

The problem with equations (117) and (118) is that experience (Sections 4.1, 6.2) suggests that discretization of the matrix $\mathfrak{C}(r_\alpha, r_\beta; \bar{n}_i, \bar{n}_j)$ or $\mathfrak{B}(r_\alpha, r_\beta; \bar{n}_i, \bar{n}_j)$ in real space is liable to introduce ringing and aliasing in Fourier space, defeating the aim of constructing an accurate Fisher matrix of the power spectrum.

A possibly more robust procedure would be to follow more closely the program described in Sections 6 and 7. In Fourier space, the FKP approximation to the pre-whitened Fisher matrix, equation (82), is

$$E(k_\alpha, k_\beta) = \int \mathfrak{B}^{-1}(k_\alpha, k_\gamma; \bar{n}_i, \bar{n}_j) j_0(k_\gamma r_{ij}) j_0(k_\beta r_{ij})\,\mathrm{d}^3 r_i\,\mathrm{d}^3 r_j \frac{4\pi k_\gamma^2\,\mathrm{d}k_\gamma}{(2\pi)^3}. \tag{119}$$

This integral might be evaluated as follows (since I have not actually carried through this program, I cannot say for sure that it would work without a hitch). First, compute the matrix of pair integrals

$$R(r; k_\alpha, k_\beta) = \int \mathfrak{B}^{-1}(k_\alpha, k_\beta; \bar{n}_i, \bar{n}_j)\delta_{3D}(r_{ij} - r)\,\mathrm{d}^3 r_i\,\mathrm{d}^3 r_j \tag{120}$$

for many pair separations $r_{ij} = r$. These pair integrals $R(r; k_\alpha, k_\beta)$, equation (120), are analogous to the FKP-weighted pair integrals $R(r; \mu_\alpha)$, equation (84). Next, cosine-transform (e.g. with FFTLOG) the pair integrals

$$\tilde{R}(k; k_\alpha, k_\beta) = 2\int_0^\infty \cos(kr) R(r; k_\alpha, k_\beta)\,\mathrm{d}r \tag{121}$$

analogously to $\tilde{R}(k; \mu_\alpha)$, equation (87). Finally, compute the

pre-whitened Fisher matrix $E(k_\alpha, k_\beta)$ by integrating

$$E(k_\alpha, k_\beta) = \frac{1}{2\pi k_\beta}\int_0^\infty [\tilde{R}(k_\gamma - k_\beta; k_\alpha, k_\gamma) - \tilde{R}(k_\gamma + k_\beta; k_\alpha, k_\gamma)]k_\gamma \,\mathrm{d}k_\gamma. \quad (122)$$

In practice, the matrix $\mathfrak{B}(k_\alpha, k_\beta; \bar{n}_i, \bar{n}_j)$ in equation (120) must be inverted on a discrete grid of wavenumbers k. Similarly, the integral over k_γ in equation (122) should be done as a discrete sum. Specifically, if the matrices are discretized (Section 2.3) on a logarithmic grid of wavenumbers, so that $R(k; k_\alpha, k_\beta)$ discretizes to $\mathbf{R}_{k_\alpha k_\beta}(k) = R(k; k_\alpha, k_\beta)4\pi(k_\alpha k_\beta)^{3/2}\Delta \ln k/(2\pi)^3$, and the Fisher matrix $E(k_\alpha, k_\beta)$ discretizes to $\mathbf{E}_{k_\alpha k_\beta} = E(k_\alpha, k_\beta)4\pi(k_\alpha k_\beta)^{3/2}\Delta \ln k/(2\pi)^3$, then equation (122) becomes

$$\mathbf{E}_{k_\alpha k_\beta} = \sum_{k_\gamma} \frac{(k_\beta k_\gamma)^{1/2}}{2\pi}[\tilde{\mathbf{R}}_{k_\alpha k_\gamma}(k_\gamma - k_\beta) - \tilde{\mathbf{R}}_{k_\alpha k_\gamma}(k_\gamma + k_\beta)]. \quad (123)$$

It may be anticipated that, as in Section 6.2, equation (123) will tend to overestimate the diagonal elements $\mathbf{E}_{k_\alpha k_\alpha}$ if the gridding of the matrix is too coarse to resolve the diagonal properly. Integrating the continuous Fisher matrix $E(k_\alpha, k_\beta)$, equation (122), over k_β yields

$$\int_0^\infty E(k_\alpha, k_\beta)k_\beta \,\mathrm{d}k_\beta = \frac{1}{\pi}\int_0^\infty\int_0^{k_\gamma} \tilde{R}(k; k_\alpha, k_\gamma)\,\mathrm{d}k\, k_\gamma \,\mathrm{d}k_\gamma. \quad (124)$$

Discretized, equation (124) becomes

$$\sum_\beta \left(\frac{k_\beta}{k_\alpha}\right)^{1/2}\mathbf{E}_{k_\alpha k_\beta} = \frac{1}{\pi}\sum_{k_\gamma}\left(\frac{k_\gamma}{k_\alpha}\right)^{1/2}\int_0^{k_\gamma}\tilde{\mathbf{R}}_{k_\alpha k_\gamma}(k)\,\mathrm{d}k. \quad (125)$$

The integral over k on the right-hand side of equation (125) can be done conveniently as a sine transform (e.g. with FFTLOG) of the pair integral

$$\int_0^{k_\gamma}\tilde{\mathbf{R}}_{k_\alpha k_\gamma}(k)\,\mathrm{d}k = 2\int_0^\infty \sin(kr)\frac{\mathbf{R}_{k_\alpha k_\gamma}(r)}{r}\mathrm{d}r. \quad (126)$$

If the sum on the left-hand side of equation (125) exceeds the right-hand side, then reduce the diagonal element $\mathbf{E}_{k_\alpha k_\alpha}$ so that the sum is satisfied. It is fine to evaluate the sum on the right-hand side of equation (125) as a discrete sum over k_γ, rather than as a continuous integral, because $\mathbf{R}_{k_\alpha k_\gamma}(k)$, equation (120), inherits its behaviour from $\mathfrak{B}_{k_\alpha k_\gamma}$, which, if constructed equation (66) from the four-point and three-point matrices M and L as discussed in Section 4.1, should behave correctly near the diagonal even if the resolution is too coarse to resolve the diagonal.

8.2 Estimate of power

The FKP approximation to the minimum variance estimator of the power spectrum, equation (44), again looks simplest when expressed in real space:

$$\hat{\xi}_\alpha = F^{-1}_{\alpha\beta}\hat{T}^\beta \quad (127)$$

with

$$\hat{T}(r_\alpha) = \int \mathfrak{C}^{-1}(r_\alpha, r_{ij}; \bar{n}_i, \bar{n}_j)\delta(\boldsymbol{r}_i)\delta(\boldsymbol{r}_j)\,\mathrm{d}^3 r_i\,\mathrm{d}^3 r_j. \quad (128)$$

As usual, the pre-whitened estimator $\hat{Y} \equiv H^{-1/2}\hat{\xi}$, equation (74), is related to the estimator $\hat{Z}$ by $\hat{Y} = E^{-1}\hat{Z}$, equation (99). The FKP approximation to the estimator $\hat{Z}$ is (equation 100)

$$\hat{Z}(r_\alpha) = \int \frac{\mathfrak{B}^{-1}(r_\alpha, r_{ij}; \bar{n}_i, \bar{n}_j)\delta(\boldsymbol{r}_i)\delta(\boldsymbol{r}_j)}{[1 + \xi(r_{ij})]^{1/2}}\mathrm{d}^3 r_i\,\mathrm{d}^3 r_j. \quad (129)$$

Equations (128) and (129) are six-dimensional integrals over pairs of volume elements ij in the survey. Once again, however, one may anticipate that discretization of the matrices $\mathfrak{C}(r_\alpha, r_\beta; \bar{n}_i, \bar{n}_j)$ or $\mathfrak{B}(r_\alpha, r_\beta; \bar{n}_i, \bar{n}_j)$ in real space would introduce ringing and aliasing in Fourier space, defeating the aim of constructing an accurate estimator of the power spectrum.

Again, it seems likely that it would be more robust to work with pre-whitened quantities in Fourier space. In Fourier space, the FKP approximation to the estimator $\hat{Z}_\alpha$ is (equation 100)

$$\hat{Z}(k_\alpha) = \int \frac{\mathfrak{B}^{-1}(k_\alpha, k_\beta; \bar{n}_i, \bar{n}_j)j_0(k_\beta r_{ij})\delta(\boldsymbol{r}_i)\delta(\boldsymbol{r}_j)}{[1 + \xi(r_{ij})]^{1/2}}\mathrm{d}^3 r_i\,\mathrm{d}^3 r_j\,\frac{4\pi k_\beta^2\,\mathrm{d}k_\beta}{(2\pi)^3}. \quad (130)$$

One way to evaluate this integral might be as follows. First, compute the matrix $\hat{S}(r; k_\alpha, k_\beta)$ of integrals over pairs of overdensities $\delta(\boldsymbol{r}_i)\delta(\boldsymbol{r}_j)$ at many separations $r_{ij} = r$:

$$\hat{S}(r; k_\alpha, k_\beta) = \int \mathfrak{B}^{-1}(k_\alpha, k_\beta; \bar{n}_i, \bar{n}_j)\delta_{3\mathrm{D}}(r_{ij} - r)\delta(\boldsymbol{r}_i)\delta(\boldsymbol{r}_j)\,\mathrm{d}^3 r_i\,\mathrm{d}^3 r_j, \quad (131)$$

which may be compared to equation (102). Next, pre-whiten (compare equation 103)

$$\hat{Z}(r; k_\alpha, k_\beta) = \frac{\hat{S}(r; k_\alpha, k_\beta)}{[1 + \xi(r)]^{1/2}} \quad (132)$$

and Fourier transform, e.g. with FFTLog (compare equation 104):

$$\hat{Z}(k; k_\alpha, k_\beta) = \int_0^\infty j_0(kr)\hat{Z}(r; k_\alpha, k_\beta)4\pi r^2\,\mathrm{d}r. \quad (133)$$

Actually it suffices to do this Fourier transform for $k = k_\beta$ only. Finally, the estimator $\hat{Z}(k_\alpha)$, equation (130), is

$$\hat{Z}(k_\alpha) = \int_0^\infty \hat{Z}(k_\beta; k_\alpha, k_\beta)\frac{4\pi k_\beta^2\,\mathrm{d}k_\beta}{(2\pi)^3}. \quad (134)$$

The integral in equation (134) should be done as a discrete sum. If discretized (Section 2.3) on a logarithmic grid of wavenumbers, so that $\hat{Z}(k_\alpha)$ discretizes to $\hat{\mathbf{Z}}_{k_\alpha} = \hat{Z}(k_\alpha)[4\pi k_\alpha^3\Delta \ln k/(2\pi)^3]^{1/2}$ and $\hat{Z}(k; k_\alpha, k_\beta)$ discretizes to $\hat{\mathbf{Z}}_{k_\alpha k_\beta}(k) = \hat{Z}(k; k_\alpha, k_\beta)4\pi(k_\alpha k_\beta)^{3/2}\Delta \ln k/(2\pi)^3$, then equation (134) is

$$\hat{\mathbf{Z}}_{k_\alpha} = \sum_{k_\beta}\hat{\mathbf{Z}}_{k_\alpha k_\beta}(k_\beta)\left[\frac{4\pi k_\beta^3\Delta \ln k}{(2\pi)^3}\right]^{1/2}. \quad (135)$$

9 RECIPES

This section summarizes the results of previous sections into logical sequences of practical steps needed to estimate the pre-whitened non-linear power spectrum from an actual galaxy survey. The end result is a set of uncorrelated pre-whitened non-linear band-powers with error bars, over some prescribed grid of wavenumbers k.

There are three versions of the recipe, gourmet (Section 9.1), fine (Section 9.2), and fastfood (Section 9.3). All the methods use the FKP approximation, equation (41). Thus one should imagine

that there is also a *haute cuisine* method, which might be brute-force, or might be some clever procedure that apodizes edges.

First a disclaimer. *The methods described herein do not take into account redshift distortions, the effects of which on the power spectrum are at least as great as those of non-linearity. There is no point in using these methods as they stand, without also taking into account redshift distortions*. However, given that a full-blown procedure including redshift distortions may well be based in part on the methods described, it seems worthwhile to lay out the steps required to implement them.

9.1 Gourmet

This version of the recipe is conceptually the simplest, but it takes the most computing power (a supercomputer would be handy). The procedure is a direct implementation of the FKP approximation to the minimum variance estimator of pre-whitened power and the associated Fisher matrix, as described in Section 8.

Naturally, if one were going to the trouble of using the gourmet recipe, then one would want to use the best possible model of the three-point and four-point correlation functions, not just the hierarchical model with constant amplitudes.

Steps 1 and 2 below require knowledge of the selection function of a survey, but no actual data. Steps 3–5 require actual data from a galaxy survey.

Step 1. Compute the FKP approximation to the asymmetric Fisher matrix $E^{\alpha\beta}$ of the pre-whitened non-linear power spectrum, as described in Section 8.1, equations (120)–(122). Equation (120) involves a five-dimensional integral over pairs ij of volume elements separated by $r_{ij} = r$ in the survey. If the selection function $\bar{n}(\boldsymbol{r})$ separates into the product of an angular mask and a radial selection function, then the three-dimensional angular integrals can be done analytically (Hamilton 1993), leaving a double integral of $\mathfrak{B}^{-1}(k_\alpha, k_\beta; \bar{n}_i, \bar{n}_j)$ over the radial directions.

Step 2. The covariance matrix $\langle \Delta\hat{X}_\alpha \Delta\hat{X}_\beta \rangle$ of the pre-whitened power is equal, equation (116), to the inverse $E^{-1}_{(\alpha\beta)}$ of the symmetrized Fisher matrix. Use this covariance matrix to construct decorrelation matrices W, as described in Paper IV, with the property that $W\langle \Delta\hat{X}\Delta\hat{X}^{\mathrm{T}} \rangle W^{\mathrm{T}}$ is diagonal in Fourier space (cf. Paper IV equation (20)). The diagonal elements of this diagonal matrix are the expected variances of the decorrelated band-powers $\hat{B} = W\hat{X}$ to be computed in step 5.

Step 3. Compute the estimator $\hat{Z}_\alpha$ as described in Section 8.2, equations (131)–(134).

Step 4. Transform $\hat{Z}^\alpha$ into the pre-whitened power $\hat{X}_\alpha$ using equations (99), (74) and (70), as stated in Section 7.5.

Step 5. Decorrelate the estimated pre-whitened power spectrum $\hat{X}$ into a set of uncorrelated band-powers $\hat{B} = W\hat{X}$, using the decorrelation matrices W computed in step 2. Bear in mind that, as usual in ML fitting, the error bars should of course be interpreted as being attached to the model, the prior band-powers B, rather than to the data, the estimated band-powers $\hat{B}$.

9.2 Fine

This method adopts the approximation made in Sections 6 and 7 that the pre-whitened reduced covariance matrix $\mathfrak{B}$ takes the simplified form (80). According to the results of Section 4, this approximation to $\mathfrak{B}$ should be quite good. If it is, then the fine method should yield results close to the gourmet method of Section 9.1, at a considerable saving in computer time.

Steps 1–5 below do not require any actual data; the steps can be used to determine in advance how well the pre-whitened power spectrum might be measured from a survey. Steps 6–9 require actual data from a galaxy survey.

Step 1. Compute a table of FKP-weighted pair integrals $R(r; \mu)$ at many separations r and several FKP constants μ. Calculating the pair integrals $R(r; \mu)$ requires knowing the selection function $\bar{n}(\boldsymbol{r})$ of a galaxy survey, but does not require actual data. This pair integral, commonly denoted $\langle RR \rangle$, is commonly computed by Monte Carlo integration, but I find it faster, more accurate and more convenient (since the program is already written) to compute the integral directly, using the procedures described by Hamilton (1993).

Step 2. Compute the pre-whitened four-point contribution $M \equiv H^{-1/2} K H^{-1/2}$ to the reduced covariance of the non-linear power spectrum. This involves adopting a prior power spectrum $\xi(k)$, and a model of the four-point correlation function η_{ijkl}. For the hierarchical model, the covariance matrices K and H are given by equations (55) and (57). Some numerical issues concerning the computation of the matrix M are discussed in Section 4.1.

Step 3. Compute the eigenfunctions ϕ_α and eigenvalues μ_α^2, equation (67), by diagonalizing the pre-whitened four-point matrix M.

Step 4. Compute the asymmetric Fisher matrix $E^{\alpha\beta}$, equation (83), of the pre-whitened non-linear power spectrum of the survey, in the representation of eigenfunctions ϕ_α of the pre-whitened four-point matrix M. This is where the pair integral $R(r, \mu)$ computed in step 1 is needed. Numerical issues are discussed in Section 6.2.

Step 5. This is the same as step 2 of the gourmet method: from the inverse $E^{-1}_{(\alpha\beta)}$ of the symmetrized Fisher matrix, construct decorrelation matrices W such that the covariance of the band-powers $\hat{B} = W\hat{X}$ is diagonal. Decorrelation is the subject of Paper IV.

Step 6. Compute a table of FKP-weighted pair densities $\hat{S}(r; \mu)$, equation (102), at many separations r and several FKP constants μ. Calculating $\hat{S}(r; \mu)$, commonly denoted $\langle DD \rangle - 2\langle DR \rangle + \langle RR \rangle$, requires actual data from a survey.

Step 7. From $\hat{S}(r; \mu_\alpha)$, compute the estimate $\hat{Z}^\alpha$, equation (105), in the representation of eigenfunctions ϕ_α, as described in Section 7.2.

Step 8. This is the same as step 4 of the gourmet method: transform $\hat{Z}^\alpha$ into the pre-whitened power $\hat{X}_\alpha$ using equations (99), (74) and (70), as stated in Section 7.5. The transformations may be done in whatever representation proves most convenient or numerically reliable. Ultimately, one wants the pre-whitened power spectrum $\hat{X}(k)$ in Fourier space.

Step 9. This is the same as step 5 of the gourmet method: decorrelate the pre-whitened power spectrum $\hat{X}$ into a set of uncorrelated band-powers $\hat{B} = W\hat{X}$, using the decorrelation matrices W computed in step 5 above.

9.3 Fastfood

For some purposes a simplified, approximate version of the procedure in Section 9.2 may be considered adequate.

The basic simplifying approximation here is that the covariance $\langle \Delta\hat{X}(k_\alpha)\Delta\hat{X}(k_\beta) \rangle$ of the pre-whitened power spectrum may be considered to be diagonal in Fourier space without further refinement. The procedure then becomes the same as the FKP procedure for Gaussian fluctuations, with the differences that (i) it is the pre-whitened power spectrum $\hat{X}(k)$, equation (69), rather

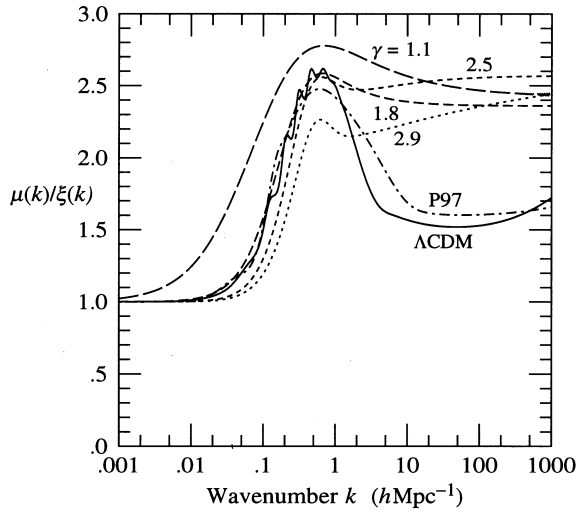

Figure 11. Ratio $\mu(k)/\xi(k)$ of the effective FKP constant $\mu(k)$ to the non-linear power spectrum $\xi(k)$, as a function of wavenumber k, for several different power spectra. The ratios should be regarded as indicative rather than definitive, because they depend on the validity of the hierarchical model (see text). The numbered curves are for power-law power spectra with correlation functions $\xi(r) = (r/5\,h^{-1}\,\mathrm{Mpc})^{-\gamma}$, the number label being the index γ. The curve labelled P97 is for the $\Omega_\mathrm{m} = 0.3$ power spectrum derived from observations by Peacock (1997), while that labelled ΛCDM is for the ΛCDM power spectrum of Eisenstein & Hu (1998).

than the power spectrum $\hat{\xi}(k)$ that is being estimated; and (ii) the FKP constants $\mu(k)$ in the FKP pair-weightings are modified from the Gaussian case where $\mu(k) = \xi(k)$.

Fig. 11 shows the ratio $\mu(k)/\xi(k)$ of the effective FKP constant $\mu(k)$ to the non-linear power spectrum $\xi(k)$ for several different power spectra. The ratios plotted in Fig. 11 should be regarded as indicative rather than definitive, because they depend on the validity of the hierarchical model with constant amplitudes $R_b = -R_a$ (as required to satisfy the Schwarz inequality, Section 3.2), which as discussed in Section 4.2 is certainly wrong at some level.

The effective FKP constant $\mu(k)$ shown in Fig. 11 is *not* the same thing as the eigenvalue μ_α of M at the nominal wavenumber k_α shown in Fig. 7. As discussed in Section 4.4, the correspondence between eigenvalue μ_α and nominal wavenumber k_α is precise only for Gaussian fluctuations.

The effective FKP constant $\mu(k)$ in Fig. 11 was calculated by going through steps 2–5 of the recipe in Section 9.2 for the case of a perfect, noiseless ($\bar{n} \to \infty$) survey of (large) volume V. In this case the Fisher matrix $E_{\alpha\beta}$ of the pre-whitened power, equation (83), reduces to

$$E_{\alpha\beta} = \frac{V}{2} M^{-1}_{\alpha\beta} \tag{136}$$

the inverse of which gives the covariance of the pre-whitened power $\hat{X}$:

$$\langle \Delta\hat{X}_\alpha \Delta\hat{X}_\beta \rangle = \frac{2}{V} M_{\alpha\beta}. \tag{137}$$

Decorrelating this covariance, equation (137), in Fourier space, as described by in Paper IV, yields band-powers $\hat{B}(k)$ the covariance of which is by construction diagonal. The diagonal values of the diagonal covariance matrix of the band-powers can be taken to define the effective FKP constants $\mu(k)$:

$$\langle \Delta\hat{B}(k_\alpha) \Delta\hat{B}(k_\beta) \rangle = \mathbf{1}_{\alpha\beta} \frac{2\,\mu(k_\alpha)^2}{V}. \tag{138}$$

With the effective FKP constants $\mu(k)$ taken as given by Fig. 11, the shortcut recipe is then as follows.

Step 1. Compute the effective spatial volume $V(k)$ of the survey for modes at wavenumber k:

$$V(k) = 2\,\mu(k)^2 R[0; \mu(k)] = \int \frac{\mathrm{d}^3 r}{[1 + \bar{n}(\boldsymbol{r})^{-1} \mu(k)^{-1}]^2}. \tag{139}$$

Step 2. The Fisher matrix (83) of the pre-whitened power reduces to

$$E(k_\alpha, k_\beta) = \frac{1}{2} \int_0^\infty j_0(k_\alpha r) j_0(k_\beta r) R[r; \mu(k_\alpha)] 4\pi r^2\,\mathrm{d}r. \tag{140}$$

If the pre-whitened power is averaged over sufficiently broad shells in k-space, then, by arguments similar to those in Section 6.3, the Fisher matrix is approximately diagonal (compare equation 95):

$$E(k_\alpha, k_\beta) \approx (2\pi)^3 \delta_{3\mathrm{D}}(k_\alpha - k_\beta) \frac{V(k_\alpha)}{2\,\mu(k_\alpha)^2}, \tag{141}$$

where $V(k_\alpha)$ is the effective volume given by equation (139). For the approximation (141) to be valid, the shells in k-space must be broad not only compared with the inverse scale of the survey (as in the Gaussian case), but also compared with the width of the four-point matrix M plotted in Figs 3 and 4. In the large-k, hierarchical limit, the width of the matrix M in k-space is comparable to an inverse correlation length, $\Delta k \sim \pi/r_0$.

Step 3. The covariance $\langle \Delta\hat{X}(k_\alpha) \Delta\hat{X}(k_\beta) \rangle$ of the pre-whitened power equals the inverse of the Fisher matrix $E(k_\alpha, k_\beta)$ given by equation (141):

$$\langle \Delta\hat{X}(k_\alpha) \Delta\hat{X}(k_\beta) \rangle \approx (2\pi)^3 \delta_{3\mathrm{D}}(k_\alpha - k_\beta) \frac{2\,\mu(k_\alpha)^2}{V(k_\alpha)}. \tag{142}$$

Define pre-whitened band-powers $\hat{B}(k)$ to be the pre-whitened power spectrum $\hat{X}(k)$ averaged over broad (as in step 2) shells of volume V_k about k,

$$\hat{B}(k) \equiv V_k^{-1} \int \hat{\xi}(k)\,\mathrm{d}V_k, \tag{143}$$

where $\mathrm{d}V_k \equiv 4\pi k^2\,\mathrm{d}k/(2\pi)^3$. The variance of the shell-averaged pre-whitened band-powers is

$$\langle \Delta\hat{B}(k)^2 \rangle \approx \frac{2\,\mu(k_\alpha)^2}{V(k) V_k}, \tag{144}$$

which is $2\,\mu(k_\alpha)^2$ divided by the effective phase volume, the product of the effective spatial volume $V(k)$, equation (139), with the Fourier volume V_k of the shell in k-space.

Step 4. As step 6 of Section 9.2: compute FKP-weighted pair densities $\hat{S}(r; \mu)$, equation (102).

Step 5. Compute the estimator $\hat{Z}(k)$:

$$\hat{Z}(k) = \int_0^\infty \frac{j_0(kr) \hat{S}[r; \mu(k)]}{[1 + \xi(r)]^{1/2}} 4\pi r^2\,\mathrm{d}r, \tag{145}$$

which may be compared to equation (104).

Step 6. As step 8 of Section 9.2: transform $\hat{Z}(k)$ to $\hat{X}(k)$ using equations (99), (74) and (70). The estimator $\hat{Y} = E^{-1} Z$ is

$$\hat{Y}(k) = \frac{2\,\mu(k)^2}{V(k)} \hat{Z}(k). \tag{146}$$

Step 7. Form pre-whitened band-powers $\hat{B}(k)$ by averaging $\hat{X}(k)$ over sufficiently broad shells in k-space, equation (143).

10 CONCLUSIONS

The main finding of this paper is that the pre-whitened non-linear power spectrum X_α defined by equation (69) has surprisingly sweet properties.

First, the covariance of the pre-whitened non-linear power is substantially narrower in Fourier space than the covariance of the non-linear power spectrum itself, Figs 3 and 4.

Secondly, in the FKP approximation, the four-point and three-point contributions M and L to the covariance of pre-whitened power are almost simultaneously diagonal (the two-point contribution is by construction the unit matrix, so is automatically diagonal), Figs 5 and 6. Thus the eigenmodes of the covariance of pre-whitened non-linear power form a set of almost uncorrelated modes somewhat analogous to the Fourier modes of power in the Gaussian case.

Thirdly, the eigenvalues μ_α and λ_α, as defined by equations (67) and (68), of the four-point and three-point pre-whitened matrices M and L are almost equal, $\mu_\alpha \approx \lambda_\alpha$ (Fig. 8), which is similar to the Gaussian case where $\mu(k) = \lambda(k) = \xi(k)$.

The second and third points above together make it possible to construct a near-minimum variance estimator, Section 7, and Fisher matrix, Section 6, of the pre-whitened non-linear power spectrum similar to the FKP estimator and Fisher matrix of the linear power spectrum in the Gaussian case.

Fourthly, all the above properties hold for all power spectra tested, including power-law non-linear power spectra $\xi(k) \propto k^n$ with indices $-2 < n < 0$ over the full range allowed by the hierarchical model, and including realistic power spectra, such as the observationally derived power spectrum of Peacock (1997), and an observationally concordant ΛCDM model of Eisenstein & Hu (1998), non-linearly evolved according to the Peacock & Dodds (1996) formula.

Fifthly, in the realistic cases of the Peacock (1997) and Eisenstein & Hu (1998) power spectra, the pre-whitened non-linear power spectrum $X(k)$ appears to be curiously close to the linear power spectrum $\xi_{\rm L}(k)$, Fig. 10.

This having been said, it should be emphasized that the above properties are all premised on the hierarchical model with constant hierarchical amplitudes, Section 3.2, which as discussed in Section 4.2 and by Scoccimarro et al. (1999, section 3.3) is certainly wrong at some level. Clearly it will be important to test how well these results stand up in N-body simulations.

In the meantime, the results of this paper raise questions. Is there some physical reason underlying the seemingly unreasonably pretty properties of the pre-whitened non-linear power spectrum? In general, modes may be statistically uncorrelated without being dynamically independent. However, the fact that the covariance of the pre-whitened power is narrow for all power spectra is suggestive: do the eigenmodes of the covariance of pre-whitened power somehow encode the information in the linear power spectrum that is ravelled by non-linear evolution in the power spectrum itself? And is there somehow a connection to the mapping between linear and non-linear power spectra found by Hamilton et al. (1991)?

I conclude with a repeat of the warning that this paper has ignored redshift distortions, light-to-mass bias and evolution, and it has assumed that the only sources of variance are cosmic variance and shot-noise variance arising from Poisson sampling of galaxies. In real galaxy surveys, all these problems must be grappled with.

ACKNOWLEDGMENTS

This work was supported by NASA Astrophysical Theory Grant NAG 5-7128. I thank Max Tegmark for much detailed comment on the original manuscript. Discussions with participants at the 1997 Aspen summer workshop on Precision Measurement of Large Scale Structure informed the content of this paper.

REFERENCES

Anderson W. L., 1982, ACM-Trans. Math. Software, 8, No. 4, 344 (Hankel transform code at http://www.acm.org/calgo/contents/588.gz)
Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 1999, MNRAS, submitted (astro-ph/9903343)
Bernardeau F., Schaeffer R., 1999, A&A, 349, 697 (astro-ph/9903387)
Binggeli B., Sandage A., Tammann G. A., 1988, ARA&A, 26, 509
Blackman R. B., Tukey J. W., 1959, The Measurement of Power Spectra from the Point of View of Communications Engineering, Dover, New York
Blanton M., Cen R., Ostriker J. P., Strauss M. A., Tegmark M., 1999, ApJ, submitted (astro-ph/9903165)
Bond J. R., Jaffe A. H., Knox L., 1998a, Phys. Rev., D57, 2117
Bond J. R., Jaffe A. H., Knox L. E., 1998b, astro-ph/9808264
Bouchet F. R., Strauss M. A., Davis M., Fisher K. B., Yahil A., Huchra J. P., 1993, ApJ, 417, 36
Candel S. M., 1981, IEEE Trans. ASSP, 29, 963
Carlberg R. G. et al., 1999, Phil. Trans. R. Soc. Lond. A, 357, 167 (astro-ph/9805131)
Cen R., Ostriker P. O., 1999, ApJ, submitted (astro-ph/9809370)
Chiu W. A., Ostriker J. P., Strauss M. A., 1998, ApJ, 494, 479
Coles P., 1993, MNRAS, 262, 1065
Coles P., Melott A., Munshi D., 1999, ApJ Lett., 521, L5 (astro-ph/9904253)
Colín P., Klypin A. A., Kravtsov A. V., Khokhlov A. M., 1999, ApJ, 523, 32 (astro-ph/9809202)
Colless M., 1999, Phil. Trans. R. Soc. Lond., 357, 105 (astro-ph/9804079)
Colley W. N., 1997, ApJ, 489, 471
Cowie L. L., Songaila A., Hu E. M., Cohen J. G., 1996, AJ, 112, 839
da Costa L. N. et al., 1998, AJ, 116, 1
Dekel A., Lahav O., 1999, ApJ, 520, 24 (astro-ph/9806193)
Dodelson S., Hui L., Jaffe A. H., 1999, preprint (astro-ph/9712074)
Eisenstein D. J., Hu W., 1998, ApJ, 496, 605
Eisenstein D. J., Hu W., Tegmark M., 1998, ApJ, 504, L57
Eisenstein D. J., Hu W., Tegmark M., 1999, ApJ, 518, 2 (astro-ph/9807130)
Falco E. E. et al., 1999, PASP, 111, 438
Fanning G., 1996, Hankel transform code at http://www.stanford.edu/~fanning/hankel/hankel.html
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23 (FKP)
Folkes S. R. et al., 1999, MNRAS, in press, astro-ph/9903456
Frieman J. A., Gaztañaga E., 1999, ApJL, 521, L83 (astro-ph/9903423)
Fry J. N., 1984, ApJ, 279, 499
Fry J. N., Gaztañaga E., 1993, ApJ, 413, 447
Gaztañaga E., 1994, MNRAS, 268, 913
Gaztañaga E., Frieman J. A., 1994, ApJ, 437, L13
Geller M. J. et al., 1997, AJ, 114, 2205
Goldberg D. M., Strauss M. A., 1998, ApJ, 495, 29
Grogin N. A., Geller M. J., Huchra J. P., 1998, ApJS, 119, 277
Gunn J. E., 1995, in Weinberg D. H., Maddox S. J., Aragón-Salamanca A., eds, Wide-Field Spectroscopy and the Distant Universe. World Scientific, Singapore, p. 3
Hamilton A. J. S., 1993, ApJ, 417, 19
Hamilton A. J. S., 1997a, MNRAS, 289, 285 (Paper I)
Hamilton A. J. S., 1997b, MNRAS, 289, 295 (Paper II)
Hamilton A. J. S., 1998, in Hamilton D., ed., The Evolving Universe. Kluwer, Dordrecht, p. 185

Hamilton A. J. S., Tegmark M., 2000, MNRAS, 312, 285 (Paper 4, this issue)
Hamilton A. J. S., Kumar P., Lu E., Matthews A., 1991, ApJ, 374, L1
Hansen E. W., 1985, IEEE Trans. ASSP, 33, 666
Hu W., Eisenstein D. J., Tegmark M., 1998, Phys. Rev. Lett., 80, 5255
Hui L., Gaztañaga E., 1999, ApJ, 519, 622 (astro-ph/9810194)
Jaynes E. T., 1996, Probability Theory: The Logic of Science, unpublished book, http://omega.math.albany.edu:8008/JaynesBook.html
Juszkiewicz R., Bouchet F. R., Colombi S., 1993, ApJ, 412, L9
Lilly S. J., Le Fèvre O., Crampton D., Hammer F., Tresse L., 1995, ApJ, 455, 50
Loveday J., Peterson B. A., Maddox S. J., Efstathiou G., 1996, ApJS, 107, 201
Mann R. G., Peacock J. A., Heavens A. F., 1998, MNRAS, 293, 209
Margon B., 1999, Phil. Trans. R. Soc. Lond. A, 357, 93 (astro-ph/9805314)
Meiksin A., White M., 1999, MNRAS, 308, 1179 (astro-ph/9812129)
Meiksin A., White M., Peacock J. A., 1999, MNRAS, 304, 851
Mo H. J., Jing Y. P., White S. D. M., 1997, MNRAS, 284, 189
Monaco P., Efstathiou G., 1999, MNRAS, 308, 763
Moscardini L., Coles P., Lucchin F., Matarrese S., 1998, MNRAS, 299, 95
Narayanan V. K., Weinberg D. H., 1999, ApJ, submitted (astro-ph/9806238)
Narayanan V. K., Berlind A., Weinberg D. H., 1999, ApJ, submitted (astro-ph/9812002)
Nusser A., Dekel A., Yahil A., 1995, ApJ, 449, 439
Ooura T., 1996, http://momonga.t.u-tokyo.ac.jp/~ooura/gamerf.html
Peacock J. A., 1997, MNRAS, 284, 885
Peacock J. A., Dodds S. J., 1996, MNRAS, 280, L19
Peebles P. J. E., 1980, The Large Scale Structure of the Universe. Princeton University Press, Princeton, NJ
Press W. H., Flannery B. P., Teukolsky S. A., Vetterling W. T., 1986, Numerical Recipes: The Art of Scientific Computing. Cambridge Univ. Press, Cambridge
Ratcliffe A., Shanks T., Parker Q. A., Broadbent A., Watson F. G., Oates A. P., Collins C. A., Fong R., 1998, MNRAS, 300, 417
Schuecker P., Ott H.-A., Seitter W. C., Ungruhe R., Duerbeck H. W., Cunow B., Spiekermann G., Duemmler R., 1998, ApJ, 496, 635
Scoccimarro R., Colombi S., Fry J. N., Frieman J. A., Hivon E., Melott A., 1998, ApJ, 496, 586
Scoccimarro R., Frieman J. A., 1999, ApJ, 520, 35 (astro-ph/9811184)
Scoccimarro R., Zaldarriaga M., Hui L., 1999, astro-ph/9901099
Shectman S. A., Landy S. D., Oemler A., Tucker D. L., Lin H., Kirshner R. P., Schechter P. L., 1996, ApJ, 470, 172
Scherrer R. J., Weinberg D. H., 1998, ApJ, 504, 607
Siegman A. E., 1977, Opt. Lett., 1, 13
Small T. A., Sargent W. L. W., Hamilton D., 1997, ApJS, 111, 1
Stirling A. J., Peacock J. A., 1996, MNRAS, 283, L99
Strauss M. A., 1999, in Dekel A., Ostriker J. P., eds, Formation of Structure in the Universe. Cambridge Univ. Press, Cambridge, p. 172 (astro-ph/9610033)
Sutherland W. et al., 1999, MNRAS, 308, 289
Swarztrauber P. N., 1979, NCAR FFT routines at http://www.netlib.org/bihar/
Talman J. D., 1978, J. Comp. Phys., 29, 35
Tegmark M., 1997a, Phys. Rev., D55, 5895
Tegmark M., 1997b, Phys. Rev. Lett., 79, 3806
Tegmark M., Peebles P. J. E., 1998, ApJ, 500, 79
Tegmark M., Taylor A. N., Heavens A. F., 1997, ApJ, 480, 22
Tegmark M., Hamilton A. J. S., Strauss M., Vogeley M., Szalay A., 1998, ApJ, 499, 555
Tresse L., 1999, in Le Fèvre O., Charlot S., eds, Formation and Evolution of Galaxies, Les Houches School Series. Springer-Verlag, in press (astro-ph/9902209)
Turner M. S., 1997, in Schramm D. N., Galeotti P., eds, Generation of Cosmological Large-Scale Structure. Kluwer, Dordrecht, p. 153
Vettolani G. et al., 1998, A&AS, 130, 323
Vogeley M. S., Szalay A. S., 1996, ApJ, 465, 34
White M., 1999, MNRAS, submitted (astro-ph/9811227)
Willmer C. N. A., 1997, AJ, 114, 898

APPENDIX A: JUSTIFICATION OF EQUATION (41)

Equation (41) is the FKP approximation, expressed in concise mathematical form. This appendix offers further details justifying this equation.

The pair–pair covariance matrix $\mathfrak{C}_{ijkl}$ can be regarded as an operator that acts on pair-functions Ψ_{kl}. It is helpful to think of Ψ_{kl} as a two-particle wavefunction (symmetric under pair exchange $k \leftrightarrow l$), and $\mathfrak{C}_{ijkl}$ as a Hermitian operator that acts on the space of such wavefunctions. The pair-wavefunctions Ψ_{kl} of interest in the present case have translation and rotation symmetry, which means that they have zero total momentum and zero total angular momentum. In the Fourier representation such wavefunctions Ψ_{kl} can be expressed in the form

$$\Psi(\boldsymbol{k}_k, \boldsymbol{k}_l) = (2\pi)^3 \delta_{3\mathrm{D}}(\boldsymbol{k}_k + \boldsymbol{k}_l)\psi(k_k), \tag{A1}$$

where $\psi(k_k)$ is a function of the scalar $k_k \equiv |\boldsymbol{k}_k|$. In a general representation, equation (A1) is

$$\Psi_{kl} = D^{\alpha}_{kl}\psi_{\alpha}, \tag{A2}$$

where D^{α}_{kl} is the operator introduced in Section 2.4, equations (28) and (29).

In the FKP approximation, the selection functions $\bar{n}_i$ and $\bar{n}_j$ upon which the pair-covariance $\mathfrak{C}_{ijkl}$ depends, equations (38) and (25), are taken to be locally constant, so that $\mathfrak{C}_{ijkl}$ also has translation and rotation symmetry, i.e. it commutes with the operators of total momentum and total angular momentum. Thus, in the FKP approximation, $\mathfrak{C}_{ijkl}$ acting on a wavefunction $D^{\alpha}_{kl}\psi_{\alpha}$ with zero momentum and angular momentum yields another wavefunction $D^{\beta}_{ij}\chi_{\beta}$ with zero momentum and angular momentum,

$$\mathfrak{C}^{kl}_{ij} D^{\alpha}_{kl}\psi_{\alpha} = D^{\beta}_{ij}\chi_{\beta}. \tag{A3}$$

Take ψ_{α} in equation (A3) to be the elements of a complete orthonormal basis of functions. Then equation (A3) implies that

$$\mathfrak{C}_{ijkl} D^{kl}_{\alpha} = D^{\beta}_{ij}\mathfrak{C}_{\beta\alpha} \tag{A4}$$

for some matrix $\mathfrak{C}_{\beta\alpha}$. Equation (A4) is the desired equation (41) that was to be justified [at least if the indices on $\mathfrak{C}_{\alpha\beta}$ are swapped in equation (A4), which is fine because $\mathfrak{C}_{\alpha\beta}$ is symmetric, as proven below]. The wavefunctions ψ_{α} and χ_{β} in equation (A3) are related by

$$\mathfrak{C}^{\alpha}_{\beta}\psi_{\alpha} = \chi_{\beta}. \tag{A5}$$

A wavefunction of the form $D^{\alpha}_{ij}\psi_{\alpha}$ is unnormalized – that is, $\psi^{\alpha} D^{ij}_{\alpha} D^{\beta}_{ij}\psi_{\beta}$ diverges – as is usual in quantum mechanics for a wavefunction that has definite momentum, and must therefore be defined over infinite space. The divergence can be tamed by regarding the wavefunction as being defined instead over an extremely large but finite volume V. Then

$$D^{ij}_{\alpha} D^{\beta}_{ij} = \mathbf{1}_{\alpha\beta} V, \tag{A6}$$

which is most easily proven from the real-space representation of D^{α}_{ij}, equation (28). Equation (A6) should be interpreted with due care. For example, equation (A6) should not be substituted into equation (45) for the Fisher matrix in the FKP approximation,

because the matrix $\mathfrak{C}^{-1\alpha\gamma}(\bar{n}_i, \bar{n}_j)$ on the right-hand side of equation (45) varies with positions i and j.

Operating on equation (A4) with D^{ij}_β implies, from equation (A6), that ($\bar{n}_i$ and $\bar{n}_j$ here are being regarded formally as fixed constants in the huge volume V)

$$D^{ij}_\beta \mathfrak{C}_{ijkl} D^{kl}_\alpha = V \mathfrak{C}_{\beta\alpha}. \tag{A7}$$

It is evident from this equation that the pair-exchange symmetry $ij \leftrightarrow kl$ of $\mathfrak{C}_{ijkl}$ implies that $\mathfrak{C}_{\alpha\beta}$ is similarly symmetric:

$$\mathfrak{C}_{\alpha\beta} = \mathfrak{C}_{\beta\alpha}. \tag{A8}$$

Equation (A7) shows that, modulo the normalization factor V, the reduced matrix $\mathfrak{C}_{\alpha\beta}$ can be regarded as the matrix elements of the operator $\mathfrak{C}_{ijkl}$ restricted to the class of wavefunctions that have zero total momentum and zero total angular momentum.

APPENDIX B: FFTLOG

B1 Introduction

FFTLog computes the fast Fourier or Hankel (=Fourier–Bessel) transform of a periodic sequence of logarithmically spaced points. FFTLog can be regarded as a natural analogue to the standard FFT, in the sense that, just as the normal FFT gives the exact (to machine precision) Fourier transform of a linearly spaced periodic sequence, so also FFTLog gives the exact Fourier or Hankel transform, of arbitrary order μ, of a logarithmically spaced periodic sequence. FFTLog shares with the normal FFT the problems of ringing (response to sudden steps) and aliasing (periodic folding of frequencies), but under appropriate circumstances FFTLog may approximate the results of a continuous Fourier or Hankel transform.

The FFTLog algorithm was originally proposed by Talman (1978). However, it seems worthwhile here to present the algorithm in some detail.

The FFTLog code may be downloaded from http://casa.colorado.edu/~ajsh/FFTLog/.

Consider the continuous Hankel (=Fourier–Bessel) transform pair

$$\tilde{a}(k) = \int_0^\infty a(r)(kr)^q J_\mu(kr) k \, \mathrm{d}r, \tag{B1}$$

$$a(r) = \int_0^\infty \tilde{a}(k)(kr)^{-q} J_\mu(kr) r \, \mathrm{d}k. \tag{B2}$$

If the substitution

$$a(r) = A(r) r^{-q} \quad \text{and} \quad \tilde{a}(k) = \tilde{A}(k) k^q \tag{B3}$$

is made, then the Hankel transform pair (B1), (B2) becomes equivalent to the transform pair

$$\tilde{A}(k) = \int_0^\infty A(r) J_\mu(kr) k \, \mathrm{d}r, \tag{B4}$$

$$A(r) = \int_0^\infty \tilde{A}(k) J_\mu(kr) r \, \mathrm{d}k. \tag{B5}$$

Although the Hankel transform (B1) with a power-law bias $(kr)^{\pm q}$ is thus equivalent in the continuous case to the unbiased Hankel transform (B4), the transforms are different when they are discretized and made periodic; for if $a(r)$ is periodic, then $A(r) = a(r) r^q$ is not periodic. FFTLog evaluates discrete Hankel transforms (B1) and (B2) with arbitrary power-law bias.

Fourier sine and cosine transforms can be regarded as special cases of Hankel transforms with $\mu = \pm 1/2$, since

$$J_{1/2}(x) = (2/\pi x)^{1/2} \sin(x), \tag{B6}$$

$$J_{-1/2}(x) = (2/\pi x)^{1/2} \cos(x). \tag{B7}$$

As first noted by Siegman (1977), if the product kr in the Hankel transform is written as $\mathrm{e}^{\ln k + \ln r}$, then the transform becomes a convolution integral in the integration variable $\ln r$ or $\ln k$. Convolution is equivalent to multiplication in the corresponding Fourier transform space. Thus the Hankel transform can be computed numerically by the algorithm: FFT → multiply by a function → FFT back. This is the idea behind a number of fast Hankel transform algorithms (Candel 1981; Anderson 1982; Hansen 1985; Fanning 1996), including FFTLog (Talman 1978).

An advantage of FFTLog, emphasized by Talman (1978), is that the order μ of the Bessel function may be any arbitrary real number. In particular, FFTLog works for 1/2-integral μ, so includes the cases of Fourier sine and cosine transforms, and spherical Hankel transforms involving the spherical Bessel functions $j_\lambda(x) \equiv (\pi/2x)^{1/2} J_{\lambda+1/2}(x)$.

B2 Normal discrete Fourier transform

First, recall the essential properties of the standard discrete Fourier transform of a periodic sequence of linearly spaced points. Suppose that $a(r)$ is a continuous, in general complex-valued, function that is periodic with period R,

$$a(r + R) = a(r). \tag{B8}$$

Without loss of generality, take the fundamental interval to be $[-R/2, R/2]$, centred at zero. Since $a(r)$ is periodic, its continuous Fourier transform contains only discrete Fourier modes $\mathrm{e}^{2\pi \mathrm{i} m r/R}$ with integral wavenumbers m. Suppose further that the function $a(r)$ is 'smooth' in the specific sense that it is some linear combination only of the N lowest frequency Fourier modes, $m = 0, \pm 1, \ldots, \pm[N/2]$, where $[N/2]$ denotes the largest integer greater than or equal to $N/2$,

$$a(r) = {\sum_m}' c_m \, \mathrm{e}^{2\pi \mathrm{i} m r/R}, \tag{B9}$$

the outermost Fourier coefficients being equal, $c_{-N/2} = c_{N/2}$, in the case of even N. The primed sum in equation (B9) signifies a sum over integral m from $-[N/2]$ to $[N/2]$, with the proviso that for even N the outermost elements of the sum receive only half-weight:

$${\sum_n}' x_n \equiv \sum_{n=-[N/2]}^{[N/2]} w_n x_n, \tag{B10}$$

with $w_n = 1$ except that $w_{-N/2} = w_{N/2} = 1/2$ if N is even.

The sampling theorem (e.g. Press et al. 1986, section 12.1) asserts that, given a function $a(r)$ satisfying equation (B9), the Fourier coefficients c_m can be expressed in terms of the values $a_n \equiv a(r_n)$ of the function $a(r)$ at the N discrete points $r_n = nR/N$ for $n = 0, \pm 1, \ldots, \pm[N/2]$. For even N, the periodicity of $a(r)$ ensures that $a_{-N/2} = a_{N/2}$. Specifically, the sampling theorem asserts that the Fourier coefficients in the expansion (B9) satisfy

$$c_m = \frac{1}{N} {\sum_n}' a_n \, \mathrm{e}^{-2\pi \mathrm{i} m n/N}, \tag{B11}$$

the discrete points a_n themselves satisfying

$$a_n = {\sum_m}' c_m \, \mathrm{e}^{2\pi \mathrm{i} mn/N} \tag{B12}$$

in accordance with equation (B9).

Equations (B11) and (B12) constitute a discrete Fourier transform pair relating two periodic, linearly spaced sequences a_n and c_m of length N. The standard FFT evaluates the discrete Fourier transform exactly (that is, to machine precision).

B3 Discrete Hankel transform

Now suppose that the function $a(r)$, instead of being periodic in ordinary space r, is periodic in logarithmic space $\ln r$, with logarithmic period L,

$$a(r\,\mathrm{e}^L) = a(r). \tag{B13}$$

Take the fundamental interval to be $[\ln r_0 - L/2, \ln r_0 + L/2]$, centred at $\ln r_0$. As in Appendix B2, the periodicity of $a(r)$ implies that its Fourier transform with respect to $\ln r$ contains only discrete Fourier modes $\mathrm{e}^{2\pi \mathrm{i} m \ln(r/r_0)/L}$ with integral wavenumbers m. Suppose further, as in Appendix B2, equation (B9), that $a(r)$ contains only the N lowest frequency Fourier modes

$$a(r) = {\sum_m}' c_m \, \mathrm{e}^{2\pi \mathrm{i} m \ln(r/r_0)/L}, \tag{B14}$$

with $c_{-N/2} = c_{N/2}$ for even N. The sampling theorem asserts that the Fourier coefficients c_m are given by

$$c_m = \frac{1}{N} {\sum_n}' a_n \, \mathrm{e}^{-2\pi \mathrm{i} mn/N}, \tag{B15}$$

where $a_n \equiv a(r_n)$ are the values of the function $a(r)$ at the N discrete points $r_n = r_0 \, \mathrm{e}^{nL/N}$ for $n = 0, \pm 1, \ldots, \pm[N/2]$,

$$a_n = {\sum_m}' c_m \, \mathrm{e}^{2\pi \mathrm{i} mn/N}. \tag{B16}$$

The continuous Hankel transform $\tilde{a}(k)$, equation (B1), of a function $a(r)$ of the form (B14) is

$$\tilde{a}(k) = {\sum_m}' c_m \int_0^\infty \mathrm{e}^{2\pi \mathrm{i} m \ln(r/r_0)/L} (kr)^q J_\mu(kr) k \, \mathrm{d}r. \tag{B17}$$

The integrals on the right-hand side of equation (B17) can be done analytically, in terms of

$$U_\mu(x) \equiv \int_0^\infty t^x J_\mu(t) \, \mathrm{d}t = 2^x \frac{\Gamma[(\mu + 1 + x)/2]}{\Gamma[(\mu + 1 - x)/2]}, \tag{B18}$$

where $\Gamma(z)$ is the usual Gamma-function. Thus equation (B17) reduces to

$$\tilde{a}(k) = {\sum_m}' c_m u_m \, \mathrm{e}^{-2\pi \mathrm{i} m \ln(k/k_0)/L}, \tag{B19}$$

where u_m is

$$u_m(\mu, q) \equiv (k_0 r_0)^{-2\pi \mathrm{i} m/L} U_\mu\left(q + \frac{2\pi \mathrm{i} m}{L}\right). \tag{B20}$$

Notice that $u_m^* = u_{-m}$, which ensures that $\tilde{a}(k)$ is real if $a(r)$ is real. Equation (B19) gives the (exact) continuous Hankel transform $\tilde{a}(k)$ of a function $a(r)$ of the form (B9). Like $a(r)$, the Hankel transform $\tilde{a}(k)$ is periodic in logarithmic space $\ln k$, with period L. The fundamental interval is $[\ln k_0 - L/2, \ln k_0 + L/2]$, centred at $\ln k_0$, which may be chosen arbitrarily (but see Appendix B5 below).

The sampling theorem requires that $u_{-N/2} = u_{N/2}$ for even N, which is not necessarily satisfied by equation (B20). However, at the discrete points $k_n = k_0 \, \mathrm{e}^{nL/N}$ considered by the sampling theorem, the contributions at $m = \pm N/2$ to the sum on the right-hand side of equation (B19) are $(-)^n c_{N/2}(u_{N/2} + u_{N/2}^*)/2$, the imaginary part of which cancels out. Thus the equality (B19) remains true at the discrete points k_n if $u_{\pm N/2}$ are replaced by their real parts,

$$u_{\pm N/2} \to \mathrm{Re}\, u_{N/2}. \tag{B21}$$

With the replacement (B21), the sampling theorem asserts that the coefficients $c_m u_m$ in the sum (B19) are determined by the values $\tilde{a}_n \equiv \tilde{a}(k_n)$ of the Hankel transform at the N discrete points $k_n = k_0 \, \mathrm{e}^{nL/N}$ for $n = 0, \pm 1, \ldots, \pm[N/2]$,

$$c_m u_m = \frac{1}{N} {\sum_n}' \tilde{a}_n \, \mathrm{e}^{2\pi \mathrm{i} mn/N}, \tag{B22}$$

$$\tilde{a}_n = {\sum_m}' c_m u_m \, \mathrm{e}^{-2\pi \mathrm{i} mn/N}. \tag{B23}$$

Putting together equations (B15), (B16), (B22) and (B23) yields the discrete Hankel transform pair

$$\tilde{a}_n = {\sum_m}' a_m \, v^+_{m+n}(\mu, q), \tag{B24}$$

$$a_m = {\sum_n}' \tilde{a}_n \, v^-_{m+n}(\mu, q), \tag{B25}$$

in which the forward discrete Hankel mode $v_n^+(\mu, q)$ is the discrete Fourier transform of $u_m(\mu, q)$ given by equations (B20) and (B21),

$$v_n^+(\mu, q) = \frac{1}{N} {\sum_m}' u_m(\mu, q) \, \mathrm{e}^{-2\pi \mathrm{i} mn/N} \tag{B26}$$

while the inverse discrete Hankel mode $v_n^-(\mu, q)$ is the discrete Fourier transform of the reciprocal $1/u_{-m}(\mu, q)$,

$$v_n^-(\mu, q) = \frac{1}{N} {\sum_m}' \frac{1}{u_{-m}(\mu, q)} \, \mathrm{e}^{-2\pi \mathrm{i} mn/N}. \tag{B27}$$

The Hankel transform matrices $v^+_{m+n}(\mu, q)$ and $v^-_{m+n}(\mu, q)$ are mutually inverse

$${\sum_l}' v^+_{m+l}(\mu, q) v^-_{l+n}(\mu, q) = \delta_{mn}, \tag{B28}$$

where δ_{mn} denotes the Kronecker delta. The forward and inverse Hankel modes have the interesting property of being self-similar: that is, Hankel modes $v^+_{m+n}(\mu, q)$ [or $v^-_{m+n}(\mu, q)$] with different indices m consist of the same periodic sequence $v_n^+(\mu, q)$ [or $v_n^-(\mu, q)$] cyclically shifted by m notches.

FFTLog evaluates the forward and inverse discrete Hankel transforms given by equations (B24) and (B25) exactly (to machine precision).

The reciprocal $1/u_{-m}(\mu, q)$ in equation (B27) is equal to $u_m(\mu, -q)$, according to equations (B18) and (B20),

$$\frac{1}{u_{-m}(\mu, q)} = u_m(\mu, -q) \qquad (m \neq N/2) \tag{B29}$$

except in the case $m = \pm N/2$ for even N, when the replacement (B21) generally invalidates equation (B29). However, in the special case where $u_{\pm N/2}$ are already real, then equation (B21) leaves $u_{\pm N/2}$ unchanged, and equation (B29) remains valid also at

$m = \pm N/2$. This special case is of particular interest, and is discussed further in Section B5 below.

In the continuous case, the inverse Hankel transform is equal to the forward transform with $q \to -q$, equations (B1) and (B2). In the discrete case this remains true for odd N, but it is not generally true for even N (the usual choice) except in the important special case discussed in Section B5.

In the general discrete case (i.e. if the condition (B30) in Section B5 is not satisfied), the inverse discrete Hankel mode $v_n^-(\mu, q)$, equation (B27), differs from the forward Hankel mode $v_n^+(\mu, -q)$, equation (B26), only for even N and only in the coefficient of the highest frequency Fourier component, $1/u_{-m}(\mu, q)$ versus $u_m(\mu, -q)$ for $m = \pm N/2$. To the extent that the highest frequency Fourier coefficient $c_{\pm N/2}$ of a sequence a_n is small, the difference between its inverse discrete Hankel transform and its forward transform with $q \to -q$ should be small.

It is possible for the inverse discrete Hankel transform to be singular, if $u_{\pm N/2}$ is purely imaginary, so that its real part vanishes, making $v_n^-(\mu, q)$ singular. As discussed in Section B5, this singularity can be avoided by choosing a low-ringing value of $k_0 r_0$, equation (B32).

The forward (inverse) discrete Hankel transforms are also singular at special values of μ and q, namely where $\mu + 1 + q$ (or $\mu + 1 - q$ in the inverse case) vanishes, because $u_0(\mu, q) = U_\mu(q)$ is singular at these points. This singularity reflects a real singularity in the corresponding continuous Hankel transform (unlike the singularity of the previous paragraph, which is an avoidable artefact of discreteness). The singularity in u_0 leads to an additive infinite constant in the discrete Hankel transform. In physical problems this additive infinite constant may somehow cancel out (for example, in the difference between two Hankel transforms). FFTLog's strategy in these singular cases is to evaluate the discrete Hankel transform with the infinite constant set to zero, and to issue a warning.

B4 FFTLog algorithm

The FFTLog algorithm for taking the discrete Hankel transform, equation (B24), of a sequence a_n of N logarithmically spaced points is:

(i) FFT a_n to obtain the Fourier coefficients c_m, equation (B15);
(ii) multiply by u_m given by equations (B20) and (B21) to obtain $c_m u_m$;
(iii) FFT $c_m u_m$ back to obtain the discrete Hankel transform $\tilde{a}_n$, equation (B23).

A variant of the algorithm is to sandwich the above operations with power-law biasing and unbiasing operations. For example, one way to take the unbiased continuous Hankel transform $\tilde{A}(k)$ of a function $A(r)$, equation (B4), is to bias $A(r)$ and $\tilde{A}(k)$ with power laws, equation (B3), and take a biased Hankel transform, equation (B1). The discrete equivalent of this is:

(i) bias A_n with a power law to obtain $a_n = A_n r_n^{-q}$, equation (B3);
(ii) FFT a_n to obtain the Fourier coefficients c_m, equation (B15);
(iii) multiply by u_m given by equations (B20) and (B21) to obtain $c_m u_m$;
(iv) FFT $c_m u_m$ back to obtain the discrete Hankel transform $\tilde{a}_n$, equation (B23);
(v) unbias $\tilde{a}_n$ with a power law to obtain $\tilde{A}_n = \tilde{a}_n k_n^{-q}$, equation (B3).

Although in the continuous limit the result would be identical to an unbiased Hankel transform, in the discrete case the result differs. With a simple unbiased discrete Hankel transform, it is the sequence A_n that is taken to be periodic, whereas in the algorithm above it is not A_n but rather a_n that is periodic.

The inverse discrete Hankel transform is accomplished by the same series of steps, except that c_m is divided instead of multiplied by u_m.

The FFTLog code is built on top of the National Center for Atmospheric Research (NCAR) suite of FFT routines (Swarztrauber 1979), and a modified version of an implementation of the complex Gamma-function from the GAMERF package by Ooura (1996).

FFTLog includes driver routines for the specific cases of the Fourier sine and cosine transforms.

B5 Low-ringing condition on $k_0 r_0$

The central values $\ln r_0$ and $\ln k_0$ of the periodic intervals in $\ln r$ and $\ln k$ may be chosen arbitrarily. However, ringing of the discrete Hankel transform may be reduced, for either even or odd N, if the product $k_0 r_0$ is chosen in such a way that the boundary points of the sequence u_m, equation (B20), are equal:

$$u_{-N/2} = u_{N/2}. \tag{B30}$$

Recall that the general procedure, for even N, was to replace $u_{\pm N/2}$ by their real part, equation (B21). The condition (B30) requires that $u_{\pm N/2}$ are already real. The condition (B30) reduces ringing because it makes the periodic sequence u_m fold smoothly across the period boundary at $m = \pm N/2$.

In addition to reducing ringing, the condition (B30) means that equation (B29) remains true also at $m = \pm N/2$, so is true for all m. In this case the inverse Hankel mode $v_n^-(\mu, q)$, equation (B27), is equal to the forward Hankel mode $v_n^+(\mu, -q)$ with q of the opposite sign

$$v_n^-(\mu, q) = v_n^+(\mu, -q) = \frac{1}{N}{\sum_m}' u_m(\mu, -q)\, \mathrm{e}^{-2\pi \mathrm{i} m n/N}. \tag{B31}$$

In other words, if condition (B30) is satisfied, then the inverse discrete Hankel transform equals the forward discrete Hankel transform with $q \to -q$. This is like the continuous Hankel transform, equations (B1), (B2), where the inverse transform equals the forward transform with $q \to -q$.

The periodicity condition (B30) on $u_{\pm N/2}$ translates, for real μ and q, into a condition on $k_0 r_0$:

$$\ln(k_0 r_0) = \frac{L}{N}\left\{\frac{1}{\pi}\mathrm{Arg}\left[U_\mu\left(q + \pi\frac{\mathrm{i}N}{L}\right)\right] + \text{integer}\right\} \tag{B32}$$

where $\mathrm{Arg} z \equiv \mathrm{Im} \ln z$ denotes the argument of a complex number, and 'integer' is any integer. In other words, to reduce ringing, it may help to choose $k_0 r_0$ so as to satisfy the condition (B32). This is not too much of a restriction, since L/N is the logarithmic spacing between points (= one notch), so the low-ringing condition (B32) allows $k_0 r_0$ to be chosen to lie within half a notch [$= L/(2N)$] of whatever number one chooses, for example within half a notch of $k_0 r_0 = 1$.

FFTLog can be set to use automatically the low-ringing value of $k_0 r_0$ nearest to any input value of $k_0 r_0$.

How else does the choice of $k_0 r_0$ affect the Hankel transform? Increasing the value of $\ln(k_0 r_0)$ by one notch L/N cyclically shifts the discrete Hankel transform $\tilde{a}_n$, equation (B23), by one notch to the left, $\tilde{a}_n \to \tilde{a}_{n-1}$. In other words, changing $\ln(k_0 r_0)$ by an

integral number of notches shifts the origin of the transform, but leaves the transform otherwise unchanged, as might have been expected.

In practice, since in most cases one is probably using the discrete Hankel transform as an approximation to the continuous transform, one would probably want to use $k_0 r_0 \approx 1$ (or 2, or π, according to taste).

B6 Unitary Hankel transform

The discrete Hankel transform with both low-ringing $k_0 r_0$ and no power-law bias, $q = 0$, is of particular interest because it is unitary, like the Fourier transform. Indeed, being also real, the low-ringing unbiased Hankel transform is orthogonal, i.e. self-inverse, like the Fourier sine and cosine transforms. This is like the continuous unbiased ($q = 0$) Hankel transform, equations (B1), (B2), which is self-inverse.

The discrete Hankel modes $v_{\rm m}(\mu, 0) = v_m^+(\mu, 0) = v_m^-(\mu, 0)$ in the low-ringing unbiased ($q = 0$) case are periodic, orthonormal, and self-similar, equation (B27),

$$\sum_l{}' v_{{\rm m}+l}(\mu, 0) v_{l+n}(\mu, 0) = \delta_{mn}. \qquad \text{(B33)}$$

Like any orthogonal transformation, the low-ringing unbiased ($q = 0$) Hankel transform commutes with the operations of matrix multiplication, inversion and diagonalization (for non-low-ringing or biased Hankel transforms, $q \neq 0$, the operations do not commute). That is, the Hankel transform of the product of two matrices is equal to the product of their Hankel transforms, and so on.

All else being equal (which it may not be), given a choice between applying an unbiased ($q = 0$) or biased ($q \neq 0$) Hankel transform, and between a low-ringing $k_0 r_0$, equation (B32), or otherwise, one would be inclined to choose the low-ringing unbiased transform, because of its orthogonality property.

B7 Example

Fig. B1 shows the correlation function $\xi(r)$ computed by FFTLog for the non-linear ΛCDM power spectrum of Eisenstein & Hu (1998) shown in Fig. 10. Two different resolutions are plotted on top of each other, a low-resolution case with 96 points over the range $r = 10^{-3}$ to $10^3\,h^{-1}$ Mpc, and a high-resolution case with 768 points over the range $r = 10^{-6}$ to $10^6\,h^{-1}$ Mpc. Both cases used an unbiased ($q = 0$) transform and a low-ringing value of $k_0 r_0$ (actually the choice of $k_0 r_0$ made little difference here).

The low- and high-resolution correlation functions shown in Fig. B1 agree well except near the edges $r \approx 10^{-3}$ and $10^3\,h^{-1}$ Mpc; in particular, the low-resolution correlation function tends to a positive constant $\approx 10^{-5}$ at $r \to 10^3\,h^{-1}$ Mpc, whereas the high-resolution correlation function is negative and declining as a power law $\propto r^{-4}$ at large r. The disagreement is caused by aliasing (see Appendix B8) of small and large separations in the low-resolution case. Aliasing is almost eliminated in the high-resolution case because the range $r = 10^{-6}$ to $10^6\,h^{-1}$ Mpc over which the transform was computed is much broader than the range plotted.

The bottom panel of Fig. B1 shows the ratio $\xi_{\rm FFT}/\xi_{\rm FFTLog}$ of the correlation function $\xi_{\rm FFT}$, computed with a normal FFT (sine transform) with 1023 points over the range $r = 0.125$ to $128\,h^{-1}$ Mpc, to the (high resolution) correlation function $\xi_{\rm FFTLog}$ computed with FFTLog. Even with 1023 points, the FFT'd correlation function rings noticeably, with an amplitude of about ±5 per cent.

In this particular instance, FFTLog outperforms the normal FFT on all counts: it is more accurate, with fewer points, over a larger range, and it shows no signs of ringing. This does not mean that FFTLog is always better than FFT. Rather, FFTLog is well matched to the problem at hand: the cosmological power spectrum extends over many orders of magnitude in wavenumber k, and varies smoothly in $\ln k$.

B8 Ringing and aliasing

FFTLog suffers from the same problems of ringing (response to sudden steps) and aliasing (periodic folding of frequencies) as the normal FFT.

Usually one is interested in the discrete Fourier or Hankel transform not for its own sake, but rather as an approximation to

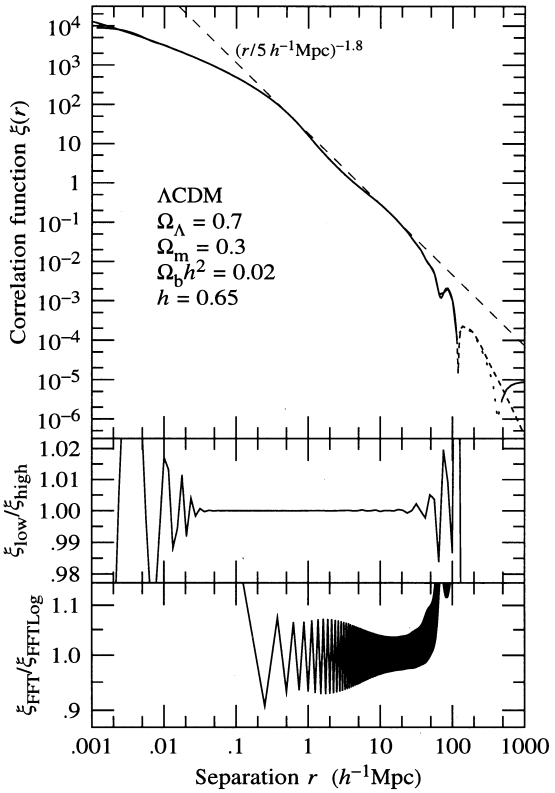

Figure B1. Correlation function $\xi(r)$ corresponding to the non-linear *COBE*-normalized ΛCDM power spectrum of Eisenstein & Hu (1998), Fig. 10. The top panel shows the correlation function computed with FFTLog at two different resolutions, plotted on top of each other: (i) with 96 points over the range $r = 10^{-3}$ to $10^3\,h^{-1}$ Mpc (low resolution), and (ii) with 768 points over the range $r = 10^{-6}$ to $10^6\,h^{-1}$ Mpc (high resolution). The lines are dashed where the correlation function is negative, at separations $r > 119\,h^{-1}$ Mpc. The low- and high-resolution curves are almost indistinguishable except at $r \gtrsim 200\,h^{-1}$ Mpc: the high-resolution curve is the one that declines as a power law $\sim r^{-4}$ at large r. The straight dashed line shows the canonical power law $(r/5\,h^{-1}\,{\rm Mpc})^{-1.8}$ for reference. The middle panel shows the ratio $\xi_{\rm low}/\xi_{\rm high}$ of the low- and high-resolution correlation functions. The bottom panel shows the ratio $\xi_{\rm FFT}/\xi_{\rm FFTLog}$ of the correlation function $\xi_{\rm FFT}$, computed with a normal FFT (sine transform) with 1023 points over the range $r = 0.125$ to $128\,h^{-1}$ Mpc, to the high-resolution correlation function $\xi_{\rm FFTLog}$ computed with FFTLog. The FFT'd correlation function $\xi_{\rm FFT}$ rings at the ±5 per cent level.

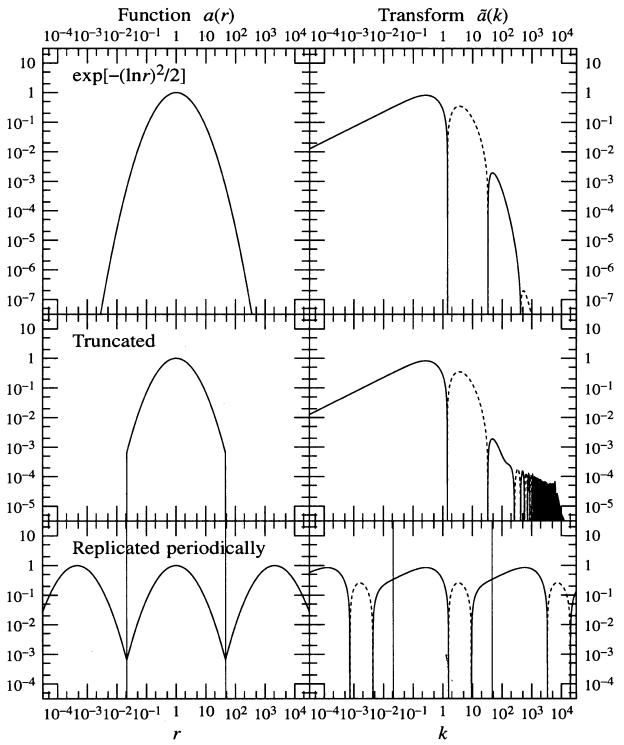

Figure B2. Illustrating the ringing and aliasing that occurs when the continuous Hankel transform of a function is approximated by the discrete Hankel transform of a finite segment of the function. Lines are dashed where values are negative. The function $a(r)$ is shown to the left, and its corresponding Hankel transform $\tilde{a}(k)$ to the right. The panels from top to bottom are: (top) the original function $a(r)$ and its Hankel transform $\tilde{a}(k)$; (middle) the truncated function $a(r)$ and its Hankel transform $\tilde{a}(k)$, which rings at high frequencies k; and (bottom) the truncated, periodically replicated function $a(r)$ and its corresponding periodically replicated Hankel transform $\tilde{a}(k)$, which is aliased. Vertical lines in the bottom panels demarcate periodic intervals.

the continuous transform. The usual procedure would be to apply the discrete transform to a finite segment of the function $a(r)$ to be transformed. For FFTLog, the procedure can be regarded as involving two steps: truncating the function to a finite logarithmic interval, which causes ringing of the transform; followed by periodic replication of the function in logarithmic space, which causes aliasing.

Fig. B2 illustrates these steps for the unbiased ($q = 0$) Hankel transform, equation (B1), of order $\mu = -1/2$ of a function that is Gaussian in the log:

$$a(r) = \exp[-(\ln r)^2/2]. \tag{B34}$$

Truncation of the function $a(r)$ leads to ringing of its transform $\tilde{a}(k)$ at high frequencies k, as seen in the middle right panel of Fig. B2. The oscillations at large k are actually uniformly spaced in k, but appear bunched up because of the logarithmic plotting.

Periodic replication means taking a sum of copies shifted by integral periods. From the definition Fig. (B1) of the continuous Hankel transform, it can be seen that periodically replicating a function $a(r)$ in logarithmic space $\ln r$ and then taking its continuous Hankel transform is equivalent to Hankel-transforming the function $a(r)$ and then periodically replicating the transform $\tilde{a}(k)$ in $\ln k$. However, truncating a function does not truncate its transform, so, whereas a truncated, periodically replicated function $a(r)$ contains contributions from only one period at each point r, the periodically replicated transform contains overlapping contributions from many periods at each point k. This is aliasing. In Fig. B2 aliasing is visible as an enhancement of the periodically replicated transform $\tilde{a}(k)$ on the high-k side of the periodic interval.

Ringing and aliasing can be reduced by taking suitable precautions.

The ringing that results from taking the discrete transform of a finite segment of a function can be reduced by arranging that the function folds smoothly from large to small scales. It may help to bias the function with a power law before transforming it, as in the second algorithm in Appendix B4. It may also help to use a low-ringing value of $k_0 r_0$, Appendix B5.

Aliasing can be reduced by enlarging the periodic interval. Aliasing can be eliminated (to machine precision) if the interval can be enlarged to the point where the transform $\tilde{a}(k)$ goes sensibly to zero at the boundaries of the period. Note that it is not sufficient to enlarge the interval to the point where $a(r)$ is sensibly zero at the period boundaries: what is important is that the transform $\tilde{a}(k)$ goes to zero at the boundaries.

This paper has been typeset from a TEX/LATEX file prepared by the author.