Mon. Not. R. Astron. Soc. **371,** 1188–1204 (2006) doi:10.1111/j.1365-2966.2006.10709.x

On measuring the covariance matrix of the non-linear power spectrum from simulations

Andrew J. S. Hamilton,[1,2]★ Christopher D. Rimes[1]★ and Román Scoccimarro[3]★

[1]*JILA, University of Colorado, 440 UCB, Boulder, CO 80309-0440, USA*
[2]*Department of Astrophysical & Planetary Sciences, University of Colorado, 391 UCB, Boulder, CO 80309-0391, USA*
[3]*Centre for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA*

Accepted 2006 June 14. Received 2006 May 22; in original form 2005 November 16

ABSTRACT

We show how to estimate the covariance of the power spectrum of a statistically homogeneous and isotropic density field from a single periodic simulation, by applying a set of weightings to the density field, and by measuring the scatter in power spectra between different weightings. We recommend a specific set of 52 weightings containing only combinations of fundamental modes, constructed to yield a minimum variance estimate of the covariance of power. Numerical tests reveal that at non-linear scales the variance of power estimated by the weightings method substantially exceeds that estimated from a simple ensemble method. We argue that the discrepancy is caused by beat-coupling, in which products of closely spaced Fourier modes couple by non-linear gravitational growth to the beat mode between them. Beat-coupling appears whenever non-linear power is measured from Fourier modes with a finite spread of wavevector, and is therefore present in the weightings method but not in the ensemble method. Beat-coupling inevitably affects real galaxy surveys, whose Fourier modes have finite width. Surprisingly, the beat-coupling contribution dominates the covariance of power at non-linear scales, so that, counter-intuitively, it is expected that the covariance of non-linear power in galaxy surveys is dominated not by small-scale structure, but rather by beat-coupling to the largest scales of the survey.

Key words: methods: data analysis – large-scale structure of Universe.

1 INTRODUCTION

The last few years have seen the emergence of a standard ΛCDM (cold dark matter) model of cosmology motivated by and consistent with a wide range of observations, including the cosmic microwave background, distant supernovae, big bang nucleosynthesis, large-scale structure, the abundance of rich galaxy clusters and local measurements of the Hubble constant (e.g. Tegmark et al. 2004b).

The power spectrum of fluctuations (of temperature, density, flux, shear etc.) is the primary statistic used to constrain cosmological parameters from observations of the cosmic microwave background (Spergel et al. 2003), of galaxies (Tegmark et al. 2004a; Cole et al. 2005; Eisenstein et al. 2005; Sanchez et al. 2006), of the Lyman α forest (Seljak et al. 2005; Viel & Haehnelt 2005; Lidz et al. 2006) and of weak gravitational lensing (Hoekstra, Yee & Gladders 2002; Pen et al. 2003; Sheldon et al. 2004; Takada & Jain 2004).

From a cosmological standpoint, the most precious data lie at large, linear scales, where fluctuations preserve the imprint of their primordial generation. A generic, albeit not universal, prediction of inflation is that primordial fluctuations should be Gaussian. At large, linear scales, observations are consistent with fluctuations being Gaussian (Komatsu et al. 2003).

However, much of the observational data, especially those involving galaxies, lies in the translinear or non-linear regime. It remains a matter of ongoing research to elucidate the extent to which non-linear data can be used to constrain cosmology (Cooray 2004).

We recently began (Rimes & Hamilton 2006) a program to measure quantitatively, from cosmological simulations, the Fisher information content of the non-linear matter power spectrum (specifically, in the first instance, the information about the initial amplitude of the linear power spectrum). For Gaussian fluctuations, the power spectrum contains all possible information about cosmological parameters. At non-linear scales, where fluctuations are non-Gaussian, it is natural to start by measuring information in the power spectrum, although it seems likely that additional information resides in the three-point and higher order correlation functions (Takada & Jain 2004; Sefusatti & Scoccimarro 2005).

Measuring the Fisher information in the power spectrum involves measuring the covariance matrix of power. For Gaussian fluctuations, the expected covariance of estimates of power is known

★E-mail: Andrew.Hamilton@colorado.edu (AJSH); rimes@colorado.edu (CDR); rs123@nyu.edu (RS)

analytically, but at non-linear scales the covariance of power must be estimated from simulations.

A common way to estimate the covariance matrix of a quantity is to measure its covariance over an ensemble of computer simulations (Meiksin & White 1999; Scoccimarro, Zaldarriaga & Hui 1999; Rimes & Hamilton 2005; Zhan & Eisenstein 2005; Zhan et al. 2005). However, a reliable estimate of covariance can be computationally expensive, requiring many, perhaps hundreds (Meiksin & White 1999; Rimes & Hamilton 2005) of realizations. On the other hand, it is physically obvious that the fluctuations in the values of quantities over the different parts of a single simulation must somehow encode the covariance of the quantities. If the covariance could be measured from single simulations, then it would be possible to measure covariance from fewer, and from higher quality, simulations. In any case, the ability to measure covariance from a single simulation can be useful in identifying simulations whose statistical properties are atypical.

A fundamental difficulty with estimating covariances from single simulations in cosmology is that the data are correlated over all scales, from small to large. As described by Künsch (1989), such correlations invalidate some of the 'jackknife' and 'bootstrap' schemes suggested in the literature. In jackknife, variance is inferred from how much a quantity varies when some segments of the data are kept, and some deleted. Bootstrap is like jackknife, except that deleted segments are replaced with other segments.

As part of the work leading to the present paper, we investigated a form of the bootstrap procedure, in which we filled each octant of a simulation cube with a block of data selected randomly from the cube. Unfortunately, the sharp edges of the blocks introduced undesirable small-scale power, which seemed to compromise the effort to measure covariance of power reliably. Such effects can be mitigated by tapering (Künsch 1989). However, it seemed to us that bootstrapping, like jackknifing, is a form of reweighting data, and that surely the best way to reweight data would be to apply the most slowly possible varying weightings. For a periodic box, such weightings would be composed of the largest scale modes, the fundamentals.

In the present paper, Section 2, we consider applying an arbitrary weighting to the density of a periodic cosmological simulation, and we show how the power spectrum (and its covariance, and the covariance of its covariance) of the weighted density are related to the true power spectrum (and its covariance, and the covariance of its covariance). We confirm mathematically the intuitive idea that weighting with fundamentals yields the most reliable estimate of covariance of power. Multiplying the density in real space by some weighting is equivalent to convolving the density in Fourier space with the Fourier transform of the weighting. This causes the power spectrum (and its covariance, and the covariance of its covariance) to be convolved with the Fourier transform of the square (and fourth, and eighth powers) of the weighting. The convolution does least damage when the weighting window is as narrow as possible in Fourier space, which means composed of fundamentals.

In Section 3, we show how to design a best set of weightings, by minimizing the expected variance of the resulting estimate of covariance of power. These considerations lead us to recommend a specific set of 52 weightings, each consisting of a combination of fundamental modes.

This paper should have stopped neatly at this point. Unfortunately, numerical simulations, described in a companion paper (Rimes & Hamilton 2006), revealed an unexpected (one might say insidious), substantial discrepancy at non-linear scales between the variance of power estimated by the weightings method and the variance of power estimated by the ensemble method. In Section 4, we argue that this discrepancy arises from beat-coupling, a non-linear gravitational coupling to the large-scale beat mode between closely spaced non-linear wavenumbers, when the power spectrum is measured from Fourier modes at anything other than infinitely sharp sets of wavenumbers. Surprisingly, in cosmologically realistic simulations, the covariance of power is dominated at non-linear scales by this beat-coupling to large scales.

We discuss the beat-coupling problem in Section 5. Beat-coupling is relevant to observations because real galaxy surveys yield Fourier modes in finite bands of wavenumber k, of width $\Delta k \sim 1/R$, where R is a characteristic linear size of the survey.

Section 6 summarizes the results.

2 ESTIMATING THE COVARIANCE OF POWER FROM AN ENSEMBLE OF WEIGHTED DENSITY FIELDS

The fundamental idea of this paper is to apply an ensemble of weightings to a (non-Gaussian, in general) density field, and to estimate the covariance of the power spectrum from the scatter in power between different weightings. This section derives the relation between the power spectrum of a weighted density field and the true power spectrum, along with its expected covariance and the covariance of its covariance.

It is shown, equations (34), (40) and (44), that the expected ((covariance of) covariance of) shell-averaged power of weighted density fields is simply proportional to the true ((covariance of) covariance of) shell-averaged power, provided that two approximations are made. The two approximations are, first, that the power spectrum and trispectrum are sufficiently slowly varying functions of their arguments, equations (17) and (37), and, secondly, that power is estimated in sufficiently broad shells in k space, equation (39). The required approximations are most accurate if the weightings contain only the largest scale Fourier modes, such as the weightings containing only fundamental modes proposed in Section 3.

As will be discussed in Section 4, the apparently innocent assumption, equation (37), that the trispectrum is a slowly varying function of its arguments is incorrect, because it sets to zero some important beat-coupling contributions. However, it is convenient to pretend in this section and the next, Sections 2 and 3, that the assumption (37) is true, and then to consider in Section 4 how the results are modified when the beat-coupling contributions to the trispectrum are included. Ultimately, we find, Section 4.5, that the weightings method remains valid when beat-couplings are included, and, Section 4.6, that the minimum variance weightings derived in Section 3, while no longer exactly minimum variance, should be close enough to remain good for practical application.

This section is necessarily rather technical, because it is necessary to distinguish carefully between various flavours of power spectrum: estimated versus expected; unweighted versus weighted; non-shell-averaged versus shell averaged. Subsections 2.1–2.7 present expressions for the various power spectra, their covariances and the covariances of their covariances. Subsections 2.8 and 2.9 show how the expressions are modified when, as is usually the case, deviations in power must be measured relative to an estimated rather than an expected value of power.

2.1 The power spectrum

Let $\rho(\boldsymbol{r})$ denote the density of a statistically homogeneous random field at position $\boldsymbol{r}$ in a periodic box. Choose the unit of length so that the box has unit side. The density $\rho(\boldsymbol{r})$ might represent, perhaps, a realization of the non-linearly evolved distribution of dark matter, or of galaxies. The density could be either continuous or discrete (particles). Expanded in Fourier modes $\rho(\boldsymbol{k})$, the density $\rho(\boldsymbol{r})$ is[1]

$$\rho(\boldsymbol{r}) = \sum_{\boldsymbol{k}} \rho(\boldsymbol{k}) \mathrm{e}^{-2\pi \mathrm{i} \boldsymbol{k}\cdot\boldsymbol{r}}. \tag{1}$$

Thanks to periodicity, the sum is over an integral lattice of wavenumbers, $\boldsymbol{k} = \{k_x, k_y, k_z\}$ with integer k_x, k_y, k_z.

The expectation value $\langle\rho(\boldsymbol{r})\rangle$ of the density defines the true mean density $\overline{\rho}$, which without loss of generality we take to equal unity

$$\overline{\rho} \equiv \langle\rho(\boldsymbol{r})\rangle = 1. \tag{2}$$

The deviation $\Delta\rho(\boldsymbol{r})$ of the density from the mean is

$$\Delta\rho(\boldsymbol{r}) \equiv \rho(\boldsymbol{r}) - \overline{\rho}. \tag{3}$$

The expectation values of the Fourier amplitudes vanish, $\langle\rho(\boldsymbol{k})\rangle = 0$, except for the zeroth mode, whose expectation value equals the mean density, $\langle\rho(\boldsymbol{0})\rangle = \overline{\rho}$. The Fourier amplitude $\rho(\boldsymbol{0})$ of the zeroth mode is the actual density of the realization, which could be equal to, or differ slightly from, the true mean density $\overline{\rho}$, depending on whether the mean density of the realization was constrained to equal the true density, or not.

Because the density field is by assumption statistically homogeneous, the expected covariance of Fourier amplitudes $\rho(\boldsymbol{k})$ is a diagonal matrix

$$\langle\Delta\rho(\boldsymbol{k}_1)\,\Delta\rho(\boldsymbol{k}_2)\rangle = 1_{\boldsymbol{k}_1+\boldsymbol{k}_2} P(\boldsymbol{k}_1). \tag{4}$$

Here $1_{\boldsymbol{k}}$ denotes the discrete delta function,

$$1_{\boldsymbol{k}} = \begin{cases} 1 & \text{if } \boldsymbol{k} = \boldsymbol{0} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

and $P(\boldsymbol{k})$ is the power spectrum. Note that there would normally be an extra factor of $\overline{\rho}^{-2}$ on the left-hand side of equation (4), but it is fine to omit the factor here because the mean density is normalized to unity, equation (2). The reason for dropping the factor of $\overline{\rho}^{-2}$ is to maintain notational consistency with equation (15) below for the power spectrum of weighted density (where the deviation in density is necessarily *not* divided by the mean).

The symmetry $P(-\boldsymbol{k}) = P(\boldsymbol{k})$ in equation (4) expresses pair exchange symmetry. Below, Section 2.5, we will assume that the density field is statistical isotropic, in which case the power is a function $P(k)$ only of the scalar wavenumber $k \equiv |\boldsymbol{k}|$, but for now we stick to the more general case where power is a function $P(\boldsymbol{k})$ of vector wavenumber $\boldsymbol{k}$.

2.2 The power spectrum of weighted density

Let $w_i(\boldsymbol{r})$ denote the ith member of a set of real-valued weighting functions, and let $\rho_i(\boldsymbol{r})$ denote the density weighted by the ith weighting,

$$\rho_i(\boldsymbol{r}) \equiv w_i(\boldsymbol{r})\rho(\boldsymbol{r}). \tag{6}$$

[1] The same symbol ρ is used in both real and Fourier space. The justification for this notation is that ρ is the same vector in Hilbert space irrespective of the basis with respect to which it is expanded. See for example Hamilton (2005) for a pedagogical exposition.

The Fourier amplitudes $\rho_i(\boldsymbol{k})$ of the weighted density are convolutions of the Fourier amplitudes of the weighting and the density:

$$\rho_i(\boldsymbol{k}) = \sum_{\boldsymbol{k}'} w_i(\boldsymbol{k}')\rho(\boldsymbol{k} - \boldsymbol{k}'). \tag{7}$$

Reality of the weighting functions implies

$$w_i(-\boldsymbol{k}) = w_i^*(\boldsymbol{k}). \tag{8}$$

The expected mean $\overline{\rho}_i(\boldsymbol{r})$ of the weighted density is proportional to the weighting,

$$\overline{\rho}_i(\boldsymbol{r}) \equiv \langle\rho_i(\boldsymbol{r})\rangle = w_i(\boldsymbol{r}), \tag{9}$$

in which a factor of $\overline{\rho}$ on the right-hand side has been omitted because the mean density has been normalized to unity, equation (2). The deviation $\Delta\rho_i(\boldsymbol{r})$ of the weighted density from the mean is

$$\Delta\rho_i(\boldsymbol{r}) \equiv \rho_i(\boldsymbol{r}) - \overline{\rho}_i(\boldsymbol{r}). \tag{10}$$

In Fourier space the expected mean $\overline{\rho}_i(\boldsymbol{k})$ of the weighted density is

$$\overline{\rho}_i(\boldsymbol{k}) \equiv \langle\rho_i(\boldsymbol{k})\rangle = w_i(\boldsymbol{k}), \tag{11}$$

and the deviation $\Delta\rho_i(\boldsymbol{k})$ of the weighted density from the mean is

$$\Delta\rho_i(\boldsymbol{k}) \equiv \rho_i(\boldsymbol{k}) - \overline{\rho}_i(\boldsymbol{k}). \tag{12}$$

The deviations $\Delta\rho_i(\boldsymbol{k})$ in the Fourier amplitudes of the weighted density are convolutions of the weighting and the deviation in the density

$$\Delta\rho_i(\boldsymbol{k}) = \sum_{\boldsymbol{k}'} w_i(\boldsymbol{k}')\Delta\rho(\boldsymbol{k} - \boldsymbol{k}'), \tag{13}$$

similarly to equation (7).

The expected covariance between two weighted densities $\rho_i(\boldsymbol{k}_1)$ and $\rho_j(\boldsymbol{k}_2)$ at wavenumbers $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$ is, from equations (4) and (13),

$$\left\langle\Delta\rho_i(\boldsymbol{k}_1)\,\Delta\rho_j(\boldsymbol{k}_2)\right\rangle = \sum_{\boldsymbol{k}'} w_i(\boldsymbol{k}')w_j(\boldsymbol{k}_1 + \boldsymbol{k}_2 - \boldsymbol{k}')P(\boldsymbol{k}_1 - \boldsymbol{k}'). \tag{14}$$

The weighting breaks statistical homogeneity, so the expected covariance matrix of Fourier amplitudes $\rho_i(\boldsymbol{k})$, equation (14), is not diagonal. Nevertheless, we *define* the power spectrum $P_i(\boldsymbol{k})$ of the ith weighted density by the diagonal elements of the covariance matrix, the variance

$$P_i(\boldsymbol{k}) \equiv \langle\Delta\rho_i(\boldsymbol{k})\,\Delta\rho_i(-\boldsymbol{k})\rangle\,. \tag{15}$$

Note that this definition (15) of the power spectrum $P_i(\boldsymbol{k})$ differs from the usual definition of power in that the deviations $\Delta\rho_i(\boldsymbol{k})$ on the right are Fourier transforms of the deviations $\Delta\rho_i(\boldsymbol{r})$ *not* divided by the mean density $\overline{\rho}_i(\boldsymbol{r}) = w_i(\boldsymbol{r})$ (dividing by the mean density would simply unweight the weighting, defeating the whole point of the procedure). The power spectrum $P_i(\boldsymbol{k})$ defined by equation (15) is related to the true power spectrum $P(\boldsymbol{k})$ by, equation (14),

$$P_i(\boldsymbol{k}) = \sum_{\boldsymbol{k}'} |w_i(\boldsymbol{k}')|^2 P(\boldsymbol{k} - \boldsymbol{k}'). \tag{16}$$

Now make the approximation that the power spectrum $P(\boldsymbol{k} - \boldsymbol{k}')$ at the wavenumber $\boldsymbol{k} - \boldsymbol{k}'$ displaced by $\boldsymbol{k}'$ from $\boldsymbol{k}$ is approximately equal to the power spectrum $P(\boldsymbol{k})$ at the undisplaced wavenumber $\boldsymbol{k}$,

$$P(\boldsymbol{k} - \boldsymbol{k}') \approx P(\boldsymbol{k}). \tag{17}$$

This approximation is good provided that the power spectrum $P(\boldsymbol{k})$ is slowly varying as a function of wavenumber $\boldsymbol{k}$, and that the displacement $\boldsymbol{k}'$ is small compared to $\boldsymbol{k}$. In Section 3, we constrain the

weightings $w_i(\boldsymbol{k}')$ to contain only fundamental modes, $\boldsymbol{k}' = \{k'_x, k'_y, k'_z\}$ with $k'_x, k'_y, k'_z = 0, \pm 1$, so that the displacement $\boldsymbol{k}'$ is as small as it can be without being zero, and the approximation (17) is therefore as good as it can be. The approximation (17) becomes exact in the case of a constant, or shot noise, power spectrum $P(\boldsymbol{k})$, except at $\boldsymbol{k} - \boldsymbol{k}' = \boldsymbol{0}$.

Under approximation (17), the power spectrum of the ith weighted density is

$$P_i(\boldsymbol{k}) \approx P(\boldsymbol{k}) \sum_{\boldsymbol{k}'} |w_i(\boldsymbol{k}')|^2 \tag{18}$$

which is just proportional to the true power spectrum $P(\boldsymbol{k})$.

Without loss of generality, let each weighting $w_i(\boldsymbol{k}')$ be normalized so that the factor on the right-hand side of equation (18) is unity,

$$\sum_{\boldsymbol{k}'} |w_i(\boldsymbol{k}')|^2 = 1. \tag{19}$$

Then the power spectrum $P_i(\boldsymbol{k})$ of the weighted density is approximately equal to the true power spectrum $P(\boldsymbol{k})$,

$$P_i(\boldsymbol{k}) \approx P(\boldsymbol{k}). \tag{20}$$

Thus, in the approximation (17) and with the normalization (19), measurements of the power spectrum $P_i(\boldsymbol{k})$ of weighted densities provide estimates of the true power spectrum $P(\boldsymbol{k})$. The plan is to use the scatter in the estimates of power over a set of weightings to estimate the covariance matrix of power.

2.3 The covariance of power spectra

Let $\widehat{P}(\boldsymbol{k})$ denote the power spectrum of unweighted density at wavevector $\boldsymbol{k}$ measured from a simulation, the hat distinguishing it from the true power spectrum $P(\boldsymbol{k})$:

$$\widehat{P}(\boldsymbol{k}) \equiv \Delta\rho(\boldsymbol{k})\Delta\rho(-\boldsymbol{k}). \tag{21}$$

Below, Section 2.5, we will invoke statistical isotropy, and we will average over a shell in k space, but in equation (21) there is no averaging because there is just one simulation, and just one specific wavenumber $\boldsymbol{k}$. Because of statistical fluctuations, the estimate $\widehat{P}(\boldsymbol{k})$ will in general differ from the true power $P(\boldsymbol{k})$, but by definition the expectation value of the estimate equals the true value, $\langle\widehat{P}(\boldsymbol{k})\rangle = P(\boldsymbol{k})$. The deviation $\Delta\widehat{P}(\boldsymbol{k})$ in the power is the difference between the measured and expected value:

$$\Delta\widehat{P}(\boldsymbol{k}) \equiv \widehat{P}(\boldsymbol{k}) - P(\boldsymbol{k}). \tag{22}$$

The expected covariance of power involves the covariance of the covariance of unweighted densities,

$$\begin{aligned}&\langle[\Delta\rho(\boldsymbol{k}_1)\Delta\rho(\boldsymbol{k}_2) - 1_{\boldsymbol{k}_1+\boldsymbol{k}_2}P(\boldsymbol{k}_1)]\\ &\times[\Delta\rho(\boldsymbol{k}_3)\Delta\rho(\boldsymbol{k}_4) - 1_{\boldsymbol{k}_3+\boldsymbol{k}_4}P(\boldsymbol{k}_3)]\rangle\\ &= (1_{\boldsymbol{k}_1+\boldsymbol{k}_3}1_{\boldsymbol{k}_2+\boldsymbol{k}_4} + 1_{\boldsymbol{k}_1+\boldsymbol{k}_4}1_{\boldsymbol{k}_2+\boldsymbol{k}_3})P(\boldsymbol{k}_1)P(\boldsymbol{k}_2)\\ &\quad + 1_{\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3+\boldsymbol{k}_4}T(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3, \boldsymbol{k}_4),\end{aligned} \tag{23}$$

which is a sum of a reducible, Gaussian part, the terms proportional to $P(\boldsymbol{k}_1)\,P(\boldsymbol{k}_2)$, and an irreducible, non-Gaussian part, the term involving the trispectrum $T(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3, \boldsymbol{k}_4)$. Equation (23) essentially defines what is meant by the trispectrum T. Exchange symmetry implies that the trispectrum function is invariant under permutations of its four arguments. The momentum-conserving delta-function $1_{k1} + \boldsymbol{k}_2 + \boldsymbol{k}_3 + \boldsymbol{k}_4$ in front of the trispectrum T expresses translation invariance.

It follows from equation (23) that the expected covariance of estimates of power is

$$\begin{aligned}&\langle\Delta\widehat{P}(\boldsymbol{k}_1)\Delta\widehat{P}(\boldsymbol{k}_2)\rangle\\ &= \left(1_{\boldsymbol{k}_1+\boldsymbol{k}_2} + 1_{\boldsymbol{k}_1-\boldsymbol{k}_2}\right)P(\boldsymbol{k}_1)^2 + T(\boldsymbol{k}_1, -\boldsymbol{k}_1, \boldsymbol{k}_2, -\boldsymbol{k}_2).\end{aligned} \tag{24}$$

2.4 The covariance of power spectra of weighted density

Similarly to equations (21) and (22), let $\widehat{P}_i(\boldsymbol{k})$ denote the power spectrum of the ith weighted density at wavevector $\boldsymbol{k}$ measured from a simulation,

$$\widehat{P}_i(\boldsymbol{k}) \equiv \Delta\rho_i(\boldsymbol{k})\Delta\rho_i(-\boldsymbol{k}), \tag{25}$$

and let $\Delta\widehat{P}_i(\boldsymbol{k})$ denote the deviation between the measured and expected value,

$$\Delta\widehat{P}_i(\boldsymbol{k}) \equiv \widehat{P}_i(\boldsymbol{k}) - P_i(\boldsymbol{k}). \tag{26}$$

The expected covariance between the power spectra of the ith and jth weighted densities is, from equations (13) and (23),

$$\begin{aligned}&\langle\Delta\widehat{P}_i(\boldsymbol{k}_1)\Delta\widehat{P}_j(\boldsymbol{k}_2)\rangle\\ &= \sum_{\boldsymbol{k}'_1+\boldsymbol{k}''_1+\boldsymbol{k}'_2+\boldsymbol{k}''_2=\boldsymbol{0}} w_i\left(\boldsymbol{k}'_1\right)w_i\left(\boldsymbol{k}''_1\right)w_j\left(\boldsymbol{k}'_2\right)w_j\left(\boldsymbol{k}''_2\right)\\ &\quad\times\left[(1_{\boldsymbol{k}_1-\boldsymbol{k}'_1+\boldsymbol{k}_2-\boldsymbol{k}'_2} + 1_{\boldsymbol{k}_1-\boldsymbol{k}'_1-\boldsymbol{k}_2-\boldsymbol{k}''_2})P\left(\boldsymbol{k}_1 - \boldsymbol{k}'_1\right)P\left(-\boldsymbol{k}_1 - \boldsymbol{k}''_1\right)\right.\\ &\quad\left. + T\left(\boldsymbol{k}_1 - \boldsymbol{k}'_1, -\boldsymbol{k}_1 - \boldsymbol{k}''_1, \boldsymbol{k}_2 - \boldsymbol{k}'_2, -\boldsymbol{k}_2 - \boldsymbol{k}''_2\right)\right].\end{aligned} \tag{27}$$

2.5 The covariance of shell-averaged power spectra

Assume now that the unweighted density field $\overline{\rho}(\boldsymbol{r})$ is statistically isotropic, so that the true power spectrum $P(k)$ is a function only of the absolute value $k \equiv |\boldsymbol{k}|$ of its argument. In estimating the power $P(k)$ from a simulation, one would typically average the measured power over a spherical shell V_k of wavenumbers in k space. Actually the arguments below generalize immediately to the case where the power is not isotropic, in which case V_k might be chosen to be some localized patch in k space. However, we shall assume isotropy, and refer to V_k as a shell.

Let $\hat{p}(k)$ denote the measured power averaged over a shell V_k about scalar wavenumber k (the estimated shell-averaged power $\hat{p}(k)$ is written in lowercase to distinguish it from the estimate $\widehat{P}(\boldsymbol{k})$ of power at a single specific wavevector $\boldsymbol{k}$):

$$\hat{p}(k) \equiv \frac{1}{N_{\boldsymbol{k}}}\sum_{\boldsymbol{k}\in V_k}\widehat{P}(\boldsymbol{k}). \tag{28}$$

Here $N_{\boldsymbol{k}}$ is the number of modes $\rho(\boldsymbol{k})$ in the shell V_k. We count $\rho(\boldsymbol{k})$ and its complex conjugate $\rho(-\boldsymbol{k})$ as contributing two distinct modes, the real and imaginary parts of $\rho(\boldsymbol{k})$. The expectation value of the estimates $\hat{p}(k)$ of shell-averaged power equals the true shell-averaged power $p(k)$,

$$\langle\hat{p}(k)\rangle = p(k) \equiv \frac{1}{N_{\boldsymbol{k}}}\sum_{\boldsymbol{k}\in V_k}P(\boldsymbol{k}). \tag{29}$$

The deviation $\Delta\hat{p}(k)$ between the measured and expected value of shell-averaged power is

$$\Delta\hat{p}(k) \equiv \hat{p}(k) - p(k) = \frac{1}{N_{\boldsymbol{k}}}\sum_{\boldsymbol{k}\in V_k}\Delta\widehat{P}(\boldsymbol{k}). \tag{30}$$

The expected covariance of shell-averaged estimates of power is, from equations (30) and (24),

$$\langle \Delta\hat{p}(k_1)\Delta\hat{p}(k_2)\rangle = \frac{1}{N_{\boldsymbol{k}_1} N_{\boldsymbol{k}_2}} \left[2 \sum_{\boldsymbol{k}_1 \in V_{k_1} \cap V_{k_2}} P(\boldsymbol{k}_1)^2 + \sum_{\boldsymbol{k}_1 \in V_{k_1},\ \boldsymbol{k}_2 \in V_{k_2}} T(\boldsymbol{k}_1, -\boldsymbol{k}_1, \boldsymbol{k}_2, -\boldsymbol{k}_2) \right]. \quad (31)$$

In the usual case, the shells V_k would be taken to be non-overlapping, in which case the intersection $V_{k_1} \cap V_{k_2}$ in equation (31) is equal either to V_{k_1} if V_{k_1} and V_{k_2} are the same shell, or to the empty set if V_{k_1} and V_{k_2} are different shells.

2.6 The covariance of shell-averaged power spectra of weighted density

Similarly to equation (28), let $\hat{p}_i(k)$ denote the measured shell-averaged power spectrum of the ith weighted density at wavenumber k,

$$\hat{p}_i(k) \equiv \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k} \in V_k} \widehat{P}_i(\boldsymbol{k}). \quad (32)$$

The expectation value of the estimates $\hat{p}_i(k)$ is (compare equation 29)

$$\langle \hat{p}_i(k)\rangle = p_i(k) \equiv \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k} \in V_k} P_i(\boldsymbol{k}). \quad (33)$$

In the approximation (17) of a slowly varying power spectrum, and with the normalization (19), the expected shell-averaged power spectrum $p_i(k)$ of the weighted density is approximately equal to the shell-averaged power spectrum $p(k)$ of the unweighted density (compare equation 20)

$$p_i(k) \approx p(k). \quad (34)$$

The deviation $\Delta\hat{p}_i(k)$ between the measured and expected values is (compare equation 30)

$$\Delta\hat{p}_i(k) \equiv \hat{p}_i(k) - p_i(k) = \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k} \in V_k} \Delta\widehat{P}_i(\boldsymbol{k}). \quad (35)$$

The expected covariance of shell-averaged power spectra of weighted densities is, from equations (35) and (27),

$$\left\langle \Delta\hat{p}_i(k_1)\Delta\hat{p}_j(k_2)\right\rangle = \frac{1}{N_{\boldsymbol{k}_1} N_{\boldsymbol{k}_2}} \times \sum_{\boldsymbol{k}_1' + \boldsymbol{k}_1'' + \boldsymbol{k}_2' + \boldsymbol{k}_2'' = \boldsymbol{0}} w_i\left(\boldsymbol{k}_1'\right) w_i\left(\boldsymbol{k}_1''\right) w_j\left(\boldsymbol{k}_2'\right) w_j\left(\boldsymbol{k}_2''\right) \sum_{\boldsymbol{k}_1 \in V_{k_1},\ \boldsymbol{k}_2 \in V_{k_2}} \times \left[\left(1_{\boldsymbol{k}_1 - \boldsymbol{k}_1' + \boldsymbol{k}_2 - \boldsymbol{k}_2'} + 1_{\boldsymbol{k}_1 - \boldsymbol{k}_1' - \boldsymbol{k}_2 - \boldsymbol{k}_2''}\right) P\left(\boldsymbol{k}_1 - \boldsymbol{k}_1'\right) P\left(-\boldsymbol{k}_1 - \boldsymbol{k}_1''\right) + T\left(\boldsymbol{k}_1 - \boldsymbol{k}_1', -\boldsymbol{k}_1 - \boldsymbol{k}_1'', \boldsymbol{k}_2 - \boldsymbol{k}_2', -\boldsymbol{k}_2 - \boldsymbol{k}_2''\right) \right]. \quad (36)$$

Assume, analogously to approximation (17) for the power spectrum, that the trispectrum function $T(\boldsymbol{k}_1 - \boldsymbol{k}_1', -\boldsymbol{k}_1 - \boldsymbol{k}_1'', \boldsymbol{k}_2 - \boldsymbol{k}_2', -\boldsymbol{k}_2 - \boldsymbol{k}_2'')$ in equation (36) is sufficiently slowly varying, and the displacements $\boldsymbol{k}_1', \boldsymbol{k}_1'', \boldsymbol{k}_2', \boldsymbol{k}_2''$ sufficiently small, that

$$T(\boldsymbol{k}_1 - \boldsymbol{k}_1', -\boldsymbol{k}_1 - \boldsymbol{k}_1'', \boldsymbol{k}_2 - \boldsymbol{k}_2', -\boldsymbol{k}_2 - \boldsymbol{k}_2'') \approx T(\boldsymbol{k}_1, -\boldsymbol{k}_1, \boldsymbol{k}_2, -\boldsymbol{k}_2). \quad (37)$$

In Section 4, we will revisit the approximation (37), and show that in fact it is not true, in a way that proves to be interesting and observationally relevant. In this section and the next, Section 3, however, we will continue to assume that the approximation (37) is valid.

In the approximations (17) and (37) that the power spectrum and trispectrum are both approximately constant for small displacements of their arguments, the covariance of shell-averaged power spectra, equation (36), becomes

$$\langle \Delta\hat{p}_i(k_1)\Delta\hat{p}_j(k_2)\rangle \approx \frac{1}{N_{\boldsymbol{k}_1} N_{\boldsymbol{k}_2}} \times \sum_{\boldsymbol{k}_1' + \boldsymbol{k}_1'' + \boldsymbol{k}_2' + \boldsymbol{k}_2'' = \boldsymbol{0}} w_i(\boldsymbol{k}_1') w_i(\boldsymbol{k}_1'') w_j(\boldsymbol{k}_2') w_j(\boldsymbol{k}_2'') \times \sum_{\boldsymbol{k}_1 \in V_{k_1},\ \boldsymbol{k}_2 \in V_{k_2}} \left[\left(1_{\boldsymbol{k}_1 - \boldsymbol{k}_1' + \boldsymbol{k}_2 - \boldsymbol{k}_2'} + 1_{\boldsymbol{k}_1 - \boldsymbol{k}_1' - \boldsymbol{k}_2 - \boldsymbol{k}_2''}\right) P(\boldsymbol{k}_1)^2 + T(\boldsymbol{k}_1, -\boldsymbol{k}_1, \boldsymbol{k}_2, -\boldsymbol{k}_2) \right]. \quad (38)$$

Consider the Gaussian (P^2) part of this expression (38). In the true covariance of shell-averaged power, equation (31), the Gaussian part of the covariance is a diagonal matrix, with zero covariance between non-overlapping shells. By contrast, the Gaussian part of the covariance of power of weighted densities, equation (38), is not quite diagonal. In effect, the Gaussian variance in each shell is smeared by convolution with the weighting function, causing some of the Gaussian variance near the boundaries of adjacent shells to leak into covariance between the shells. In Section 3, we advocate restricting the weightings $w_i(\boldsymbol{k})$ to contain only fundamental modes, which keeps smearing to a minimum. Whatever the case, if each shell V_k is broad compared the extent of the weightings $w_i(\boldsymbol{k})$ in k space, then the smearing is relatively small, and can be approximated as zero. Mathematically, this broad-shell approximation amounts to approximating

$$\sum_{\boldsymbol{k}_1 \in V_{k_1},\ \boldsymbol{k}_2 \in V_{k_2}} 1_{\boldsymbol{k}_1 - \boldsymbol{k}_1' + \boldsymbol{k}_2 - \boldsymbol{k}_2'} + 1_{\boldsymbol{k}_1 - \boldsymbol{k}_1' - \boldsymbol{k}_2 - \boldsymbol{k}_2''} \approx \sum_{\boldsymbol{k}_1 \in V_{k_1},\ \boldsymbol{k}_2 \in V_{k_2}} 1_{\boldsymbol{k}_1 + \boldsymbol{k}_2} + 1_{\boldsymbol{k}_1 - \boldsymbol{k}_2} = \sum_{\boldsymbol{k}_1 \in V_{k_1} \cap V_{k_2}} 2. \quad (39)$$

In the broad-shell approximation (39), the expected covariance of shell-averaged power spectra of weighted densities, equation (38), simplifies to

$$\langle \Delta\hat{p}_i(k_1)\Delta\hat{p}_j(k_2)\rangle \approx f_{ij} \langle \Delta\hat{p}(k_1)\Delta\hat{p}(k_2)\rangle, \quad (40)$$

where the factor f_{ij} is

$$f_{ij} \equiv \sum_{\boldsymbol{k}_1' + \boldsymbol{k}_1'' + \boldsymbol{k}_2' + \boldsymbol{k}_2'' = \boldsymbol{0}} w_i(\boldsymbol{k}_1') w_i(\boldsymbol{k}_1'') w_j(\boldsymbol{k}_2') w_j(\boldsymbol{k}_2''). \quad (41)$$

In real (as opposed to Fourier) space, the factor f_{ij} is

$$f_{ij} = \int w_i(\boldsymbol{r})^2 w_j(\boldsymbol{r})^2 \, \mathrm{d}^3 r. \quad (42)$$

Equation (40) is the most basic result of the present paper. It states that the expected covariance between estimates of power from various weightings is proportional to the true covariance matrix of power. The nice thing about the result (40) is that the constant of proportionality f_{ij} depends only on the weightings $w_i(\boldsymbol{k})$ and $w_j(\boldsymbol{k})$, and is independent both of the power spectrum $P(\boldsymbol{k})$ and of the wavenumbers k_1 and k_2 in the covariance $\langle \Delta\hat{p}_i(k_1)\Delta\hat{p}_j(k_2)\rangle$.

2.7 The covariance of the covariance of shell-averaged power spectra of weighted density

Equation (40) provides the formal mathematical justification for estimating the covariance of power from the scatter in estimates of

power over an ensemble of weightings of density. In Section 3, we will craft the weightings $w_i(\boldsymbol{k})$ so as to minimize the expected variance of the estimated covariance of power. The resulting weightings are 'best possible', within the framework of the technique. To determine the minimum variance estimator, it is necessary to have an expression for the (co)variance of the covariance of power, which we now derive.

The expected covariance between estimates $\Delta\hat{p}_i(k_1)\Delta\hat{p}_i(k_2)$ of covariance of power is a covariance of covariance of covariance of densities, an eight-point object. This object involves, in addition to the eight-point function, a linear combination of products of lower-order functions adding to eight points. The types of terms are (cf. Verde & Heavens 2001)

$$2^4, \quad 2\cdot 3^2, \quad 2^2\cdot 4, \quad 2\cdot 6, \quad 3\cdot 5, \quad 4^2, \quad 8 \tag{43}$$

in which 2^4 signifies a product of four two-point functions, $2\cdot 3^2$ signifies a product of a two-point function with two three-point functions, and so on, up to eight, which signifies the eight-point function. We do not pause to write out all the terms explicitly, because in the same slowly varying and broad-shell approximations that led to equation (40), the covariance of the covariance of power spectra of weighted densities simplifies to

$$\begin{aligned}&\langle[\Delta\hat{p}_i(k_1)\Delta\hat{p}_i(k_2)-\langle\Delta\hat{p}_i(k_1)\Delta\hat{p}_i(k_2)\rangle]\\&\quad\times[\Delta\hat{p}_j(k_3)\Delta\hat{p}_j(k_4)-\langle\Delta\hat{p}_j(k_3)\Delta\hat{p}_j(k_4)\rangle]\rangle\\&\quad\approx g_{ij}\langle[\Delta\hat{p}(k_1)\Delta\hat{p}(k_2)-\langle\Delta\hat{p}(k_1)\Delta\hat{p}(k_2)\rangle]\\&\quad\times[\Delta\hat{p}(k_3)\Delta\hat{p}(k_4)-\langle\Delta\hat{p}(k_3)\Delta\hat{p}(k_4)\rangle]\rangle,\end{aligned} \tag{44}$$

where g_{ij} is, analogously to equation (41),

$$\begin{aligned}g_{ij}\equiv&\sum_{\boldsymbol{k}_1'+\boldsymbol{k}_1''+\boldsymbol{k}_2'+\boldsymbol{k}_2''+\boldsymbol{k}_3'+\boldsymbol{k}_3''+\boldsymbol{k}_4'+\boldsymbol{k}_4''=\boldsymbol{0}}\\&\times w_i(\boldsymbol{k}_1')w_i(\boldsymbol{k}_1'')w_i(\boldsymbol{k}_2')w_i(\boldsymbol{k}_2'')w_j(\boldsymbol{k}_3')w_j(\boldsymbol{k}_3'')w_j(\boldsymbol{k}_4')w_j(\boldsymbol{k}_4'').\end{aligned} \tag{45}$$

In real (as opposed to Fourier) space, the factors g_{ij} are

$$g_{ij}=\int w_i(\boldsymbol{r})^4 w_j(\boldsymbol{r})^4\,\mathrm{d}^3r. \tag{46}$$

Equation (44) states, analogously to equation (40), that the expected covariance of covariance of power spectra of weighted densities is proportional to the true covariance of covariance of power. As with the factors f_{ij}, equation (41), the constants of proportionality g_{ij}, equation (45), depend only on the weightings $w_i(\boldsymbol{k})$ and $w_j(\boldsymbol{k})$, and are independent of the power spectrum $P(\boldsymbol{k})$ or of any of the higher order functions, and are also independent of the wavenumbers $k_1,\ldots,k_4$ in the covariance, a gratifyingly simple result.

2.8 Subtracting the mean power

The deviation $\Delta\hat{p}_i(k)$ of the shell-averaged power spectrum of the ith weighted density was defined above, equation (35), to be the difference between the measured value $\hat{p}_i(k)$ and the expected value $p_i(k)$ of shell-averaged power. However, the expected power spectrum $p_i(k)$ (the true power spectrum) is probably unknown. Even if the true power spectrum is known in the linear regime (because the simulation was set up with a known linear power spectrum), the true power spectrum in the non-linear regime is not known precisely, but must be estimated from the simulation.

In practice, therefore, it is necessary to measure the deviation in power not from the true value, but rather from some estimated mean value. Two strategies naturally present themselves. The first strategy is to take the mean power spectrum to be the measured power spectrum $\hat{p}(k)$ of the unweighted density of the simulation. In this case the deviation $\Delta\hat{p}_i'(k)$ between the measured shell-averaged power spectra of the weighted and unweighted densities is (the deviation $\Delta\hat{p}_i'(k)$ is primed to distinguish it from the deviation $\Delta\hat{p}_i(k)$, equation 35)

$$\Delta\hat{p}_i'(k)\equiv\hat{p}_i(k)-\hat{p}(k). \tag{47}$$

The second strategy is to take the mean power spectrum to be the average over weightings of the measured power spectra of weighted densities, $N^{-1}\sum_i\hat{p}_i(k)$. In this case the deviation $\Delta\hat{p}_i'(k)$ between the measured shell-averaged power spectra and their average is (with the same primed notation for the deviation $\Delta\hat{p}_i'(k)$ as in equation 47; it is up to the user to decide which strategy to adopt)

$$\Delta\hat{p}_i'(k)\equiv\hat{p}_i(k)-\frac{1}{N}\sum_i\hat{p}_i(k). \tag{48}$$

The advantage of the first strategy, equation (47), is that the power spectrum $\hat{p}(k)$ of the unweighted density is the most accurate (by symmetry) estimate of the power spectrum that can be measured from a single simulation. Its disadvantage is that measurements of power spectra of weighted densities yield (slightly) biased estimates of the power spectrum of unweighted density, because the approximation (17) can lead to a slight bias if, as is typical, the power spectrum $P(\boldsymbol{k})$ is not constant. In other words, the approximation $p_i(k)\approx p(k)$, equation (34), is not an exact equality. Although the bias is likely to be small, it contributes systematically to estimates of deviations of power, causing the covariance of power to be systematically overestimated. The second strategy, equation (48), is unaffected by this bias, but the statistical uncertainty is slightly larger. Probably the sensible thing to do is to apply both strategies, and to check that they yield consistent results.

To allow a concise expression for the covariance of power to be written down, it is convenient to introduce $v_i(\boldsymbol{k})$, defined to be the Fourier transform of the squared real-space weighting, $v_i(\boldsymbol{r})\equiv w_i(\boldsymbol{r})^2$,

$$v_i(\boldsymbol{k})\equiv\sum_{\boldsymbol{k}'+\boldsymbol{k}''=\boldsymbol{k}}w_i(\boldsymbol{k}')w_i(\boldsymbol{k}''). \tag{49}$$

The normalization condition (19) on the weightings $w_i(\boldsymbol{k})$ is equivalent to requiring

$$v_i(\boldsymbol{0})=1. \tag{50}$$

In terms of $v_i(\boldsymbol{k})$, the factors f_{ij}, equation (41), relating the expected covariance matrix of power spectra of weighted densities to the true covariance matrix of power are

$$f_{ij}=\sum_{\boldsymbol{k}}v_i(\boldsymbol{k})v_j(-\boldsymbol{k}). \tag{51}$$

An expression is desired for the covariance of power in terms of the deviations $\Delta\hat{p}_i'(k)$, equation (47) or (48), instead of $\Delta\hat{p}_i(k)$. For this, a modified version of $v_i(\boldsymbol{k})$ is required. For strategy one, equation (47),

$$v_i'(\boldsymbol{k})=\begin{cases}0 & (\boldsymbol{k}=\boldsymbol{0}),\\ v_i(\boldsymbol{k}) & (\boldsymbol{k}\neq\boldsymbol{0}),\end{cases} \tag{52}$$

whereas for strategy two, equation (48),

$$v_i'(\boldsymbol{k})=v_i(\boldsymbol{k})-\frac{1}{N}\sum_i v_i(\boldsymbol{k}). \tag{53}$$

In either case, the expected covariance $\langle\Delta\hat{p}'_i(k_1)\Delta\hat{p}'_j(k_2)\rangle$ of estimates of shell-averaged power spectra is related to the true covariance $\langle\Delta\hat{p}(k_1)\Delta\hat{p}(k_2)\rangle$ of shell-averaged power by (compare equation 40)

$$\left\langle\Delta\hat{p}'_i(k_1)\Delta\hat{p}'_j(k_2)\right\rangle \approx f'_{ij}\left\langle\Delta\hat{p}(k_1)\Delta\hat{p}(k_2)\right\rangle, \tag{54}$$

where the factors f'_{ij} are (compare equation 51)

$$f'_{ij} = \sum_{\boldsymbol{k}} v'_i(\boldsymbol{k})v'_j(-\boldsymbol{k}). \tag{55}$$

The approximation (54) is valid under the same assumptions made in deriving the approximation (40), namely the slowly varying approximations (17) and (37), and the broad-shell approximation (39).

2.9 Subtracting the mean covariance of power

The expression (45) for the covariance of covariance of power must likewise be modified to allow for the fact that the deviations in power must be measured as deviations not from the true power spectrum but from either (strategy 1) the power spectrum of the unweighted density, or (strategy 2) the averaged power spectrum of the weighted densities.

For this purpose it is convenient to define $u_i(\boldsymbol{k})$ to be the Fourier transform of the fourth power of the real-space weighting, $u_i(\boldsymbol{r}) \equiv v_i(\boldsymbol{r})^2 = w_i(\boldsymbol{r})^4$,

$$u_i(\boldsymbol{k}) \equiv \sum_{\boldsymbol{k}'+\boldsymbol{k}''=\boldsymbol{k}} v_i(\boldsymbol{k}')v_i(\boldsymbol{k}''). \tag{56}$$

In terms of $u_i(\boldsymbol{k})$, the factors g_{ij}, equation (45), relating the expected covariance of covariance of power spectra of weighted densities to the true covariance of covariance of power are

$$g_{ij} = \sum_{\boldsymbol{k}} u_i(\boldsymbol{k})u_j(-\boldsymbol{k}). \tag{57}$$

To write down an expression for the covariance of the covariance of the deviations $\Delta\hat{p}'_i(k)$ instead of $\Delta\hat{p}_i(k)$, define a modified version $u'_i(\boldsymbol{k})$ of $u_i(\boldsymbol{k})$ by

$$u'_i(\boldsymbol{k}) \equiv \sum_{\boldsymbol{k}'+\boldsymbol{k}''=\boldsymbol{k}} v'_i(\boldsymbol{k}')v'_i(\boldsymbol{k}'') \tag{58}$$

which is the same as equation (56) but with primed $v'_i(\boldsymbol{k})$, equation (52) or (53), in place of $v_i(\boldsymbol{k})$. Then the covariance of the covariance of the deviations $\Delta\hat{p}'_i(k)$ is related to the true covariance of covariance of shell-averaged power by (compare equation 44)

$$\begin{aligned}&\left\langle\left[\Delta\hat{p}'_i(k_1)\Delta\hat{p}'_i(k_2) - \left\langle\Delta\hat{p}'_i(k_1)\Delta\hat{p}'_i(k_2)\right\rangle\right]\right.\\ &\quad\left.\times\left[\Delta\hat{p}'_j(k_3)\Delta\hat{p}'_j(k_4) - \left\langle\Delta\hat{p}'_j(k_3)\Delta\hat{p}'_j(k_4)\right\rangle\right]\right\rangle\\ &\quad\approx g'_{ij}\langle[\Delta\hat{p}(k_1)\Delta\hat{p}(k_2) - \langle\Delta\hat{p}(k_1)\Delta\hat{p}(k_2)\rangle]\\ &\quad\left.\times[\Delta\hat{p}(k_3)\Delta\hat{p}(k_4) - \langle\Delta\hat{p}(k_3)\Delta\hat{p}(k_4)\rangle]\right\rangle,\end{aligned} \tag{59}$$

where the factors g'_{ij} are (compare equation 57)

$$g'_{ij} = \sum_{\boldsymbol{k}} u'_i(\boldsymbol{k})u'_j(-\boldsymbol{k}). \tag{60}$$

Equation (59) gives the expected covariance of the difference between the estimate of covariance $\Delta\hat{p}'_i(k_1)\Delta\hat{p}'_i(k_2)$ and its expectation value $\langle\Delta\hat{p}'_i(k_1)\Delta\hat{p}'_i(k_2)\rangle$, but this latter expectation value is again an unknown quantity. What can actually be measured is the difference between the estimate $\Delta\hat{p}'_i(k_1)\Delta\hat{p}'_i(k_2)$ and its average over weightings $N^{-1}\sum_i \Delta\hat{p}'_i(k_1)\Delta\hat{p}'_i(k_2)$. To write down an expression for the covariance of the covariance relative to the weightings-averaged covariance rather than the expected covariance, define a modified version $u''_i(\boldsymbol{k})$ of $u'_i(\boldsymbol{k})$, equation (58), by

$$u''_i(\boldsymbol{k}) \equiv u'_i(\boldsymbol{k}) - \frac{1}{N}\sum_i u'_i(\boldsymbol{k}). \tag{61}$$

Then the covariance of the covariance of the deviations $\Delta\hat{p}'_i(k)$ is related to the true covariance of covariance of shell-averaged power by (compare equations 44 and 59)

$$\begin{aligned}&\left\langle\left[\Delta\hat{p}'_i(k_1)\Delta\hat{p}'_i(k_2) - \frac{1}{N}\sum_k \Delta\hat{p}'_k(k_1)\Delta\hat{p}'_k(k_2)\right]\right.\\ &\quad\left.\times\left[\Delta\hat{p}'_j(k_3)\Delta\hat{p}'_j(k_4) - \frac{1}{N}\sum_l \Delta\hat{p}'_l(k_3)\Delta\hat{p}'_l(k_4)\right]\right\rangle\\ &\quad\approx g''_{ij}\langle[\Delta\hat{p}(k_1)\Delta\hat{p}(k_2) - \langle\Delta\hat{p}(k_1)\Delta\hat{p}(k_2)\rangle]\\ &\quad\times[\Delta\hat{p}(k_3)\Delta\hat{p}(k_4) - \langle\Delta\hat{p}(k_3)\Delta\hat{p}(k_4)\rangle]\rangle,\end{aligned} \tag{62}$$

where the factors g''_{ij} are (compare equations 57 and 60)

$$g''_{ij} = \sum_{\boldsymbol{k}} u''_i(\boldsymbol{k})u''_j(-\boldsymbol{k}). \tag{63}$$

Approximations (59) and (62) are valid under the same approximations as approximations (40) and (44), namely the slowly varying approximations (17) and (37), and the broad-shell approximation (39).

3 MINIMUM VARIANCE WEIGHTINGS

It was shown in Section 2 that the expected covariance between shell-averaged power spectra of weighted densities is proportional to the true covariance of shell-average power, equation (54). It follows that the scatter in estimates of power from different weightings can be used to estimate the true covariance of power. In this section, we use minimum variance arguments to derive a set of 52 weightings, equation (71), which we recommend, Section 3.6, for practical application.

In this section as in the previous one, Section 2, we continue to ignore the beat-coupling contributions to the (covariance of) covariance of power. These beat-couplings are discussed in Section 4, which in Section 4.6 concludes that the minimum variance weightings derived in the present section, although no longer precisely minimum variance, should be satisfactory for practical use.

3.1 Fundamentals and symmetries

In the first place, we choose to use weightings $w_i(\boldsymbol{k})$ that contain only combinations of fundamental modes, that is, $\boldsymbol{k} = \{k_x, k_y, k_z\}$ with k_x, k_y, k_z running over 0, ±1. By restricting the weightings to fundamental modes only, we ensure that the two approximations required for equation (54) to be valid are as good as can be. The first approximation was the slowly varying approximation, that both the power spectrum $P(\boldsymbol{k})$ and the trispectrum $T(\boldsymbol{k}_1, -\boldsymbol{k}_1, \boldsymbol{k}_2, -\boldsymbol{k}_2)$ remain approximately constant, equations (17) and (37), when their arguments are displaced by the extent of the weightings $w_i(\boldsymbol{k}')$, that is, by amounts $\boldsymbol{k}'$ for which $w_i(\boldsymbol{k}')$ is non-zero. The second approximation was the broad-shell approximation, that the shells V_k over which the estimated power $\hat{p}_i(k)$ is averaged are broad compared to the extent of the weightings $w_i(\boldsymbol{k}')$, which reduces the relative importance of smearing of Gaussian variance from the edges of adjacent shells into covariance between the shells.

In the second place, we choose to use weightings that are symmetrically related to each other, which seems a natural thing to do given

the cubic symmetry of a periodic box. Choosing a symmetrically related set of weightings not only simplifies practical application of the procedure, but also simplifies the mathematics of determining a best set of Fourier coefficients $w_i(\boldsymbol{k})$, as will be seen in Section 3.2.

There are 48 rotational and reflectional transformations of a cube, corresponding to choosing the x-axis in any of six directions, then the y-axis in any of four directions perpendicular to the x-axis and finally the z-axis in either of the two directions perpendicular to the x- and y-axes.

To the rotational and reflectional transformations, we adjoin the possibility of translations by a fraction (half, quarter, eighth) of a box along any of the three axes, for a net total of $48 \times 8^3 = 24\,576$ possible transformations. In practice, however, the minimum variance weightings $w_i(\boldsymbol{k})$ presented in Section 3.3 prove to possess a high degree of symmetry, greatly reducing the number of distinct weightings.

3.2 How to derive minimum variance weightings

For brevity, let $\widehat{X}_i$ denote an estimate of the covariance of shell-averaged power from the ith weighted density (the arguments k_1 and k_2 on $\widehat{X}_i$ are suppressed, since they play no role in the arguments that follow),

$$\widehat{X}_i \equiv \frac{1}{f'} \Delta\hat{p}'_i(k_1)\Delta\hat{p}'_i(k_2). \tag{64}$$

The quantity f' here is any diagonal element,

$$f' \equiv f'_{ii}, \tag{65}$$

of the matrix of factors f'_{ij} defined by equation (55); the diagonal elements f'_{ii} are identically equal for all i because the weightings $w_i(\boldsymbol{k})$ are by assumption symmetrically related. The factor $1/f'$ in equation (64) ensures that $\widehat{X}_i$ is, in accordance with equation (54), an estimate of the true covariance of shell-averaged power, which we abbreviate X,

$$\langle\widehat{X}_i\rangle \approx X \equiv \langle\Delta\hat{p}(k_1)\Delta\hat{p}(k_2)\rangle\,. \tag{66}$$

The approximation (66) is valid under the assumptions made in deriving equation (54), namely the slowly varying approximations (17) and (37), and the broad-shell approximation (39).

Let N denote the number of weightings. Because the weightings are by assumption symmetrically related, it follows immediately that the best estimate of the true covariance of shell-averaged power $\langle\Delta\hat{p}(k_1)\Delta\hat{p}(k_2)\rangle$ will be a straight average over the ensemble of weightings

$$\widehat{X} = \frac{1}{N}\sum_i \widehat{X}_i. \tag{67}$$

It remains to determine the best Fourier coefficients $w_i(\boldsymbol{k})$ for a representative weighting i. The best set is that which minimizes the expected variance $\langle\Delta\widehat{X}^2\rangle \equiv \langle(\widehat{X} - X)^2\rangle$ of the estimate (67). According to equation (59), this expected variance $\langle\Delta\widehat{X}^2\rangle$ is approximately proportional to a factor that depends on the weightings

$$\langle\Delta\widehat{X}^2\rangle = \frac{1}{N^2}\sum_{ij}\langle\Delta\widehat{X}_i\Delta\widehat{X}_j\rangle \propto \frac{1}{(f'N)^2}\sum_{ij} g'_{ij}, \tag{68}$$

multiplied by another factor that is independent of weightings, namely the true covariance of covariance of power, the expression to the right of the coefficient g'_{ij} in equation (59). Note that the variance $\langle\Delta\widehat{X}^2\rangle$ is the expected variance $\langle(\widehat{X} - X)^2\rangle$ about the true value X, so it is g'_{ij}, equation (60), not g''_{ij}, equation (63), that appears in equation (68).

Equation (68) shows that minimizing the variance $\langle\Delta\widehat{X}^2\rangle$ with respect to the coefficients $w_i(\boldsymbol{k})$ of the weightings is equivalent to minimizing the quantity on the right-hand side of the proportionality (68). From equations (55), (58) and (60) it follows that this factor can be written as

$$\frac{1}{(f'N)^2}\sum_{ij} g'_{ij} = \frac{1}{u'(\boldsymbol{0})^2}\sum_{\boldsymbol{k}} |u'(\boldsymbol{k})|^2, \tag{69}$$

where $u'(\boldsymbol{k})$ denotes the average of $u'_i(\boldsymbol{k})$, equation (58), over weightings

$$u'(\boldsymbol{k}) \equiv \frac{1}{N}\sum_i u'_i(\boldsymbol{k}). \tag{70}$$

Note that $f' = u'(\boldsymbol{0})$. Equation (69) shows that minimizing the variance $\langle\Delta\widehat{X}^2\rangle$ involves computing $u'(\boldsymbol{k})$, equation (70). We evaluate $u'(\boldsymbol{k})$ using an algebraic manipulation program (MATHEMATICA) as follows.

A representative weighting $w_i(\boldsymbol{k})$ contains 27 non-zero Fourier coefficients, since by assumption it contains only combinations of fundamental modes. The coefficients $w_i(\boldsymbol{k})$ and $w_i(-\boldsymbol{k})$, which are complex conjugates of each other, effectively contribute two coefficients, the real and imaginary parts of $w_i(\boldsymbol{k})$.

First, evaluate $v_i(\boldsymbol{k})$, equation (49), in terms of the coefficients $w_i(\boldsymbol{k})$ of the representative weighting. The $v_i(\boldsymbol{k})$ are non-zero for 125 values of $\boldsymbol{k}$, whose components k_x, k_y, k_z run over 0, ±1, ±2. Each $v_i(\boldsymbol{k})$ is a quadratic polynomial in the 27 Fourier coefficients.

Next, modify $v_i(\boldsymbol{k})$ to get $v'_i(\boldsymbol{k})$, equation (52), by setting the coefficient for $\boldsymbol{k} = \boldsymbol{0}$ to zero. Again, each $v'_i(\boldsymbol{k})$ is a quadratic polynomial in the 27 Fourier coefficients. For definiteness, we adopt strategy one, equation (52), rather than strategy two, equation (53). That is, we assume that the deviation $\Delta\hat{p}'_i(k)$ in the power spectrum of the ith weighting of density is being measured relative to the power spectrum of the unweighted density, rather than relative to the average of the power spectra of the weighted densities. In the end it turns out, Section 3.4, that the minimum variance solution is the same for both strategies, so there is no loss in restricting to strategy one.

Next, evaluate $u'_i(\boldsymbol{k})$, equation (58). The $u'_i(\boldsymbol{k})$ are non-zero for 729 values of $\boldsymbol{k}$, whose components k_x, k_y, k_z run over $0, \pm1, \ldots, \pm4$. Each $u'_i(\boldsymbol{k})$ is a quartic polynomial in the 27 Fourier coefficients.

Next, evaluate $u'(\boldsymbol{k})$, equation (70), the average of $u'_i(\boldsymbol{k})$ over weightings i. Consider first averaging $u'_i(\boldsymbol{k})$ over the 48 different rotational and reflectional transformations of the weighting. The averaged result $u'(\boldsymbol{k})$ possesses rotational and reflectional symmetry, so that $u'(\boldsymbol{k})$ is equal to its value at $\boldsymbol{k}$ with components permuted and reflected in such a way that $0 \leqslant k_x \leqslant k_y \leqslant k_z \leqslant 4$, of which there are 35 distinct cases. The rotationally and translationally symmetrized function $u'(\boldsymbol{k})$ can be computed by averaging the values of $u'_i(\boldsymbol{k})$ at 729 values of $\boldsymbol{k}$ into 35 distinct bins.

The symmetrized function satisfies $u'(\boldsymbol{k}) = u'(-\boldsymbol{k})$, so is necessarily real. Thus the absolute value sign around $u'(\boldsymbol{k})^2$ in equation (69) can be omitted.

Now consider averaging the $u'_i(\boldsymbol{k})$ over translations by half a box in each dimension. There are $2^3 = 8$ such translations, and each translation is characterized by a triple s_x, s_y, s_z giving the number of half boxes translated in each dimension, either zero or one for each component. The effect of the translation is to multiply each coefficient $w_i(\boldsymbol{k})$ by $(-)^{s_xk_x+s_yk_y+s_zk_z}$, that is, by ± 1 according to whether $s_xk_x + s_yk_y + s_zk_z$ is even or odd. The sign change carries through the definitions (49) of $v_i(\boldsymbol{k})$ and (52) of $v'_i(\boldsymbol{k})$ to the definition (58) of $u'_i(\boldsymbol{k})$, and thence to the definition (70) of $u'(\boldsymbol{k})$. That is, the effect of a translation by half a box is to multiply $u'(\boldsymbol{k})$ by $(-)^{s_xk_x+s_yk_y+s_zk_z}$. It

follows that, after averaging over translations, $u'(\boldsymbol{k})$ vanishes if any component of $\boldsymbol{k}$ is odd, leaving only cases where all components of $\boldsymbol{k}$ are even. Consequently, $u'(\boldsymbol{k})$ need be evaluated only at the 125 wavevectors $\boldsymbol{k}$ all of whose components are even. The symmetrized function $u'(\boldsymbol{k})$ can be computed by averaging the values of $u'(\boldsymbol{k})$ at the 125 values of $\boldsymbol{k}$ into the 10 distinct bins with even $0 \leqslant k_x \leqslant k_y \leqslant k_z \leqslant 4$. It is amusing that increasing the number of weightings (by a factor of 8, if all translations yield distinct weightings) actually decreases the computational work required to find the best Fourier coefficients $w_i(\boldsymbol{k})$.

Adjoining translations by a quarter of a box simplifies the problem of finding the minimum variance solution for the coefficients $w_i(\boldsymbol{k})$ even further. There are $4^3 = 64$ such translations, and each translation is characterized by a triple s_x, s_y, s_z, each component running over 0–3, giving the number of quarter boxes translated in each dimension. The effect of the translation is to multiply each coefficient $w_i(\boldsymbol{k})$ by $\mathrm{i}^{s_x k_x + s_y k_y + s_z k_z}$. The effect propagates through to the symmetrized function $u'(\boldsymbol{k})$, which is therefore non-zero only for the 27 wavevectors $\boldsymbol{k}$ all of whose components are multiples of 4. The symmetrized function $u'(\boldsymbol{k})$ can be computed by averaging the values of $u'(\boldsymbol{k})$ at the 27 values of $\boldsymbol{k}$ into the four distinct bins with $0 \leqslant k_x \leqslant k_y \leqslant k_z \leqslant 4$ and each component a multiple of 4.

One more step, adjoining translations by an eighth of a box, reduces the problem of finding the minimum variance solution to a triviality. After adjoining translations by an eighth of a box, the symmetrized function $u'(\boldsymbol{k})$ vanishes except at $\boldsymbol{k} = \boldsymbol{0}$. The function to be minimized, the right-hand side of equation (69), is therefore identically equal to 1, and any arbitrary weighting therefore yields a minimum variance solution. Though amusing, the result is not terribly useful, because it involves a vast number, $48 \times 8^3 = 24\,576$, of weightings. Physically, if there are enough weightings, then together they exhaust the information about the covariance of power, however badly crafted the weightings may be. As will be seen in Section 3.3, there are much simpler solutions that achieve the absolute minimum possible variance, for which the right-hand side of equation (69) equals 1, with far fewer weightings.

The argument above has shown that the problem of finding the minimum variance solution for $w_i(\boldsymbol{k})$ attains its simplest non-trivial form if the weightings are generated from a representative weighting by rotations, reflections and translations by quarter of a box, a total of $48 \times 4^3 = 3\,072$ symmetries. In this case, the weighting-dependent factor in the variance of covariance of power, the right-hand side of equation (69), becomes a rational function, a ratio of two eighth-order polynomials in the 27 Fourier coefficients $w_i(\boldsymbol{k})$, the numerator being a sum $\sum u'(\boldsymbol{k})^2$ of squares of four quartics, and the denominator $u'(\boldsymbol{0})^2$ the square of a quartic. It is this function that we minimize in Section 3.3 to find a best set of weightings.

The minimum variance solution is independent of the overall normalization of the coefficients $w_i(\boldsymbol{k})$, since the quantity being minimized, the ratio on the right-hand side of equation (69), is independent of the normalization of $w_i(\boldsymbol{k})$. Once the minimum variance solution for the coefficients $w_i(\boldsymbol{k})$ has been found, the coefficients can be renormalized to satisfy the normalization condition (19) that ensures that the estimates $\hat{p}_i(k)$ of the shell-averaged power spectra of weighted densities are estimates of the true shell-averaged power $p(k)$, equations (33) and (34).

3.3 Minimum variance weightings

The previous subsection, Section 3.2, described how to obtain the coefficients $w_i(\boldsymbol{k})$ that minimize the expected variance of the estimate of covariance of shell-averaged power that comes from averaging over an ensemble of weightings that contain only combinations of fundamental modes, and that are symmetrically related to each other by rotations, reflections and translations by quarter of a box.

Numerically, we find not one but three separate sets of minimum variance weightings (with hindsight, the sets are simple enough that they might perhaps have been found without resort to numerics). Each set consists of symmetrical transformations of a weighting generated by a single mode, namely $\{1, 0, 0\}$, $\{1, 1, 0\}$ and $\{1, 1, 1\}$, respectively, for each of the three sets. Because each individual weighting has a rather high degree of symmetry, each set has far fewer than the $48 \times 4^3 = 3\,072$ weightings expected if all symmetrical transformations yielded distinct weightings. Each of the three sets is generated by the weighting

$$w_i(\boldsymbol{k}) = \begin{cases} \mathrm{e}^{\pm \mathrm{i}\pi/8}/\sqrt{2} & \text{if } \boldsymbol{k} = \pm\boldsymbol{k}_I, \\ 0 & \text{otherwise,} \end{cases} \tag{71}$$

where $\boldsymbol{k}_i$ is one of the three possibilities

$$\boldsymbol{k}_i = \begin{cases} \{1, 0, 0\} & \text{set one: 12 weightings,} \\ \{1, 1, 0\} & \text{set two: 24 weightings,} \\ \{1, 1, 1\} & \text{set three: 16 weightings.} \end{cases} \tag{72}$$

In real space, the weighting $w_i(\boldsymbol{r})$ corresponding to $w_i(\boldsymbol{k})$ of equation (71) is

$$w_i(\boldsymbol{r}) = \sqrt{2}\cos\left[2\pi\left(\boldsymbol{k}_i \cdot \boldsymbol{r} + \frac{1}{16}\right)\right]. \tag{73}$$

Fig. 1 illustrates two representative examples of the weightings w_i ($\boldsymbol{r}$). The complete set of 12 (24, 16) weightings for each set is obtained as follows. In set one (two, three), a factor of 6 (12, 8) comes from the cubic (dodecahedral, octahedral) symmetry of permuting and reflecting the components k_x, k_y, k_z of $\boldsymbol{k}$, or equivalently the components x, y, z of $\boldsymbol{r}$. A further factor of 2 comes from multiplying $w_i(\pm\boldsymbol{k})$ by $\pm\mathrm{i}$, equivalent to translating by quarter of a box, or $1/16 \to 5/16$ in equation (73).

The three minimum variance solutions are absolute minimum variance, in the sense that each set not only minimizes the expression on the right-hand side of equation (69), but it solves $u'(\boldsymbol{k}) = 0$ for $\boldsymbol{k} \neq \boldsymbol{0}$. This means that it is impossible to find better solutions in which all the weightings are symmetrically related to each other, which is the condition under which equation (69) was derived.

With the minimum variance solutions in hand, it is possible to go back and examine the covariance $\langle \Delta\hat{p}'_i \Delta\hat{p}'_j \rangle$, equation (54), between estimates of power from different weightings i and j, either within the same set, or across two different sets. Estimates of power between two different sets are uncorrelated: the covariance $\langle \Delta\hat{p}'_i \Delta\hat{p}'_j \rangle$ is zero if i and j are drawn from two different sets. If on the other hand the weightings i and j are drawn from the same set, then it turns out that only half of the weightings, the 6 (12, 8) weightings related by the cubic (dodecahedral, octahedral) symmetry of permuting and reflecting k_x, k_y, k_z, yield distinct estimates of deviation in power. The covariance matrix $\langle \Delta\hat{p}'_i \Delta\hat{p}'_j \rangle$ of estimates of power between the 6 (12, 8) cubically (dodecahedrally, octahedrally) related weightings is proportional to the unit matrix. However, translating a weighting by quarter of a box, $w_j(\pm\boldsymbol{k}) = \pm\mathrm{i}w_i(\pm\boldsymbol{k})$, yields an estimate of deviation of power that is minus that of the original weighting, $\Delta\hat{p}'_j = -\Delta\hat{p}'_i$. Actually, this is exactly true only if the slowly varying and thick-shell approximations are exactly true (of course, the thick-shell approximation is never exactly true). Thus translating a weighting by quarter of a box should yield an estimate of deviation in power that is highly anticorrelated with the original; which should provide a useful check of the procedure.

Translating a weighting by half a box simply changes its sign, $w_j(\pm\boldsymbol{k}) = -w_i(\pm\boldsymbol{k})$. This yields an estimate of deviation of power that equals exactly (irrespective of approximations) that of the original weighting, so yields no distinct estimate of deviation in power. These redundant translations by half a box have already been omitted from the set of 12 (24, 16) weightings.

The value of f', the factor that converts, equation (64), estimates $\widehat{X}_i$ of the covariance of power from a weighted density field to an estimate of the true covariance of power is

$$f' = \frac{1}{2}, \tag{74}$$

the same factor for each of the three sets.

The expected covariance matrix $\langle \Delta\widehat{X}_i \Delta\widehat{X}_j \rangle$ of estimates of covariance of power equals g'_{ij} times the true covariance of covariance of power, according to equation (44). The factors g'_{ij}, equation (60), are

$$g'_{ij} = \begin{cases} 3/8 & \text{if } \boldsymbol{k}_i = \boldsymbol{k}_j, \\ 1/8 & \text{if } \boldsymbol{k}_i = -\boldsymbol{k}_j, \\ 1/4 & \text{otherwise.} \end{cases} \tag{75}$$

Equation (75) is valid for weightings i, j both within the same set and across different sets. The case $\boldsymbol{k}_i = \boldsymbol{k}_j$ in equation (75) occurs not only when $i = j$, but also when the weightings i and j are related by translation by quarter of a box. The case $\boldsymbol{k}_i = -\boldsymbol{k}_j$ in equation (75) occurs not only when the weightings i and j are parity conjugates of each other, but also when they are parity confugates translated by quarter of a box.

The factors g''_{ij}, equation (63), which relate the covariance $\langle (\widehat{X}_i - \widehat{X})(\widehat{X}_j - \widehat{X}) \rangle$ of estimates $\widehat{X}_i$ relative to their measured mean $\widehat{X}$, equation (67), as opposed to their expected mean X, equation (66), are

$$g''_{ij} = \begin{cases} 1/8 & \text{if } \boldsymbol{k}_i = \boldsymbol{k}_j, \\ -1/8 & \text{if } \boldsymbol{k}_i = -\boldsymbol{k}_j, \\ 0 & \text{otherwise.} \end{cases} \tag{76}$$

An estimate of the uncertainty in the estimate $\widehat{X}$ can be deduced by measuring the variance $N^{-1}\sum_i (\widehat{X}_i - \widehat{X})^2$ in the fluctuations about the measured mean $\widehat{X}$. There is of course no point in attempting to estimate the uncertainty from $N^{-2}\sum_{ij} \Delta\widehat{X}_i \Delta\widehat{X}_j$, which is identically zero. The true variance $\langle \Delta\widehat{X}^2 \rangle$ can be estimated from the measured variance $N^{-1}\sum_i (\widehat{X}_i - \widehat{X})^2$ by

$$\langle \Delta\widehat{X}^2 \rangle \approx \frac{2}{N} \sum_i (\widehat{X}_i - \widehat{X})^2 \tag{77}$$

in which the factor of 2 comes from (but note the caveat at the end of Section 4.6)

$$\frac{N^{-2}\sum_{ij} g'_{ij}}{N^{-1}\sum_i g''_{ii}} = \frac{1/4}{1/8} = 2 \tag{78}$$

which corrects for the neglected covariance in the measured variance.

3.4 Minimum variance weightings for strategy two

The minimum variance weightings derived above assumed, for definiteness, strategy one, in which the deviation $\Delta\hat{p}'_i(\boldsymbol{k})$ in power is taken to be relative to the power spectrum of the unweighted density, equation (47). An alternative strategy, strategy two, is to take the deviation $\Delta\hat{p}'_i(\boldsymbol{k})$ in power to be relative to the average of the power spectra of the weighted densities, equation (48). Strategy two yields an estimate of covariance of power that has potentially less systematic bias, but potentially greater statistical uncertainty.

As it happens, the minimum variance solution for strategy one, Section 3.3, proves also to solve the minimum variance problem for strategy two. Thus the minimum variance solution weightings are the same for both strategies. Mathematically, expectation values of covariances for the two methods differ in that $v'(\boldsymbol{k})$ is given for strategy one by equation (52), and for strategy two by equation (53). However, for the minimum variance weightings $w_i(\boldsymbol{k})$ of strategy one, equation (71) and its symmetrical transformations, it turns out that $N^{-1}\sum_i v_i(\boldsymbol{k})$, the term subtracted from $v_i(\boldsymbol{k})$ in strategy two, equation (53), is equal to $v_i(\boldsymbol{0})$ if $\boldsymbol{k} = \boldsymbol{0}$, and zero otherwise. This is exactly the same as the term subtracted from $v_i(\boldsymbol{k})$ in strategy one, equation (52). It follows that $v'(\boldsymbol{k})$ is the same for the two strategies.

Although the minimum variance set of weightings is the same for both strategies, the two strategies will in general yield different estimates of the covariance of power.

3.5 More minimum variance weightings

The three minimum variance sets of weightings found (numerically) in Section 3.3 all take the same form, equation (71), differing only in that they are generated by a different single mode, with wavevectors $\{1, 0, 0\}$, $\{1, 1, 0\}$ and $\{1, 1, 1\}$, respectively. One can check that the result generalizes to higher order weightings, in which the wavevector $\boldsymbol{k}_i$ in equation (71) is any wavevector with integral components (such as $\{2, 0, 0\}$, $\{2, 1, 0\}$ and so on). That is, for any wavevector $\boldsymbol{k}_i$ with integral components, the weightings generated from the weighting of equation (71) by rotations, reflections and translations by quarter of a box, form a minimum variance set. All the results of Section 3.3 (and Section 3.4) carry through essentially unchanged. In particular, all equations (73)–(78) remain the same.

The disadvantage of including higher order weightings is that the estimates $\widehat{X}_i$ of the covariance of power become increasingly inaccurate as the wavenumber $|\boldsymbol{k}_i|$ of the weighting increases, because the slowly varying approximations (17) and (37), and the broad-shell approximation (39), become increasingly poor as $|\boldsymbol{k}_i|$ increases.

The advantage of including higher order weightings is that the more weightings, the better the statistical estimate, at least in principle. However, the gain from more weightings is not as great as one might hope. The Cramér–Rao inequality (Kendall & Stuart 1967; see e.g. Hamilton 2005 for a pedagogical derivation) states that the inverse variance of the best possible unbiased estimate $\widehat{X}$ of the parameter X must be less than or equal to the Fisher information F (see Tegmark, Taylor & Heavens 1997) in the parameter X:

$$\langle \Delta\widehat{X}^2 \rangle^{-1} \leqslant F \equiv -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial X^2} \right\rangle, \tag{79}$$

where $\mathcal{L}$ is the likelihood function. To the extent that the estimates $\widehat{X}_i$ are Gaussianly distributed (that is, the likelihood function is a Gaussian in the estimates $\widehat{X}_i$,

$$\mathcal{L} \propto \exp\left[-\frac{1}{2} \sum_{ij} \langle \Delta\widehat{X}_i \Delta\widehat{X}_j \rangle^{-1} (\widehat{X}_i - X)(\widehat{X}_j - X) \right] \tag{80}$$

with covariance $\langle \Delta\widehat{X}_i \Delta\widehat{X}_j \rangle$ independent of X), which could be a rather poor approximation, the Fisher information F in the parameter X approximates the sum of the elements of the inverse covariance matrix,

$$F \approx \sum_{ij} \langle \Delta\widehat{X}_i \Delta\widehat{X}_j \rangle^{-1}. \tag{81}$$

In the present case, the covariance matrix $\langle \Delta\widehat{X}_i \Delta\widehat{X}_j \rangle$ is proportional to g'_{ij}, so in approximation that $\widehat{X}_i$ are Gaussianly distributed, the Fisher information F is proportional to

$$F \propto\propto \sum_{ij} {g'_{ij}}^{-1}. \tag{82}$$

With the coefficients g'_{ij} given by equation (75), the quantity on the right-hand side of equation (82) proves to be a constant, independent of the number of estimates $\widehat{X}_i$,

$$\sum_{ij} {g'_{ij}}^{-1} = 4. \tag{83}$$

This constancy of the Fisher information F with respect to the number of estimates suggests that there is no gain at all in adjoining more and more estimates. However, this conclusion is true only to the extent, first, that the slowly varying and broad-shell approximations are good, and, secondly, that the estimates $\widehat{X}_i$ are Gaussianly distributed, neither of which assumptions necessarily holds. All one can really conclude is that the gain in statistical accuracy from including more estimates is likely to be limited.

There is however another important consideration besides the accuracy of the estimate of the covariance matrix of power: it is desirable that the estimated covariance matrix be, like the true covariance matrix, strictly positive definite, that is, it should have no zero (or negative) eigenvalues. As noted by Pan & Szapudi (2005), if a matrix is estimated as an average over N estimates, then its rank can be no greater than N. Thus, to obtain a positive definite covariance matrix of power for N shells of wavevector, at least N distinct estimates $\widehat{X}_i$ are required.

In Section 3.6, we recommend estimating the covariance of power from an ensemble of $12 + 24 + 16 = 52$ weightings. This will yield a positive definite covariance matrix only if the covariance of power is estimated over no more than 52 shells of wavenumber. Since, as noted in Section 3.3, weightings related by translation by quarter of a box yield highly anticorrelated estimates of power, hence highly correlated estimates of covariance of power, a more conservative approach would be to consider that the 52 weightings yield only 26 effectively distinct estimates of covariance of power, so that the covariance of power can be estimated over no more than 26 shells of wavenumber. If (strategy two) the deviation of power is measured relative to the measured mean over symmetrically related weightings, a (slightly) different mean for each of the three sets of weightings, then 3 degrees of freedom are lost, and the covariance of power can be estimated over no more than $52 - 3 = 49$ shells of wavenumber, or more conservatively over no more than $26 - 3 = 23$ shells of power.

3.6 Recommended strategy

Here is a step-by-step recipe for applying the weightings method to estimate the covariance of power from a periodic simulation.

(i) Select the weightings w_i. We recommend the minimum variance sets of weightings given by equation (71) and its symmetrical transformations. If the weightings are restricted to contain only combinations of fundamental modes, then there are three such sets of weightings, equation (72), and the three sets together provide $N = 12 + 24 + 16 = 52$ distinct weightings.

(ii) For each weighting, measure the shell-averaged power spectrum $\hat{p}_i(k)$ of the weighted density field, equations (32) and (25).

(iii) For each weighting, evaluate the deviation $\Delta\hat{p}_i(k)$ in the shell-averaged power as the difference between $\hat{p}_i(k)$ and either (strategy one) the shell-averaged power $\hat{p}(k)$ of the unweighted density, or (strategy two) the mean $N^{-1}\sum_i \hat{p}_i(k)$ over symmetrically related weightings. The advantage of strategy one is that the statistical error is potentially smaller, whereas the advantage of strategy two is that the systematic bias is potentially smaller. In strategy two, it makes sense to subtract the mean separately for each symmetrically related set of weightings, because the systematic bias is (slightly) different for each set. We recommend trying both strategies one and two, and checking that they yield consistent results.

(iv) Estimate the covariance matrix of shell-averaged power from the average over all N (52) weightings,

$$\left\langle \Delta\hat{p}(k_1)\Delta\hat{p}(k_2) \right\rangle_{\rm est} = \frac{2}{N}\sum_i \Delta\hat{p}_i(k_1)\Delta\hat{p}_i(k_2). \tag{84}$$

The factor of 2 in equation (84) is $1/f' = 2$, equation (74), necessary to convert the average over weightings to an estimate of the true covariance of power, equation (64).

4 BEAT-COUPLING

This paper should have ended at this point. Unfortunately, numerical tests, described in detail in the companion paper (Rimes & Hamilton 2006), revealed a serious problem.

Fig. 2 shows the problem. It shows the median and quartiles of variance of power measured by the weightings method in each of 25 ART ΛCDM simulations of $128\,h^{-1}$ Mpc box size, compared to the variance of power measured over the ensemble of the same 25 simulations. Although the two methods agree at linear scales, the weightings method gives a systematically larger variance at non-linear

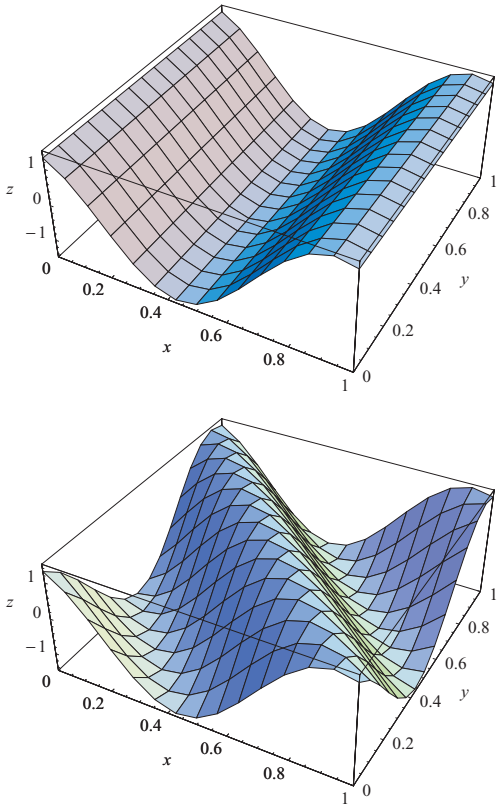

Figure 1. Representative minimum variance weightings $w_i(\boldsymbol{r})$, equation (73), for the cases (top) $\boldsymbol{k}_i = \{1, 0, 0\}$, and (bottom) $\boldsymbol{k}_i = \{1, 1, 0\}$. They are just single Fourier modes, appropriately scaled and phased.

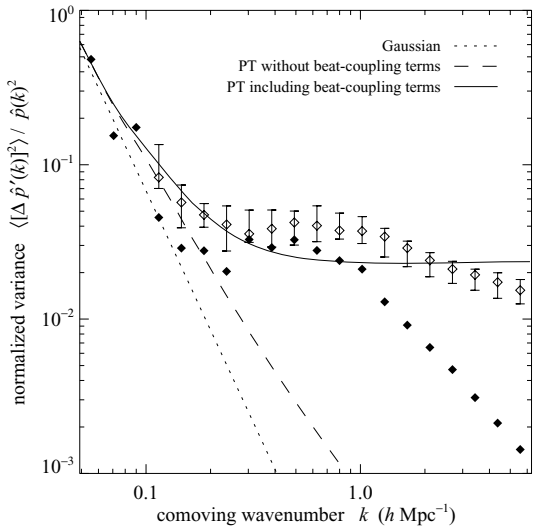

Figure 2. Comparison between the normalized variance of power measured from 25 ART ΛCDM simulations by (symbols with error bars, indicating median and quartiles) the weightings method and (plain symbols) the ensemble method. The two methods disagree substantially at non-linear scales. Lines show the normalized variance predicted by PT both with (solid line) and without (dashed line) the large-scale beat-coupling contribution. The dotted line shows the expected Gaussian contribution to the variance. This figure is a condensed version of fig. 5 of Rimes & Hamilton (2006).

scales. The discrepancy reaches almost an order of magnitude at the smallest scales measured, $k \sim 5\,h^{-1}$ Mpc. The reader is referred to Rimes & Hamilton (2006) for details of the simulations and their results.

This section diagnoses and addresses the problem. The next section, Section 5, discusses the problem and its relevance to observations.

4.1 The cause of the problem: beat-coupling

The physical cause of the problem illustrated in Fig. 2 traces to a non-linear coupling of products of Fourier modes closely spaced in wavenumber to the large-scale beat mode between them. This beat-coupling, as we refer to it, occurs only when power is measured from Fourier modes with a finite spread in wavevector, and therefore appears in the weightings method (and in observations – see Section 5.1) but not in the ensemble method. The beat-coupling is surprisingly large to the point that, as seen in Fig. 2, it actually dominates the variance of power at non-linear scales.

More specifically, in the ensemble method, the power spectrum of a periodic simulation is measured from the variance $\Delta\rho(\boldsymbol{k})\Delta\rho(-\boldsymbol{k})$ of Fourier modes. In the weightings method on the other hand, the power spectrum receives contributions not only from the variance, but also from the covariance $\Delta\rho(\boldsymbol{k})\Delta\rho(-\boldsymbol{k}-\boldsymbol{\varepsilon})$ between modes a small wavevector $\boldsymbol{\varepsilon}$ apart. This covariance vanishes in the mean, but it couples to large-scale modes $\Delta\rho(\boldsymbol{\varepsilon})$ through quadratic non-linearities. That is, the correlation between the product $\Delta\rho(\boldsymbol{k})\,\Delta\rho(-\boldsymbol{k}-\boldsymbol{\varepsilon})$ and the large-scale mode $\Delta\rho(\boldsymbol{\varepsilon})$ is the bispectrum

$$\langle\Delta\rho(\boldsymbol{k})\Delta\rho(-\boldsymbol{k}-\boldsymbol{\varepsilon})\Delta\rho(\boldsymbol{\varepsilon})\rangle = B(\boldsymbol{k}, -\boldsymbol{k}-\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}). \tag{85}$$

The bispectrum is zero for Gaussian fluctuations, but is driven away from zero by non-linear gravitational growth.

4.2 Tetrahedron

The place where, prior to this section, we inadvertently discarded the large-scale beat-coupling is equation (37), where we made the seemingly innocent approximation that the trispectrum $T(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3, \boldsymbol{k}_4)$ is a slowly varying function of what appears to be its arguments, $\boldsymbol{k}_1$–$\boldsymbol{k}_4$. This assumption is false, as we now show.

For a statistically isotropic field (as considered in this paper), the trispectrum depends on six scalar arguments. This follows from the fact that a spatial configuration of four points is determined by the six lengths of the sides of the tetrahedron whose vertices are the four points. In Fourier space, the configuration is an object four of whose sides are equal to the wavevectors $\boldsymbol{k}_1$–$\boldsymbol{k}_4$. The object forms a closed tetrahedron (because $\sum_i \boldsymbol{k}_i = \boldsymbol{0}$), whose shape is determined by the six lengths of the sides of the tetrahedron.

Fig. 3 illustrates the configuration of interest in the present paper, that for the trispectrum in equation (36). Rewritten as a function of six scalar arguments, the trispectrum of equation (36) is

$$\begin{aligned} &T(\boldsymbol{k}_1 - \boldsymbol{k}_1', -\boldsymbol{k}_1 - \boldsymbol{k}_1'', \boldsymbol{k}_2 - \boldsymbol{k}_2', -\boldsymbol{k}_2 - \boldsymbol{k}_2'') \\ &= T(|\boldsymbol{k}_1 - \boldsymbol{k}_1'|, |\boldsymbol{k}_1 + \boldsymbol{k}_1''|, |\boldsymbol{k}_2 - \boldsymbol{k}_2'|, |\boldsymbol{k}_2 + \boldsymbol{k}_2''|, |\boldsymbol{k}_1 - \boldsymbol{k}_2 - \boldsymbol{k}_1' - \boldsymbol{k}_2''|, \varepsilon), \end{aligned} \tag{86}$$

where the wavevector $\boldsymbol{\varepsilon}$ is defined by

$$\boldsymbol{\varepsilon} \equiv -(\boldsymbol{k}_1' + \boldsymbol{k}_1'') = \boldsymbol{k}_2' + \boldsymbol{k}_2'' \tag{87}$$

which is small but not necessarily zero. The invalid approximation (37) is equivalent to approximating

$$\begin{aligned} &T\left(|\boldsymbol{k}_1 - \boldsymbol{k}_1'|, |\boldsymbol{k}_1 + \boldsymbol{k}_1''|, |\boldsymbol{k}_2 - \boldsymbol{k}_2'|, |\boldsymbol{k}_2 + \boldsymbol{k}_2''|, |\boldsymbol{k}_1 - \boldsymbol{k}_2 - \boldsymbol{k}_1' - \boldsymbol{k}_2''|, \varepsilon\right) \\ &\quad \approx T(k_1, k_1, k_2, k_2, |\boldsymbol{k}_1 - \boldsymbol{k}_2|, 0). \end{aligned} \tag{88}$$

The problem with this approximation is apparent. Although primed wavenumbers are small compared to unprimed ones, so that the approximation in the first five arguments is reasonable, in the last argument it is not valid to approximate a finite wavenumber ε, however small, by zero. A valid approximation is, rather,

$$\begin{aligned} &T\left(|\boldsymbol{k}_1 - \boldsymbol{k}_1'|, |\boldsymbol{k}_1 + \boldsymbol{k}_1''|, |\boldsymbol{k}_2 - \boldsymbol{k}_2'|, |\boldsymbol{k}_2 + \boldsymbol{k}_2''|, |\boldsymbol{k}_1 - \boldsymbol{k}_2 - \boldsymbol{k}_1' - \boldsymbol{k}_2''|, \varepsilon\right) \\ &\quad \approx T(k_1, k_1, k_2, k_2, |\boldsymbol{k}_1 - \boldsymbol{k}_2|, \varepsilon). \end{aligned} \tag{89}$$

As an example of the large-scale beat-coupling contributions to the trispectrum that arise from the beat wavevector $\boldsymbol{\varepsilon}$, consider perturbation theory (PT).

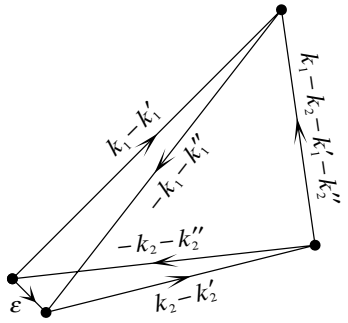

Figure 3. Four-point configuration of wavevectors for the trispectrum in equation (36), which describes the covariance of power spectra of weighted densities. The short leg ε, equation (87), produces a beat-coupling to large scales.

4.3 Perturbation theory

In PT, the trispectrum can be split into snake and star contributions (Scoccimarro et al. 1999; Sefusatti & Scoccimarro 2005):

$$\begin{aligned} T(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3, \boldsymbol{k}_4) \\ &= 4\, P(k_1)P(k_2)P(k_{13})F_2(\boldsymbol{k}_1, -\boldsymbol{k}_{13})F_2(\boldsymbol{k}_2, \boldsymbol{k}_{13}) \\ &\quad + \text{cyclic (12 snake terms)} \\ &\quad + P(k_1)P(k_2)P(k_3)[F_3(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) + \text{perm. (6 terms)}] \\ &\quad + \text{cyclic (4 star terms)}, \end{aligned} \tag{90}$$

where $\boldsymbol{k}_{ij} \equiv \boldsymbol{k}_i + \boldsymbol{k}_j$, and the second-order PT kernel F_2 is given by

$$F_2(\boldsymbol{k}_1, \boldsymbol{k}_2) = \frac{5}{7} + \frac{x}{2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + \frac{2}{7}\, x^2 \tag{91}$$

with $x \equiv \hat{\boldsymbol{k}}_1 \cdot \hat{\boldsymbol{k}}_2$.

In the case of interest, where the trispectrum is that of equation (86), four of the 12 snake terms produce a coupling to large scales, those where the beat wavenumber k_{13} in equation (90) is small. In the (valid) approximation (89), the pertinent PT trispectrum is

$$\begin{aligned} T(|\boldsymbol{k}_1 - \boldsymbol{k}_1'|, |\boldsymbol{k}_1 + \boldsymbol{k}_1''|, |\boldsymbol{k}_2 - \boldsymbol{k}_2'|, |\boldsymbol{k}_2 + \boldsymbol{k}_2''|, |\boldsymbol{k}_1 - \boldsymbol{k}_2 - \boldsymbol{k}_1' - \boldsymbol{k}_2''|, \varepsilon) \\ &\approx T(k_1, k_1, k_2, k_2, |\boldsymbol{k}_1 - \boldsymbol{k}_2|, 0) \\ &\quad + 16\, P(k_1)P(k_2)P(\varepsilon)F_2(\boldsymbol{k}_1, -\boldsymbol{\varepsilon})F_2(\boldsymbol{k}_2, \boldsymbol{\varepsilon}) \end{aligned} \tag{92}$$

in which the term on the last line represents the large-scale beat-coupling contribution incorrectly ignored by the approximation (88). In equation (36) for the covariance of shell-averaged power, this trispectrum, equation (92), is angle averaged over the directions of $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$. The angle-averaged second-order PT kernel is

$$\int F_2(\boldsymbol{k}, \boldsymbol{\varepsilon})\, \frac{do_{\boldsymbol{k}}}{4\pi} = \frac{17}{21} \tag{93}$$

and it follows that the last line of equation (92), when angle-averaged is $16(17/21)^2 P(k_1)P(k_2)\, P(\varepsilon)$.

Following the same arguments that led from equations (36) to (40), and then to equation (54), but with the beat-coupling term now correctly retained in the trispectrum, one finds that equation (54) for the expected covariance of shell-averaged power spectra of weighted densities is modified to

$$\begin{aligned} \langle \Delta\hat{p}_i'(k_1)\Delta\hat{p}_j'(k_2)\rangle &\approx f_{ij}' \langle \Delta\hat{p}(k_1)\Delta\hat{p}(k_2)\rangle \\ &\quad + 4\, R_a P(k_1)P(k_2) \sum_{\boldsymbol{k}} v_i'(\boldsymbol{k}) v_j'(-\boldsymbol{k}) P(k), \end{aligned} \tag{94}$$

where $v_i'(\boldsymbol{k})$ is defined by equations (49) and (52) or (53), and the constant R_a is

$$R_a = 4\left(\frac{17}{21}\right)^2 \approx 2.62. \tag{95}$$

The reason for writing equation (94) in this form, with the constant R_a separated out, is that, as will be seen in Section 4.4, the same expression remains valid in the hierarchical model, but with R_a the four-point hierarchical snake amplitude.

Fig. 2 includes lines showing the predicted PT result for the variance of shell-averaged power of weighted density, equation (36), both with (solid lines) and without (dashed lines) beat-coupling. The PT variance with beat-coupling was obtained by numerically integrating the PT expression (90) for the trispectrum (86) in equation (36) (that is, without making the approximation 89 or 94), with the minimum variance weightings (71), and then multiplying by the factor $1/f' = 2$, equation (74). From this the PT variance without beat-coupling was obtained by setting $P(\varepsilon) = 0$. The variance without beat-coupling agreed well with a direct PT evaluation of equation (24).

Fig. 2 shows that the beat-coupling contribution predicted by PT seems to account reasonably well for the extra variance that appears at non-linear scales in the weightings versus the ensemble method.

We will return to equation (94) in Section 4.5, but first consider the hierarchical model as a prototype of the trispectrum beyond PT.

4.4 Hierarchical model

PT is valid only in the translinear regime. The behaviour of the trispectrum in the fully non-linear regime is less well understood. Available observational and N-body evidence (Colombi, Bouchet & Hernquist 1996; Hui & Gaztañaga 1999; Scoccimarro & Frieman 1999; Baugh et al. 2004; Croton et al. 2004) is consistent with a hierarchical model of higher order correlations. In the hierarchical model (Peebles 1980), the trispectrum is a sum of snake and star terms:

$$\begin{aligned} T(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3, \boldsymbol{k}_4) \\ &= R_a[P(k_1)P(k_2)P(k_{13}) + \text{cyclic (12 snake terms)}] \\ &\quad + R_b[P(k_1)P(k_2)P(k_3) + \text{cyclic (4 star terms)}]. \end{aligned} \tag{96}$$

The PT trispectrum, equation (90), shows a hierarchical structure with hierarchical amplitudes R_a and R_b that are not constant, but rather depend on the shape of the trispectrum tetrahedron. At highly non-linear scales, Scoccimarro & Frieman (1999) suggested an ansatz, dubbed hyperextended perturbation theory (HEPT), that the hierarchical amplitudes go over to the values predicted by PT for configurations collinear in Fourier space. For power-law power spectra $P(k) \propto k^n$, HEPT predicts four-point amplitudes

$$R_a = R_b = \frac{54 - 27\, 2^n + 2\, 3^n + 6^n}{2\,(1 + 6\, 2^n + 3\, 3^n + 6\, 6^n)}. \tag{97}$$

As pointed out by Scoccimarro & Frieman (1999) and Hamilton (2000), HEPT is not entirely consistent because it predicts a covariance of power $\langle \Delta\hat{p}(k_1)\Delta\hat{p}(k_2)\rangle$ that violates the Schwarz inequality when $k_1 \gg k_2$.

In the hierarchical model with constant hierarchical amplitudes, four of the 12 snake terms produce a coupling to large scales in the trispectrum of interest, equation (86). In the (valid) approximation (89), the hierarchical trispectrum is

$$\begin{aligned} T\left(|\boldsymbol{k}_1 - \boldsymbol{k}_1'|, |\boldsymbol{k}_1 + \boldsymbol{k}_1''|, |\boldsymbol{k}_2 - \boldsymbol{k}_2'|, |\boldsymbol{k}_2 + \boldsymbol{k}_2''|, |\boldsymbol{k}_1 - \boldsymbol{k}_2 - \boldsymbol{k}_1' - \boldsymbol{k}_2''|, \varepsilon\right) \\ &\approx T(k_1, k_1, k_2, k_2, |\boldsymbol{k}_1 - \boldsymbol{k}_2|, 0) \\ &\quad + 4\, R_a\, P(k_1)P(k_2)P(\varepsilon) \end{aligned} \tag{98}$$

in which the term on the last line represents the large-scale beat-coupling contribution.

The hierarchical trispectrum (98) looks similar to (slightly simpler than) the PT trispectrum (92). Following the same arguments as before, one recovers the same expression (94) for the expected covariance of shell-averaged power spectra of weighted densities.

4.5 Covariance of shell-averaged power spectra including large-scale coupling

Suppose that either the PT, Section 4.3, or the hierarchical model, Section 4.4, offers a reliable guide to the coupling of the non-linear

trispectrum to large scales, so that equation (94) is a good approximation to the expected covariance of shell-averaged power spectra of weighted densities.

Make the further assumption that the power spectrum is approximately constant over the large-scale wavevectors represented in $\dot{w}(\boldsymbol{k})$:

$$P(k) \approx P(2k_b) = \text{constant} \qquad \text{for } v_i'(\boldsymbol{k}) \neq 0, \tag{99}$$

where k_b is the wavenumber at the box scale. The factor of 2 in $2\ k_b$ in equation (99) appears as a reminder that the wavevectors $\boldsymbol{k}$ in $\dot{w}(\boldsymbol{k})$ are, equations (49) and (52) or (53), sums of pairs of wavenumbers $\boldsymbol{k}'$ represented in the weighting $w_i(\boldsymbol{k}')$. For example, if the weightings are taken to be the minimum variance weightings given by equation (71), then $k_b = k_i$, where k_i is the wavenumber of the weighting. Approximation (99) is in the same spirit as, but distinct from the earlier approximation (17) that the power spectrum is a slowly varying function. Note that equation (99) does *not* require that $P(0) \approx P(2\ k_b)$ (which would certainly not be correct, because $P(0) = 0$), because $\dot{w}(\boldsymbol{0})$ is zero, which is true a priori in strategy one, equation (52), and ends up being true a posteriori in strategy two, equation (53), by the argument in Section 3.4.

In the approximation (99), the summed expression on the right-hand side of equation (94) is

$$\sum_{\boldsymbol{k}} v_i'(\boldsymbol{k}) v_j'(-\boldsymbol{k}) P(k) \approx f_{ij}' P(2k_b), \tag{100}$$

and equation (94) reduces to

$$\left\langle \Delta\hat{p}_i'(k_1) \Delta\hat{p}_j'(k_2) \right\rangle \approx f_{ij}' [\langle \Delta\hat{p}(k_1) \Delta\hat{p}(k_2) \rangle + 4\, R_a P(k_1) P(k_2) P(2k_b)] \tag{101}$$

with the term on the last line being the large-scale beat-coupling contribution.

Equation (101) provides the fundamental justification for the weightings method when beat-coupling is taken into account. It states that the covariance of shell-averaged power spectra of weighted densities is proportional to the sum of the true covariance $\langle \Delta\hat{p}(k_1) \Delta\hat{p}(k_2) \rangle$ of shell-averaged power, and a beat-coupling term $4R_a P(k_1) P(k_2) P(2k_b)$ proportional to power at (twice) the box wavenumber k_b. The crucial feature of equation (101) is that the constant of proportionality f_{ij}', equation (55), depends only on the weightings $w_i(\boldsymbol{k})$, and is independent either of the power spectrum $P(k)$ or of the wavenumbers k_1 and k_2.

In the limit of infinite box size, the beat-coupling contribution to the covariance of power spectra of weighted densities in equation (101) goes to zero, $P(2k_b) \to P(0) = 0$ as $k_b \to 0$, and the covariance becomes proportional to the true covariance $\langle \Delta\hat{p}(k_1) \Delta\hat{p}(k_2) \rangle$ of power. However, in cosmologically realistic simulations, such as illustrated in Fig. 2 and discussed further in Section 5, the beat-coupling contribution, far from being small, is liable to dominate at non-linear scales.

Beyond PT or the hierarchical model, the weightings method remains applicable just so long as the hierarchical amplitude R_a in equation (101) is independent of the weightings ij. In general, R_a could be any arbitrary function of k_1, k_2 and the box wavenumber k_b.

4.6 Not quite minimum variance weightings

Section 3 derived sets of minimum variance weightings valid when the covariance, and the covariance of covariance of power spectra of weighted densities took the separable forms given by equations (54) and (59). When beat-scale coupling is included, the covariance of power, equation (101), still takes the desired separable form (as long as the hierarchical amplitude R_a is independent of the weightings ij), but the covariance of covariance of power (equation A3 of Appendix A) does not.

In Appendix A, we discuss what happens to the minimum variance derivation of Section 3 when beat-coupling is included. We argue that the minimum variance weightings of Section 3.3 are no longer exactly minimum variance, but probably remain near minimum variance, and therefore fine to use in practice.

The factor of 2 on the right-hand side of equation (77) is no longer correct when beat-coupling is included, but may remain a reasonable approximation.

5 DISCUSSION

As shown in Section 4, the covariance of non-linear power receives beat-coupling contributions from large scales whenever power is measured from Fourier modes $\rho(\boldsymbol{k})$ that have a finite spread in wavevector $\boldsymbol{k}$, as opposed to being delta functions at single discrete wavevectors. Physically, the large-scale beat-coupling arises because a product $\Delta\rho(\boldsymbol{k}) \Delta\rho(-\boldsymbol{k} - \boldsymbol{\varepsilon})$ of Fourier amplitudes of closely spaced wavevectors couples by non-linear gravitational growth to the beat mode $\Delta\rho(\boldsymbol{\varepsilon})$ between them.

The beat-coupling contribution does not appear when covariance of power is measured from ensembles of periodic box simulations, because in that case power is measured from products of Fourier amplitudes $\Delta\rho(\boldsymbol{k}) \Delta\rho(-\boldsymbol{k})$ at single discrete wavevectors. Here the 'beat' mode is the mean mode, $\boldsymbol{k} - \boldsymbol{k} = \boldsymbol{0}$, whose fluctuation is by definition always zero, $\Delta\rho(\boldsymbol{0}) = 0$.

There is on the other hand a beat-coupling contribution when covariance of power is measured by the weightings method, because the Fourier modes of weighted density are spread over more than one wavevector.

For weightings constructed from combinations of fundamental modes, as recommended in Section 3, the covariance of power spectra of weighted densities receives beat-coupling contributions from power near the box fundamental k_b. The beat-coupling and normal contributions to the variance $\langle \Delta\hat{p}_i'(k)^2 \rangle$ of non-linear power are in roughly the ratio $P(2k_b)/P(k)$ of power at the box scale to power at the non-linear scale, according to equation (101).

In cosmologically realistic simulations, box sizes are typically around the range 10^2–$10^3\ h^{-1}$ Mpc. This is just the scale at which the power spectrum goes through a broad maximum. For example, in observationally concordant ΛCDM models, power goes through a broad maximum at $k_{\rm peak} \approx 0.016\, h\,{\rm Mpc}^{-1}$ (e.g. Tegmark et al. 2004a), corresponding to a box size $4\pi/k_{\rm peak} \sim 800\, h^{-1}$ Mpc. Power at the maximum is about 25 times greater than power at the onset $k_{\rm trans} \approx 0.3\, h\,{\rm Mpc}^{-1}$ of the translinear regime, $P(k_{\rm peak})/P(k_{\rm trans}) \approx 25$, and the ratio $P(k_{\rm peak})/P(k)$ of power increases at more non-linear wavenumbers k.

It follows that in cosmologically realistic simulations the beat-coupling contribution to the covariance of power is liable to dominate the normal contribution. This is consistent with the numerical results illustrated in Fig. 2 and discussed by Rimes & Hamilton (2006), which show that the variance of power measured by the weightings method (which includes beat-coupling contributions) substantially exceeds, at non-linear scales, the variance of power measured by the ensemble method (which does not include beat-coupling contributions).

5.1 Relevance to real galaxy surveys

In real galaxy surveys, measured Fourier modes inevitably have finite width $|\Delta \boldsymbol{k}| \sim 1/R$, where R is a characteristic linear size of the survey. The characteristic size R varies from 100 to a few $1000\,h^{-1}$ Mpc (an upper limit is set by the comoving horizon distance, which is about $10^4\,h^{-1}$ Mpc in the concordant ΛCDM model).

It follows that the covariance of non-linear power measured in real galaxy surveys is liable to be dominated not by the 'true' covariance of power (the covariance of power in a perfect, infinite survey), but rather by the contribution from beat-coupling to power at the scale of the survey.

This means that one must take great care in using numerical simulations to estimate or to predict the covariance of non-linear power expected in a galaxy survey. The scatter in power over an ensemble of periodic box simulations will certainly underestimate the covariance of power by a substantial factor at non-linear scales, because of the neglect of beat-coupling contributions.

A common and in principle reliable procedure is to estimate the covariance of power of a galaxy survey from mock surveys 'observed' with the same selection rules as the real survey from numerical simulations large enough to encompass the entire survey (e.g. Coil, Davis & Szapudi 2001; Padilla & Baugh 2003; Tegmark et al. 2004a; Yan, White & Coil 2004; Blaizot et al. 2005; Cole et al. 2005; Eisenstein et al. 2005; Frith, Shanks & Outram 2005; Park et al. 2005).

It is important that numerical simulations be genuinely large enough to contain a mock survey. One should be wary about estimating covariance of power from mock surveys extracted from small periodic boxes replicated many times (e.g. Yang et al. 2004), since such boxes are liable to be missing power at precisely those wavenumbers, the inverse scale size of the mock survey, where beat-coupling should in reality be strongest. Beat-coupling arises from a real gravitational coupling to large-scale modes, and the simulation from which a mock survey is extracted must be large enough to contain such modes.

Further, it would be wrong to take, say, a volume-limited subsample of a galaxy survey, and then to estimate the covariance of power from an ensemble of periodic numerical simulations whose size is that of the volume-limited subsample. A volume-limited subsample of observational data retains beat-coupling contributions to the covariance of power, whereas periodic box simulations do not.

6 SUMMARY

This paper falls into two parts. In the first part, Sections 2 and 3, we proposed a new method, the weightings method, that yields an estimate of the covariance of the power spectrum of a statistically homogeneous and isotropic density field from a single periodic box simulation. The procedure is to apply a set of weightings to the density field, and to measure the covariance of power from the scatter in power over the ensemble of weightings. In Section 2, we developed the formal mathematical apparatus that justifies the weightings method, and in Section 3, we derived sets of weightings that achieve minimum variances estimates of covariance of power. Section 3.6 gives a step-by-step recipe for applying the weightings method. We recommend a specific set of 52 minimum variance weightings containing only combinations of fundamental modes.

In the second part of this paper, Sections 4 and 5, we discuss an unexpected glitch in the procedure that emerged from the periodic box numerical simulations described in the companion paper (Rimes & Hamilton 2006). The numerical simulations showed that, at non-linear scales, the covariance of power measured by the weightings method substantially exceeded that measured over an ensemble of independent simulations.

In Section 4, we argue from PT that the discrepancy between the weightings and ensemble methods arises from 'beat-coupling', in which products of closely spaced Fourier modes couple by non-linear gravitational growth to the large-scale beat mode between them. Beat-coupling is present whenever non-linear power is measured from Fourier modes that have a finite spread in wavevector, as opposed to being delta functions at single discrete wavevectors. Beat-coupling affects the weightings method, because Fourier modes of weighted densities have a finite width, but not the ensemble method, because the Fourier modes of a periodic box are delta functions of wavevector.

As discussed in Section 5, beat-coupling inevitably affects real galaxy surveys, whose Fourier modes necessarily have a finite width of the order of the inverse scale size of the survey. Surprisingly, at non-linear scales, beat-coupling is liable to dominate the covariance of power of a real survey. One would have thought that the covariance of power at non-linear scales would be dominated by structure at small scales, but this is not true. Rather, the covariance of non-linear power is liable to be dominated by beat-coupling to power at the largest scales of the survey.

A common and valid procedure for estimating the covariance of power from a real survey is the mock survey method, in which artificial surveys are 'observed' from large numerical simulations, with the same selection rules as the real survey. It is important that mock surveys be extracted from genuinely large simulations, not from many small periodic simulations stacked together, since stacked simulations miss the large-scale power essential to beat-coupling.

Finally, it should be remarked that, although this paper has considered only the covariance of the power spectrum, it is likely that, in real galaxy surveys and cosmologically realistic simulations, beat-coupling contributions dominate the non-linear variance and covariance of most other statistical measures, including higher order n-point spectra such as the bispectrum and trispectrum, and n-point correlation functions in real space, including the two-point correlation function.

ACKNOWLEDGMENTS

We thank Nick Gnedin, Matias Zaldarriaga and Max Tegmark for helpful conversations, and Anatoly Klypin and Andrey Kravtsov for making the MPI implementation of ART available to us and for help with its application. GRAFIC is part of the COSMICS package, which was developed by Edmund Bertschinger under NSF grant AST-9318185. The simulations used in this work were performed at the San Diego Supercomputer Centre using resources provided by the National Partnership for Advanced Computational Infrastructure under NSF cooperative agreement ACI-9619020. This work was supported by NSF grant AST-0205981 and by NASA ATP award NAG5-10763.

REFERENCES

Baugh C. M. et al. (29 authors; 2dFGRS team), 2004, MNRAS, 351, L44
Blaizot J., Wadadekar Y., Guiderdoni B., Colombi S., Bertin E., Bouchet F. R., Devriendt J. E. G., Hatton S., 2005, MNRAS, 360, 159
Coil A. L., Davis M., Szapudi I., 2001, PASP, 113, 1312
Cole S. et al. (31 authors; 2dFGRS team), 2005, MNRAS, 362, 505
Colombi S., Bouchet F. R., Hernquist L., 1996, ApJ, 465, 14

Cooray A., 2004, MNRAS, 348, 250
Croton D. J. et al. (29 authors; 2dFGRS team), 2004, MNRAS, 352, 1232
Eisenstein D. J. et al. (48 authors; SDSS collaboration), 2005, ApJ, 633, 560
Frith W. J., Shanks T., Outram P. J., 2005, MNRAS, 361, 701
Hamilton A. J. S., 2000, MNRAS, 312, 257
Hamilton A. J. S., 2006, in Martinez V., ed., Proc. Int. Summer School, Data Analysis in Cosmology. Lecture Notes in Physics. Springer-Verlag, Berlin, in press (astro-ph/050360)
Hoekstra H., Yee Y., Gladders M., 2002, ApJ, 577, 595
Hui L., Gaztañaga E., 1999, ApJ, 519, 622
Kendall M. G., Stuart A., 1967, The Advanced Theory of Statistics. Hafner Publishing, New York
Komatsu E. et al. (15 authors; WMAP team), 2003, ApJS, 148, 119
Künsch H. R., 1989, Ann. Stat., 17, 1217
Lidz A., Heitmann K., Hui L., Habib S., Rauch M., Sargent W. L. W., 2006, ApJ, 638, 27
Meiksin A., White M., 1999, MNRAS, 308, 1179
Padilla N. D., Baugh C. M., 2003, MNRAS, 343, 796
Pan J., Szapudi I., 2005, MNRAS, 362, 1363
Park C. et al. (SDSS collaboration), 2005, ApJ, 633, 11
Peebles P. J. E., 1980, The Large Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
Pen U.-L., Lu T., van Waerbeke L., Mellier Y., 2003, MNRAS, 346, 994
Rimes C. D., Hamilton A. J. S., 2005, MNRAS, 360, L82
Rimes C. D., Hamilton A. J. S. 2006, MNRAS, in press (doi: 10.1111/j.1365-2966.2006.10710.x)
Sanchez A. G., Baugh C. M., Percival W. J., Peacock J. A., Padilla N. D., Cole S., Frenk C. S., Norberg P., 2006, MNRAS, 366, 189
Scoccimarro R., Frieman J. A., 1999, ApJ, 520, 35
Scoccimarro R., Zaldarriaga M., Hui L., 1999, ApJ, 527, 1
Sefusatti E., Scoccimarro R., 2005, Phys. Rev. D, 71, 063 001
Seljak U. et al. (21 authors; SDSS team), 2005, Phys. Rev. D, 71, 103 515
Sheldon E. S. et al., 2004, AJ, 127, 2544
Spergel D. N. et al. (17 authors; WMAP team), 2003, ApJS, 148, 175
Takada M., Jain B., 2004, MNRAS, 348, 897
Tegmark M., Taylor A., Heavens A., 1997, ApJ, 480, 22
Tegmark M. et al. (62 authors; SDSS collaboration), 2004a, ApJ, 606, 702
Tegmark M. et al. (64 authors; SDSS collaboration), 2004b, Phys. Rev. D, 69, 103 501
Verde L., Heavens A. F., 2001, ApJ, 553, 14
Viel M., Haehnelt M. G., 2006, MNRAS, 365, 231
Yan R., White M., Coil A. L., 2004, ApJ, 607, 739
Yang X., Mo H. J., Jing Y. P., van den Bosch F. C., Chu Y., 2004, MNRAS, 350, 1153
Zhan H., Eisenstein D., 2005, MNRAS, 357, 1387
Zhan H., Davé R., Eisenstein D., Katz N., 2005, MNRAS, 363, 1145

APPENDIX A: MINIMUM VARIANCE WEIGHTINGS AND BEAT-COUPLING

This appendix describes how the beat-coupling contributions to covariance of power discussed in Section 4 modify the minimum variance arguments in Section 3. The conclusion is that the minimum variance weightings given in Section 3.3 are no longer exact minimum variance, but are probably near minimum variance, and therefore fine to use in practice.

With no approximations at all, the covariance of covariance of shell-averaged power spectra of weighted densities (with the deviations $\Delta\hat{p}'_i(k)$ in power being taken relative to the measured rather than the expected mean power, equations 47 or 48), takes the generic form

$$\begin{aligned}&\left\langle \left[\Delta\hat{p}'_i(k_1)\Delta\hat{p}'_i(k_2) - \left\langle \Delta\hat{p}'_i(k_1)\Delta\hat{p}'_i(k_2)\right\rangle\right]\right.\\&\quad\left.\times\left[\Delta\hat{p}'_j(k_3)\Delta\hat{p}'_j(k_4) - \left\langle \Delta\hat{p}'_j(k_3)\Delta\hat{p}'_j(k_4)\right\rangle\right]\right\rangle\\&\quad= \sum_{\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4=\mathbf{0}} v'_i(\varepsilon_1)v'_i(\varepsilon_2)v'_j(\varepsilon_3)v'_j(\varepsilon_4)\\&\quad\times \sum_{\boldsymbol{k}_1\in V_{k_1},\ldots,\ \boldsymbol{k}_4\in V_{k_4}} E(\boldsymbol{k}_1-\boldsymbol{k}'_1, -\boldsymbol{k}_1-\boldsymbol{k}''_1, \ldots, \boldsymbol{k}_4-\boldsymbol{k}'_4, -\boldsymbol{k}_4-\boldsymbol{k}''_4),\end{aligned} \tag{A1}$$

where $v'(\boldsymbol{k})$ is defined by equations (49) and (52) or (53), the wavevectors ε_n are defined by

$$\varepsilon_n \equiv -\boldsymbol{k}'_n - \boldsymbol{k}''_n \qquad \text{for } n = 1, \ldots, 4, \tag{A2}$$

and E is an eight-point object, a sum of products of n-point functions adding up to eight points, as enumerated in equation (43).

Fig. A1 illustrates the configuration of the eight-point function that contributes to the eight-point object E in equation (A1). The short legs $\varepsilon_1, \ldots, \varepsilon_4$ of the configuration constitute a tetrahedron, whose six sides generate beat-couplings to large scale. None of the legs ε_n is zero, because a zero leg would make zero contribution to equation (A1), since $v'(\mathbf{0}) = 0$, equation (52) or, a posteriori, equation (53); but it is possible for the sum of a pair of the short legs to be zero. It is these configurations, where the sum of a pair of short legs is zero, that prevent equation (A1) from being separated, as in equation (59), into a product of a factor g'_{ij} that depends only on the weightings and a factor that is independent of the weightings. Note that it is only the eight-point function itself, not lower-order functions, that prevent separability: lower-order functions depend on at most three of the four ε_n, and a triangle with three non-zero sides has (of course) no zero sides.

In either PT or the hierarchical model, and in the various valid approximations made in this paper (first, that n-point spectra are slowly varying functions of their arguments *except* that small arguments ε are *not* replaced by zero, and secondly, that shells are broad), the covariance of covariance of shell-averaged power spectra of weighted densities, equation (A1), reduces to

$$\begin{aligned}&\left\langle \left[\Delta\hat{p}'_i(k_1)\Delta\hat{p}'_i(k_2) - \left\langle \Delta\hat{p}'_i(k_1)\Delta\hat{p}'_i(k_2)\right\rangle\right]\right.\\&\quad\left.\times\left[\Delta\hat{p}'_j(k_3)\Delta\hat{p}'_j(k_4) - \left\langle \Delta\hat{p}'_j(k_3)\Delta\hat{p}'_j(k_4)\right\rangle\right]\right\rangle\\&\quad\approx \lambda g'_{ij} - \mu f'_{ii} f'_{jj} - \nu f'^2_{ij},\end{aligned} \tag{A3}$$

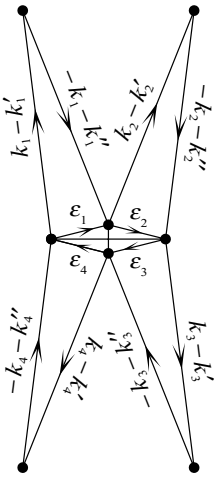

Figure A1. Eight-point configuration of wavevectors contributing to the covariance of covariance of power spectra of weighted densities, equation (A1). The central tetrahedron of short legs produces beat-couplings to large scales.

where f'_{ij} and g'_{ij} are given by equations (55) and (60). The quantities λ, μ, ν in equation (A3) are each functions of k_1, k_2, k_3, k_4 and the box wavenumber k_b, but, importantly, are independent of the weightings ij. The $f'_{ii}\,f'_{jj}$ (respectively f'^2_{ij}) term in equation (A3) arises from terms where $\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 = \mathbf{0}$ (respectively $\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_3 = \mathbf{0}$ or $\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_4 = \mathbf{0}$) in equation (A1). All three terms on the right-hand side of equation (A3) contain large-scale beat-coupling contributions, proportional to one, two or three factors of large-scale power. The second (μ) and third (ν) terms in equation (A1) are written with negative signs because their effect is such as to cancel some of the beat-coupling terms appearing in the first (λ) term (that is, some of the beat-coupling terms proportional to $P(2\,k_b)$ in λ should really be proportional to $P(0) = 0$; the μ and ν terms remove these terms). It is to be expected that λ, $\lambda - \mu$ and $\lambda - \nu$ are all positive. At linear scales, where fluctuations are Gaussian, the beat-couplings generated by non-linear evolution are small, so that $\lambda \gg \mu, \nu$. At non-linear scales, however, μ and ν could be an appreciable fraction of λ.

The derivation of minimum variance weightings in Section 3 involved summing over weightings ij, equation (68). Consider the corresponding double sum over ij of equation (A3). The sum over g'_{ij} yields the same result as before, equation (69). Adjoining the sum over $f'_{ii}\,f'_{jj}$ from equation (A3) modifies equation (69) to

$$\frac{1}{(f'N)^2}\sum_{ij}[g'_{ij} - (\mu/\lambda)f'_{ii}f'_{jj}] = \frac{1}{u'(\mathbf{0})^2}\left[\left(1-\frac{\mu}{\lambda}\right)u'(\mathbf{0})^2 + \sum_{\boldsymbol{k}\neq\mathbf{0}}|u'(\boldsymbol{k})|^2\right], \tag{A4}$$

because $f'_{ii} = f' = u'(\mathbf{0})$ for all i. The minimum variance weightings given in Section 3.3 were absolute minimum variance in the sense that $u'(\boldsymbol{k}) = 0$ for $\boldsymbol{k} \neq \mathbf{0}$. The same minimum variance weightings continue to achieve absolute minimum variance for equation (A3), reducing its right-hand side to the irreducible minimum $1 - \mu/\lambda$.

Thus the minimum variance weightings of Section 3.3 remain minimum variance as long as only the first two terms (λ and μ) of equation (A3) are considered. The third (ν) term breaks the minimum variance derivation. However, this third term is likely to be subdominant compared to the first two. The quantity f'_{ij} in the third term of equation (A3) is proportional to the covariance of power between weightings i and j, equation (101), and the Schwarz inequality guarantees that

$$f'^2_{ij} \leqslant f'_{ii}f'_{jj} \tag{A5}$$

so that there is a natural tendency for the third term of equation (A3) to be dominated by the second. The only way for the third term to be large is for the power spectra from different weightings ij to be highly correlated with each other. Physically, however, the most accurate estimate of covariance of power should come from averaging over many uncorrelated weightings, in which case $f'^2_{ij} \ll f'_{ii}f'_{jj}$ for most weightings $i \neq j$. Thus, as just stated, it is to be expected that the third term should be subdominant compared to the first two.

In summary, to the extent that either PT or the hierarchical model provide a reliable guide to the behaviour of high-order correlations, and to the extent that the third term of equation (A3) is subdominant, as it should be, the minimum variance weightings of Section 3.3 should remain near minimum variance, good enough for practical application.

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