MEASURING THE GALAXY POWER SPECTRUM AND SCALE-SCALE CORRELATIONS WITH MULTIORESOLUTION-DECOMPOSED COVARIANCE. I. METHOD

LI-ZHI FANG
Department of Physics, University of Arizona, Tucson, AZ 85721; fanglz@physics.arizona.edu

AND

LONG-LONG FENG
Center for Astrophysics, University of Science and Technology of China, Hefei, Anhui 230026, Peoples Republic of China; and National Astronomical Observatories, Chinese Academy of Science, Chao-Yang District, Beijing, 100012; fengll@physics.arizona.edu

Received 1999 November 10; accepted 2000 March 20

ABSTRACT

We present a method for measuring the galaxy power spectrum based on multiresolution analysis of the discrete wavelet transformation (DWT). Apart from the technical advantages of the computational feasibility for data sets with a large volume and complex geometry, the DWT scale-by-scale decomposition provides a physical insight into the covariance matrix of the cosmic mass field. Since the DWT representation has a strong capability for suppressing the off-diagonal components of the covariance for self-similar clustering, the DWT covariance for all popular models of the cold dark matter cosmogony is generally diagonal, or j (scale) diagonal in the scale range in which the second or higher order scale-scale correlations are weak. In this range, the DWT covariance gives a lossless estimation of the power spectrum, which is equal to the corresponding Fourier power spectrum banded with a logarithmical scaling. This DWT estimator is optimized in the sense that the spatial resolution is automatically adaptive to the perturbation wavelength to be studied. In the scale range in which the scale-scale correlation is significant, the accuracy of a power spectrum detection depends on the scale-scale or band-band correlations. In this case, for a precision measurements of the power spectrum, or a precision comparison of the observed power spectrum with models, a measurement of the scale-scale or band-band correlations is needed. We show that the DWT covariance can be employed to measure both the band–power spectrum and second-order scale-scale correlation. We also present the DWT algorithm of the binning and Poisson sampling with real observational data. We show that the so-called alias effect appeared in usual binning schemes can exactly be eliminated by the DWT binning. Since the Poisson process possesses diagonal covariance in the DWT representation, the Poisson sampling and selection effects on the power spectrum and second order scale-scale correlation detection are suppressed into a minimum. Moreover, the effect of the non-Gaussian features of the Poisson sampling can also be calculated in this frame. The DWT method is open, i.e., one can add further DWT algorithms on the basic decomposition in order to estimate other effects on the power spectrum detection, such as non-Gaussian correlations and bias models.

Subject headings: cosmology: theory — large-scale structure of universe

1. INTRODUCTION

Measuring the galaxy power spectrum has been and continues to be a central subject of the study of large-scale structure. Although the power spectrum is only a second-order statistical measure of the deviations of a random field, \( \delta(x) \), of mass density from homogeneity, it directly reflects the physical scales of the processes that affect structure formation. Mathematically, the positive-definiteness of the power spectrum is useful for constraining the parameter space in comparing predictions with data. Since the ongoing and upcoming redshift surveys of galaxies will provide data on galaxy distributions with highly improved quality and greater quantities, this invites us to develop methods for measuring the power spectrum that are more precise and computationally efficient.

Different methods for the power-spectrum measurements adopt different representations or decompositions of the covariance, \( \text{Cov} = \langle \delta(x) \delta(x') \rangle \), where \( \langle \ldots \rangle \) stands for an ensemble average. For a representation given by a set of basis functions \( \psi_i(x) \) (sometimes referred to as the weight function), the random field is described by the variables

\[
X_i = \int \delta(x) \psi_i(x) \, dx ,
\]

and the covariance is given by \( \text{Cov}_{ij} = \langle X_i X_j \rangle \). If the covariance in this representation is exactly or approximately diagonalized, the diagonal elements \( \langle |X_i|^2 \rangle \) would be a fair estimate of the power spectrum or band power spectrum. Thus, measuring the power spectrum mathematically is almost a synonym for diagonalizing the covariance of the density field \( \delta(x) \), or calculating the eigenvalues of the covariance matrix.

Traditionally, the Fourier decomposition, and then the Fourier power spectrum, are the popular tool for analyzing a cosmic density field, because the Fourier transform retains the translation invariance of a homogeneous and isotropic universe. However, the observed samples given by redshift surveys are not translation invariant, due to the selection effect and irregular geometry of the surveys. To effectively compare the predicted power spectrum with the observed galaxy distributions, the basis functions of the decomposition should be chosen to incorporate the selection effect,
sampling, and complex geometry of the data. As a result, various decompositions for measuring the galaxy power spectrum have been proposed (Tegmark et al. 1998 and reference therein). An ideal estimator of the power spectrum should match the following conditions:

1. \( X_i \)'s are independent from each other, i.e., the data are decomposed into mutually exclusive chunks.
2. \( X_i \)'s retain all the information of the original data, i.e., the decomposed chunks are collectively exhaustive.
3. It is computationally feasible.
4. It allows us to take into account the systematic effects, such as redshift distortion, evolution, morphology dependence, Galactic extinction, etc.

These ideal estimators are believed to be information lossless, i.e., retaining all the information on the power spectrum present in the original data.

In this paper we study the estimator based on the multi-scale decomposition, i.e., the discrete wavelet transform (DWT) representation. The DWT power-spectrum estimator has been applied to measure the power spectrum from samples of the Lyz forests of QSO absorption spectra (Pando & Fang 1998a). The result has demonstrated that the DWT power-spectrum estimator can match the conditions listed above; in particular, it is very helpful in overcoming the difficulties of complex geometry and sampling. Within the framework of DWT, this paper will present a general working scheme for extracting the statistical characteristics from the observational data, in which the selection effect, sampling, and binning are addressed.

It has been recognized recently that the non-Gaussian behavior of \( X_i \) is substantial for a precise measurement of the power spectrum. The accuracy of a power spectrum estimation is significantly affected by the so-called power-spectrum correlations induced by nonlinear clustering (Meiksin & White 1999; Scoccimarro, Zaldarriaga, & Hui 1999). The power-spectrum correlation is also found to be essential for recovering the initial power spectrum by a Gaussianization of observed distribution (Weinberg 1992; Narayanan & Weinberg 1998; Feng & Fang 2000). Thus, beyond the conditions mentioned above for an ideal power-spectrum estimator, one should add one more requirement: that the power-spectrum correlation caused by the nonlinear clustering and Poisson sampling are calculable. We will show that the power-spectrum correlations, or the scale-scale correlations, can be calculated in the DWT analysis.

Moreover, for popular models of the cold dark matter (CDM) cosmogony, including the standard cold dark matter models (SCDM), open CDM (ODM), and flat CDM (ACDM), the scale-scale correlations have been found to be negligible on large scales, and the nonlocal scale-scale correlations are also negligible even on small scales (Feng, Deng, & Fang 2000). That is, the effect of the power-spectrum correlations is largely suppressed in the DWT representation. We show how to take advantage of this suppression for a scale-by-scale approach to measuring the power spectrum.

The paper is organized as follows. Section 2 gives a brief description of the DWT decomposition of the covariance of density random field. The physical meaning and mathematical properties of the \( j \) diagonal and \( j \) off-diagonal components of the covariance will also be discussed. In § 3, an optimized-band power-spectrum estimator based on the DWT \( j \) diagonal covariance is proposed. In addition, the scale-scale correlation extracted from the \( j \) off-diagonal components of the covariance is investigated. This correlation gives the scale range in which the power spectra obtained by the \( j \) diagonalization are information lossless. We then present the algorithm for estimating the DWT band power spectrum from observed galaxy catalogs. It includes the DWT binning (§ 4) and the DWT technique for dealing with Poisson sampling and selection (§ 5). A discussion and conclusions are given in § 6. A brief introduction to DWT analysis is given in the Appendix.

2. COVARIANCE OF DENSITY FLUCTUATIONS IN THE DWT REPRESENTATION

2.1. DWT Decomposition of Density Fields

For the sake of simplicity, we analyze a one-dimensional density distribution, \( \rho(x) \), in the range \( 0 < x < L \), which is assumed to be a stationary random field. The density contrast is defined by \( \delta(x) = \langle \rho(x) - \bar{\rho} \rangle / \bar{\rho} \), where \( \bar{\rho} = \langle \rho(x) \rangle \), and \( \langle \ldots \rangle \) stands for ensemble average. It would be straightforward to extend most results to two and three dimensions. Some specific problems related to higher dimension extension will be discussed in § 6. In addition, the redshift distortion will not be taken into account in this paper.

To ensure that a multiscale decomposition of \( \delta(x) \) will be information-lossless, the natural working scheme is adopt discrete wavelet transformation (DWT) within the framework of multiresolution analysis (MRA). The mathematical construction of MRA theory is briefly sketched in the Appendix.

Let \( \delta^P(x) \) be the periodic extension of \( \delta(x) \), i.e., \( \delta^P(x) = \delta(x - \lfloor x/L \rfloor L) \), where \( \lfloor \ldots \rfloor \) denotes the integer part of \( \eta \). From equation (A36), the density contrast, \( \delta^P(x) \), can be decomposed in term of an orthonormal wavelet basis:

\[
\delta^P(x) = \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \tilde{e}_{j,l} \psi_{j,l}(x) .
\]  

The wavelet function coefficient (WFC), \( \tilde{e}_{j,l} \), is given by the inner product of

\[
\tilde{e}_{j,l} = \langle \psi_{j,l} | \delta \rangle \equiv \int_{-\infty}^{\infty} \delta^P(x) \psi_{j,l}(x) dx ,
\]

which describes the density fluctuation on a scale \( L/2^j \) at position \( L/2^j \). The WFCs are the variables of the random field in the DWT representation. The original distributions can be exactly and unredundantly reconstructed from these decomposed variables.

By using the periodized wavelet function defined by

\[
\psi^P_{j,l}(x) = \left( \frac{2}{L} \right)^{1/2} \sum_{n=-\infty}^{\infty} \psi \left( \frac{x + n L}{L} - l \right) ,
\]

where \( \psi \) is the basic wavelet function (eq. [A21]), equation (1) becomes

\[
\delta^P(x) = \sum_{j=0}^{\infty} \sum_{l=0}^{2^j - 1} \tilde{e}_{j,l} \psi^P_{j,l}(x) .
\]

The WFC can then be computed by

\[
\tilde{e}_{j,l}^P = \int_{0}^{L} \delta^P(x) \psi^P_{j,l}(x) dx .
\]

We will always use the periodized functions below, and we drop the superscript \( P \).
Furthermore, \( \Psi_{j,l}(x) \) is admissible (eq. [A27]), which implies that \( \psi_{j,l}(x) \) has zero mean if it is integrable,
\[
\int \psi_{j,l}(x) \, dx = 0 . \tag{7}
\]
It then follows from equation (2) that
\[
\langle \hat{\epsilon}_{j,l} \rangle = 0 . \tag{8}
\]

The Fourier decomposition of the field \( \delta(x) \) is given by
\[
\delta(x) = \sum_{n=-\infty}^{\infty} \delta_n e^{2\pi i nx/L} , \tag{9}
\]
where \( n \) is an integer, and the Fourier coefficient, \( \delta_n \), is
\[
\delta_n = \langle \delta(x) \rangle = \frac{1}{L} \int_{0}^{L} \delta(x) e^{-2\pi i nx/L} \, dx . \tag{10}
\]
Since both the bases of the Fourier transform and the DWT are orthogonal and complete in the space of one-dimensional functions with period length \( L \), we have
\[
\sum_{j=0}^{\infty} \sum_{l=0}^{2^{j-1}} \langle \delta(x) \rangle = \delta_{k,k'} , \tag{11}
\]
where \( \delta_{k,k'} \) is the Kronecker delta function, and \( \langle \delta(x) \rangle \) is the Fourier transform of the wavelet \( \psi_{j,l} \) given by
\[
\hat{\psi}_{j,l}(n) \equiv \langle \delta(x) \rangle = \frac{1}{L} \int_{0}^{L} \psi_{j,l}(x) e^{-2\pi i nx/L} \, dx . \tag{12}
\]

Considering that the wavelet \( \psi_{j,l}(x) \) is related to the basic wavelet \( \psi(n) \) by equation (A11), equation (12) can be rewritten as
\[
\hat{\psi}_{j,l}(n) = \left( \frac{2}{L} \right)^{-1/2} \hat{\psi}(n/2) e^{-2\pi i nl/2L} , \tag{13}
\]
where \( \hat{\psi}(n) \) is the Fourier transform of the basic wavelet,
\[
\hat{\psi}(n) = \int_{0}^{L} \psi(x) e^{-2\pi i nx/L} \, dx . \tag{14}
\]
Substituting equation (9) into equation (6) yields
\[
\hat{\epsilon}_{j,l} = \sum_{n=-\infty}^{\infty} \delta_n \int_{0}^{L} \psi_{j,l}(x) e^{-2\pi i nx/L} \, dx = \sum_{n=-\infty}^{\infty} \delta_n \hat{\psi}_{j,l}(n) . \tag{15}
\]
Similarly, inserting equation (5) into equation (10), we have
\[
\delta_n = \frac{1}{L} \sum_{j=0}^{\infty} \sum_{l=0}^{2^{j-1}} \hat{\epsilon}_{j,l} \hat{\psi}_{j,l}(n) = \sum_{j=0}^{\infty} \sum_{l=0}^{2^{j-1}} \left( \frac{1}{2L} \right)^{1/2} \hat{\epsilon}_{j,l} e^{-2\pi i nl/2L} \hat{\psi}(n/2) , \tag{16}
\]
\( n \neq 0 \).

Equations (15) and (16) show that both the Fourier variables, \( \delta_n \), and the DWT variables, \( \hat{\epsilon}_{j,l} \), are complete.

However, the statistical properties of the Fourier mode \( n \) and the DWT mode \( (j, l) \) are quite different. For a non-Gaussian field consisting of randomly homogeneously distributed clumps with a non-Gaussian probability distribution function (PDF), the one-point distributions of the real and imaginary components of the Fourier modes could be still Gaussian. This is because the Fourier modes are subject to the central limit theorem of random fields (Adler 1981). Even though the non-Gaussian clumps are correlated, the central limit theorem still holds if the two-point correlation function of the clumps approaches zero sufficiently fast (Fan & Bardeen 1995). Thus, the non-Gaussian information could be lost in the Fourier representation if the phases of the Fourier coefficients are missing.

On the other hand, the DWT basis does not suffer from the central limit theorem. A key condition necessary for the central limit theorem to hold is that the modulus of the decomposition basis is less than \( C/L^{1/2} \), where \( L \) is the size of the sample and \( C \) is a constant (Ivanov & Leonenko 1989). The Fourier basis obviously satisfies this condition, because \( (1/L^{1/2}) | \sin 2\pi nx/L | < C/L^{1/2} \), where \( C \) is independent of \( x \) and \( n \). However, the DWT basis is compactly supported (see Appendix), and its modulus does not satisfy the condition of being less than \( C/L^{1/2} \). Consequently, for the non-Gaussian fields, the one-point distributions of the Fourier variables, \( \delta_n \), could be Gaussian, while for the DWT variable \( \hat{\epsilon}_{j,l} \) the one-point distributions show non-Gaussianity (Pando & Fang 1998b).

2.2. The WFC Covariance and DWT Power Spectrum

In the DWT representation, the covariance, \( \langle \delta(x) \delta(x') \rangle \), is expressed by a matrix \( \langle \hat{\epsilon}_{j,l} \hat{\epsilon}_{j',l'} \rangle \) with subscripts \( (j, l) \) and \( (j', l') \). The elements of \( j = j' \) and \( l = l' \) will be called diagonals, while \( j \neq j' \) will be called the \( j \) diagonals. The Parseval theorem for the DWT decomposition is (Fang & Thews 1998)
\[
\frac{1}{L} \int_{0}^{L} | \delta(x) |^2 \, dx = \sum_{j=0}^{\infty} \frac{1}{L} \sum_{l=0}^{2^{j-1}} | \hat{\epsilon}_{j,l} |^2 , \tag{17}
\]
which implies that the power of perturbations can be divided into modes, \( (j, l) \). The term \( | \hat{\epsilon}_{j,l} |^2 \) describes the power of the mode \( (j, l) \). One can then define the DWT power spectrum by the diagonals of the covariance matrix, i.e.,
\[
P_{j,l} = \langle \hat{\epsilon}_{j,l} \hat{\epsilon}_{j,l} \rangle . \tag{18}
\]
Since the random variables \( \hat{\epsilon}_{j,l} \) are complete, one can define a Gaussian field \( \delta(x) \) by requiring that all the variables \( \hat{\epsilon}_{j,l} \) are distributed as a Gaussian process with the covariance
\[
\langle \hat{\epsilon}_{j,l} \hat{\epsilon}_{j',l'} \rangle = P_{j,l} \delta_{j,j'} \delta_{l,l'} , \tag{19}
\]
and the zero ensemble average of all higher order cumulants of \( \hat{\epsilon}_{j,l} \). Thus, a Gaussian field is completely described by its DWT power spectrum, \( P_{j,l} \). For a homogeneous Gaussian field, the DWT power spectrum, \( P_{j,l} \), is \( l \)-independent, i.e., \( P_{j,l} = P_j \).

Using equations (15) and (16), the covariance in the Fourier and DWT representations can be converted from one form to another by
\[
\langle \hat{\delta}_n \hat{\delta}_{n'} \rangle = \sum_{j=0}^{\infty} \sum_{l=0}^{2^{j-1}} \sum_{i=0}^{2^{j-1}} \langle \hat{\epsilon}_{j,l} \hat{\epsilon}_{j',l'} \rangle \hat{\psi}_{j,l}(n) \hat{\psi}_{j',l'}(n') \tag{20}
\]
and conversely,
\[
\langle \hat{\epsilon}_{j,l} \hat{\epsilon}_{j',l'} \rangle = \sum_{n,n'} \langle \hat{\delta}_n \hat{\delta}_{n'} \rangle \hat{\psi}_{j,l}'(n) \hat{\psi}_{j',l'}(n') . \tag{21}
\]
\(^1\) The DWT power spectrum, or called scalogram, has been extensively applied in signal analysis (e.g., Mallat 1999).
Therefore, for a homogeneous Gaussian field given by the DWT power spectrum \( P_j \), equation (20) implies
\[
\langle \delta_\ell \delta_\ell^* \rangle = P(n)\delta_{n,n'} \, ,
\]  
(22)
where
\[
P(n) = \sum_{j=0}^{\infty} P \left| \hat{\psi}(\frac{n}{2^j}) \right|^2 .
\]  
(23)

In the derivation of equation (22), we used
\[
\sum_{l=0}^{2^j-1} e^{-i2\pi(n-n')/2^j} = 2\delta_{n,n'} .
\]  
(24)

Equation (22) shows that for a homogeneous Gaussian \( P_j \), the Fourier power spectrum, \( P(n) \), is uniquely determined by the DWT power spectrum, \( P_j \).

However, the reversed relation does not exist, i.e., one cannot show that the DWT covariance is given by equation (19) with \( P_{j,l} = P_j \) if the Fourier covariance is given by equation (22). This indicates that the Fourier and WFC covariances are not equivalent. For instance, fields consisting of homogeneously distributed non-Gaussian clumps generally do not satisfy equation (19) with an \( l \)-independent \( P_{j,l} \), but do so for equation (22). That is, equation (19) with an \( l \)-independent \( P_{j,l} \) places stronger constraints on the random field than equation (22), and therefore equation (22) will hold when equation (19) with an \( l \)-independent \( P_{j,l} \) holds, but the converse is not generally true.

2.3. \( j \) Off-Diagonals of the WFC Covariance

We now identify the physical meaning of the \( j \) off-diagonal components of the WFC covariance.

When the “fair sample hypothesis” (Peebles 1980) holds, or equivalently, the random field is ergodic, then the \( 2^j \) WFCs \( \tilde{\epsilon}_{j,l} \), \( l = 0, \ldots, 2^j - 1 \), for a given \( j \) can be taken as \( 2^j \)-independent measurements, because they are measured by projecting onto the mutually orthogonal basis \( \psi_{j,l}(x) \). Accordingly, the \( 2^j \) WFCs form a statistical ensemble on the scale \( j \). This ensemble actually represents the one-point distribution of the fluctuations of the DWT modes at a given scale \( j \). The average over \( l \) is thus a fair estimation of the ensemble average.

For a Gaussian field, these one-point distributions are Gaussian. However, even if the one-point distributions for all \( j \) are Gaussian, the density field \( \delta(x) \) could still be non-Gaussian. That is simply due to the statistical properties of the WFCs \( \tilde{\epsilon}_{j,l} \) for indices \( j \) and \( l \) that are independent. It is easy to construct a density field \( \delta(x) \) for which the WFCs \( \tilde{\epsilon}_{j,l} \) are Poisson or Gaussian in their one-point distributions with respect to \( l \), yet highly non-Gaussian in terms of \( j \) (Greiner, Lipa, & Carruthers 1995). A simple example is demonstrated as follows. Suppose the one-point distribution of the \( 2^j \) WFCs \( \tilde{\epsilon}_{j,l} \), on a scale \( j \), is Gaussian. If the WFCs on the scale \( j + 1 \) are incorporated with those on the scale \( j \), e.g.,
\[
\tilde{\epsilon}_{j+1,2l} = a\tilde{\epsilon}_{j,l} \, ,
\]
\[
\tilde{\epsilon}_{j+1,2l+1} = b\tilde{\epsilon}_{j,l} \, ,
\]  
(25)
where \( a \) and \( b \) are arbitrary constants, the one-point distribution of the \( 2^{j+1} \) WFCs \( \tilde{\epsilon}_{j+1,l} \) is also Gaussian. However, the coherent structure given by equation (25) leads to a strong correlation between \( \tilde{\epsilon}_{j+1,l} \) and \( \tilde{\epsilon}_{j,l} \), i.e., the scale \( j + 1 \) fluctuations are always proportional to those on the scale \( j \) at the same position. This is a local scale-scale correlation. One can also design a nonlocal scale-scale correlation by
\[
\tilde{\epsilon}_{j+1,2l+1} = a\tilde{\epsilon}_{j,l} \, ,
\]
\[
\tilde{\epsilon}_{j+1,2l+1} = b\tilde{\epsilon}_{j,l} \, ,
\]  
(26)
where \( l \) is a fixed number. Equation (26) leads to a strong correlation between the fluctuations on scales \( j + 1 \) and \( j \), but at two places with distance \( l \).

Hence, in terms of the DWT representation, a homogeneous Gaussian field requires that (1) the one-point distributions of the WFCs with respect to \( l \) are Gaussian, and (2) the distributions of WFCs with different \( j \)’s are uncorrelated, such that
\[
\langle \tilde{\epsilon}_{j+1,l} \tilde{\epsilon}_{j',l'} \rangle = 0 \, .
\]  
(27)

Correspondingly, in the Fourier representation, a Gaussian field also has two requirements: (1) the one-point distributions of the amplitudes of the Fourier mode \( \delta_\ell \) are Gaussian; and (2) the phases of \( \delta_\ell \) are random. Therefore, equation (27) is the DWT counterpart of the Fourier random phase. However, it is difficult, practically impossible, to capture the phase information of each Fourier mode. The local scale-scale correlation is overlooked with the Fourier covariance.

In summary, the \( j \) off-diagonals of the WFC covariance provide information on the scale-scale correlation. This non-Gaussian feature arises from mode-mode coupling of gravitational clustering, and cannot be measured by the higher order cumulants of the one-point distribution for a given scale \( j \), but rather must be measured by the cross-correlation between the different scales. The covariance of a system without scale-scale correlations will be \( j \)-diagonal, i.e.,
\[
\langle \tilde{\epsilon}_{j,l} \tilde{\epsilon}_{j',l'} \rangle = \langle \tilde{\epsilon}_{j,l} \rangle \langle \tilde{\epsilon}_{j',l'} \rangle = 0 \, , \quad j \neq j' \, ,
\]  
(28)
where equation (8) has been used at the last step.

3. STATISTICAL INFORMATION EXTRACTED FROM THE WFC COVARIANCE

3.1. \( j \)-Diagonalization of the WFC Covariance

It has been known that the DWT is a powerful tool for data compression. For very wide types of stochastic clustering processes, the off-diagonal components of the covariance are strongly suppressed in the DWT representation. This suppression is especially efficient for self-similar clustering. For instance, one can show analytically that the covariance in the DWT representation is exactly diagonal for some popular hierarchical models of structure formation, such as the block model and its variants (Meneveau & Sreenivasan 1987; Cole & Kaiser 1988). In this respect, the DWT basis represents the adequate normal coordinates. In other words, the DWT analysis can be understood as a proper orthonormal decomposition (POD), or a Karhunen-Loève transformation (e.g., Aubry et al. 1988), in regard to the second-order correlations of these stochastic clustering processes.

For more realistic models and observed samples, the WFC covariance is not fully diagonal, but mostly \( j \)-diagonal. In fact, this character has been evident from the measurement of the fourth-order scale-scale correlation in the observational samples such as the Lyα forest lines.
(Pando et al. 1998a), the transmitted flux of QSO absorption spectra (Feng & Fang 2000), and the APM bright galaxy catalog (Feng, Deng, & Fang 2000). A common conclusion is that the scale-scale correlations are very weak, and negligible on large scales, i.e., \( \langle \xi_{j,1} \xi_{j,1}^* \rangle = \langle \xi_{j,1} \rangle^2 \langle \xi_{j,1}^* \rangle = 0 \) for \( j \neq j' \) and \( j, j' \leq J_{ss} \), where \( J_{ss} \) denotes the scale above which the scale-scale correlation is not significant. It is also true for the mass distributions and two- and three-dimensional mock catalog of galaxies in the CDM family of models (Feng et al. 2000). This result indicates \( \text{Sv8} \models \text{models} \) (Feng et al. 2000). This result indicates \( \text{Sv8} \models \text{models} \).

This result indicates \( \text{Sv8} \models \text{models} \). It is also true for the mass distributions and two- and three-dimensional mock catalog of galaxies in the CDM family of models (Feng et al. 2000). The extracted from one realization of an ergodic field. The statistically valuable band power spectrum that can be obtained from this is now referred to as the DWT power spectrum. As we have only one realization of the cosmic mass distribution, i.e., \( \text{Sv8} \), the banded power spectrum? The DWT representation provides a natural and reasonable approach to the banding.

The DWT power spectrum is conveniently expressed in the Fourier representation, any statistical estimator designed to measure the power spectrum from real data should have a simple relation with the Fourier power spectrum. Since we have only one realization of the cosmic mass field, no ensemble is available for each mode \( n \). One cannot measure the Fourier power spectrum \( P(n) \), since it is from the variance of the amplitude \( |\delta_n| \) of mode \( n \). Generally, a power-spectrum estimator measures the banded power spectrum as

\[
P_j = \sum_n W(n)P(n),
\]

where \( W(n) \) is a window function localized in the \( n \) (or Fourier) space. The problem that arises here is, what is the criterion for a reasonable banding? and how to optimize the banded power spectrum? The DWT representation provides a natural and reasonable approach to the banding.

As discussed in \( \S \) 2.2, for an ergodic field, the \( 2^j \) WFCs \( \hat{\xi}_{j,i} \) at a given \( j \) form a one-point distribution of the fluctuations at the scale \( j \). Therefore, the DWT power spectrum at the scale \( j \) can be defined as the variance of the one-point distribution, i.e.,

\[
P_j = \frac{1}{2} \sum_{i=0}^{2^{j-1}} (\hat{\xi}_{j,i} - \langle \hat{\xi}_{j,i} \rangle)^2.
\]

Because of the zero mean of WFC \( \langle \hat{\xi}_{j,i} \rangle \) (eq. [8]), \( P_j \) can be written, statistically, as

\[
P_j = \frac{1}{2} \sum_{i=0}^{2^{j-1}} |\hat{\xi}_{j,i}|^2 = \frac{1}{2} \sum_{i=0}^{2^{j-1}} P_{j,i},
\]

which is an ergodicity-allowed spatial average of \( P_{j,i} \), and is usually referred to as the DWT power spectrum. As we show below, equation (31) gives an estimator of the band-averaged Fourier power spectrum.

The DWT power spectrum given by equation (31) is certainly less detailed than the power spectrum \( P(n) \) or \( P_{j,i} \). However, the numbers \( P_{j,i} \) are probably the maximum of the statistically valuable band power spectrum that can be extracted from one realization of an ergodic field. The optimum of this banding can be seen via the phase space \( \{x, k\} \), where the wavenumber \( k = 2\pi n/L \). Generally, a set of orthogonal and complete bases of multiresolution analysis decomposes the entire phase space into elements with different shapes, but their volume always satisfies the uncertainty relation \( \Delta x \Delta k \geq 2\pi \). The ordinary Fourier transform is not a multiresolution decomposition, but always takes the highest resolution of \( k \), i.e., \( \Delta k \to 0 \), and lowest resolution of \( x \), \( \Delta x \to \infty \).

To apply the ergodicity, we chopped the survey volume \( L \) into pieces \( \Delta x \). If \( \Delta x \) is too large, or \( L/\Delta x \) too small, the ensemble contains few members, and thus there will be larger vertical errors placed on the estimated power spectrum. In order to minimize this error, we can make the size of chopped pieces \( \Delta x \) small. Correspondingly, the width of the window function \( \Delta k = 2\pi/\Delta x \) will broaden, and the scale resolution will be poor, i.e., there will be a large horizontal error bar placed on the estimated power spectrum. Thus, the optimal chopping can be achieved by a compromise between these two trade-off factors, \( L/\Delta x \) and \( \Delta k \). Generally, \( 1/\Delta x \) is proportional to the resolvable wavenumber, i.e.,

\[
1/\Delta x \propto k;
\]

therefore, the optimized banding \( \Delta k \Delta x = 2\pi \) requires

\[
\frac{\Delta k}{k} = \Delta \log k \approx 1.
\]

That is, the optimized banding is in logarithmic spacing. To detect small-scale fluctuations (larger wavenumber \( k \)), the size of the pieces \( \Delta x \) is chosen to be smaller. To detect large-scale fluctuations (smaller wavenumber), the size of the pieces \( \Delta x \) is chosen to be larger. The wavelets \( \psi_{j,(x)} \) are constructed by dilating (i.e., changing the scale of) the generating function by a factor of \( 2^j \) (see the Appendix). Therefore, we have \( \Delta \log k \sim 1 \). In this sense, the DWT is an optimized multiscale decomposition (Farge 1992). Because the set of wavelet bases is complete, one cannot have more independent bands than \( P_j \).

Under the assumption of a homogeneous Gaussian field, the DWT power-spectrum equation (31) can be rewritten as

\[
P_j = \frac{1}{2} \sum_{n=-\infty}^{\infty} \left| \hat{\psi} \left( \frac{n}{2^j} \right) \right|^2 P(n),
\]

where equations (11), (21), and (22) have been used. Comparing with equation (29), clearly \( P_j \) is a band-averaged Fourier power spectrum with the window function \( W(n) = \frac{1}{2} \left| \hat{\psi} \left( \frac{n}{2^j} \right) \right|^2 \).

Generally, the function \( \hat{\psi}(n) \) is non-zero in two narrow wavenumber ranges centered at \( n = \pm n_p \), with width \( \Delta n_p \). Therefore, \( P_j \) is the band spectrum centered at

\[
\ln n_j = j \log 2 + \log n_p,
\]

with the bandwidth

\[
\Delta \log n = \frac{\Delta n_p}{n_p},
\]

which stays constant logarithmically. Equations (36) and (37) show that the countable data set \( \{P_j, j = 1, 2, \ldots\} \) represents a scale-by-scale band-averaged Fourier power spectrum with the logarithmic spacing of the wavenumber. \( P_j \) is completely determined by the Fourier power spectrum, and
therefore it should be effective for constraining the parameters contained in the Fourier power spectrum.

The band power spectrum equation (31) can also be written as, alternatively,

$$P_j = \frac{1}{2} \text{tr} \text{Cov}_j \ (38)$$

where the matrix \(\text{Cov}_j\) is the \(j\) submatrix of the covariance, i.e.,

$$\text{Cov}_j = \tilde{\varepsilon}_{j,t} \tilde{\varepsilon}_{j,t}^\dagger \ (39)$$

Therefore, \(P_j\) exhausts all the information of the \(j\) diagonals of the WFC covariance. Equation (38) shows that we actually do not need to diagonalize each \(j\) submatrix, since \(P_j\) is given by the trace of the \(j\) submatrix.

3.3. Scale-Scale Correlations in Second and Higher Orders

In the range of \(j > J_{\text{max}}\), the scale-scale correlations become significant, and the DWT covariance will no longer be diagonal or \(j\)-diagonal.

In this scale range, we should somewhat diagonalize the DWT covariance. However, the scale-scale correlation may lead to large errors of the diagonalization, even rendering the diagonalization impossible. Let us consider the example of the scale-scale correlation given by equation (25). In this case, the variable \(\tilde{\varepsilon}_{j+1, t}\) is actually linearly dependent on \(\tilde{\varepsilon}_{j, t+\Delta l}\), and therefore the matrix \(\langle \tilde{\varepsilon}_{j+1, t} \tilde{\varepsilon}_{j, t} \rangle\) is singular. It cannot be diagonalized. For instance, for scales \(j = 1, 2\), the covariance matrix is now

$$\begin{pmatrix}
\varepsilon_{1,0} & \varepsilon_{1,0} & \varepsilon_{2,0} & \varepsilon_{2,1} \\
\varepsilon_{1,0} & \varepsilon_{1,0} & \varepsilon_{2,0} & \varepsilon_{2,1} \\
\varepsilon_{1,1} & \varepsilon_{1,1} & \varepsilon_{2,0} & \varepsilon_{2,1} \\
\varepsilon_{1,1} & \varepsilon_{1,1} & \varepsilon_{2,0} & \varepsilon_{2,1}
\end{pmatrix} = \varepsilon_{1,0}^2 \begin{pmatrix}
1 & a & b \\
a & a^2 & ab \\
b & ab & b^2
\end{pmatrix} .$$

(40)

Obviously, this matrix cannot be diagonalized.

More seriously, if the matrix elements have some uncorrelated errors due to measurements, i.e., \(\tilde{\varepsilon}_{j,t} \tilde{\varepsilon}_{j,t}^\dagger = \Delta \tilde{\varepsilon}_{j,t} \tilde{\varepsilon}_{j,t}^\dagger\), the matrix given in equation (40) looks diagonalizable. However, in this case the minors of the matrix are given by the errors \(\Delta \tilde{\varepsilon}_{j,t} \tilde{\varepsilon}_{j,t}^\dagger\), and therefore the diagonalization will be largely contaminated by the errors.

This example indicates that when the scale-scale correlations appear, the number of the independent variables, and then the signal-to-noise ratio, will decrease. We should not extract the statistical properties of the covariance by a diagonalization.

Fortunately, our ultimate goal is not mathematical diagonalization, but discrimination among physical models of the structure formation. An alternative to the full diagonalization is to take the following two measures: (1) Using the \(j\)-diagonals of each \(j\) to calculate the band power spectrum \(P_j\) (eq. [31]); (2) using the \(j\)-off-diagonals to calculate the second-order scale-scale correlations. The second-order scale-scale correlation is defined as

$$C_{j,j'}(\Delta l) = \frac{1}{2} \sum_{l=0}^{2j-1} \varepsilon_{j,l} \tilde{\varepsilon}_{j',l}^\dagger , \quad j > j' ,$$

\(l = \text{mod} \left( \frac{l}{2^{j'-j}} \right) + \Delta l . \ (41)$$

Like the band power spectrum (eqs. [30] and [31]), \(C_{j,j'}(\Delta l)\) is defined by an ergodicity-allowed average. The term \(C_{j,j'}(\Delta l)\) measures the second-order correlation between fluctuations on scales \(j\) and \(j'\) at positions \(l\) and \(l'\). Since the cosmic density field is homogeneous, the correlation depends only on the difference between \(l\) and \(l'\), i.e., \(\Delta l/2^{j'}\). For an initially Gaussian field, the scale-scale correlations are developed during the nonlinear evolution of the gravitational clustering.

Now we can use the two statistics \(P_j\) and \(C_{j,j'}(\Delta l)\) to discriminate among models. In fact, a discrimination with these two statistics would be worth more than a full diagonalization. For instance, the model-predicted galaxy power spectra on smaller scales are generally degenerate with respect to cosmological parameters, i.e., models with different cosmological parameters can yield the same galaxy power spectrum. This is because one can always choose the bias model parameters to fit the prediction with the observations. Therefore, to remove the degeneracy, an independent measure for constraining the bias models is necessary. The scale-scale correlation is found to be sensitive to the bias model (Feng et al. 2000). Thus, for model discrimination, the \(j\)-diagonal power spectrum plus scale-scale correlation would be more useful than a full diagonalization.

In a word, in the scale range of \(j > J_{\text{max}}\), we will extract the valid statistical information from the covariance by \(P_j\) and \(C_{j,j'}(\Delta l)\).

It should be pointed out that even when all \(C_{j,j'}(\Delta l)\) vanish, one cannot conclude that the system is scale-scale uncorrelated. In other words, that a decomposition \(X_i\) yields a diagonal covariance does not mean that the modes \(X_i\) are really statistically uncorrelated. There are many clustering models that have diagonal covariance, but mode-mode statistics are correlated on higher orders (Greiner et al. 1995). A diagonal decomposition means only that mode-mode coupling is uncorrelated at second order.

The higher order generalization of \(C_{j,j'}(\Delta l)\) is straightforward. For instance, one can measure the fourth-order scale-scale correlations by

$$C_{j,j'}^2(\Delta l) = \frac{1}{2^4} \sum_{l=0}^{2^{j'-j}-1} \varepsilon_{j,l} \varepsilon_{j',l}^\dagger , \quad j > j' ,$$

\(l = \text{mod} \left( \frac{l}{2^{j'-j}} \right) + \Delta l . \ (42)$$

This correlation, \(C_{j,j'}^2(\Delta l = 0)\), is essentially the same as the so-called band-band correlation, defined by

$$T = \frac{\langle P_j P_{j+1} \rangle}{\langle P_j \rangle \langle P_{j+1} \rangle} . \ (43)$$

It has been shown that the precision of the Fourier band power-spectrum estimator depends on the band-band correlation \(T\) (Meiksin & White 1999). In the DWT representation, we arrive at the similar conclusion that when \(C_{j,j'}(\Delta l)\) or \(C_{j,j'}^2(\Delta l)\) are nonzero, i.e., when the DWT covariance is not \(j\) diagonal, we should test models by both the band power spectrum and scale-scale correlations. For samples of large-scale structure, the scale-scale correlation \(C_{j,j}(\Delta = 0)\) has been found to be significant on scales of less than about 10 \(h^{-1}\) Mpc (Pando et al. 1998a; Feng et al. 2000).

4. THE DWT ALGORITHM OF DATA BINNING

In the following two sections, we discuss the algorithm for estimating the band power spectrum, \(P_j\), and scale-scale
correlations, $C_{j_L}(\Delta l)$, from galaxy redshift surveys and other samples of large-scale structure.

If the position measurement is perfectly precise, the observed galaxy distribution can be written as

$$\rho(x) = \sum_{i=1}^{N_g} w_i \delta_p(x - x_i) ,$$

(44)

where $N_g$ is the total number of galaxies, $x_i$ is the position of the $i$th galaxy, $0 \leq x_i \leq L$, $w_i$ is its weight, and $\delta_p$ is the Dirac $\delta$ function. However, the position measurement has error due to finite spatial resolution, and therefore the distribution is usually somewhat given by a binned histogram.

The binning is performed by a convolution of the data with a binning function $W(x)$ as

$$\tilde{\rho}(x) = \Pi(x) \int W(x - x') \rho(x') dx' ,$$

(45)

where $\Pi(x)$ is the sampling function, defined as $\Pi(x) = \sum_l \delta_p(x - IL/2^l)$, where $l$ labels the $l$th bin. Obviously, the mesh-defined density distribution is given by $\tilde{\rho}(x) = \sum_l \rho_l \delta_p(x - IL/2^l)$, where $\rho_l = \int W((IL/2^l - x) \rho(x') dx'$ is a mass assignment at the $l$th bin.

It is well known that the binning equation (45) will result in spurious features of the Fourier power spectrum on scales around the Nyquist frequency of the fast Fourier transform (FFT) grid (e.g., Jing 1992; Percival & Walden 1993; Baugh & Efstathiou 1994). Mathematically, equation (45) implies a decomposition by the weight function, $W(x)$. In other words, $W(IL/2^l - x)$ plays the role of a scaling function (or sampling function). If the scaling functions are orthogonal and complete, then one cannot recover the original field without distortion. This may cause some spurious features, such as the aliasing effect in the FFT. In the DWT analysis, the binning or sampling is always done by an orthogonal and complete decomposition; thus, one can expect that the spurious features and false correlations can be completely avoided.

### 4.1. Binning with Wavelets

The WFCs $e_{j_l}$ are assigned at regular grids $l = 0, \ldots, 2^{-l}$. This is actually a binning of data. In this case, the binning is automatically realized by the orthogonal projection onto wavelet space, and no extra weight function is required. The result is that the contamination due to the sampling error is naturally eliminated.

With equation (6), one can directly calculate the WFCs of the galaxy distribution equation (44) by

$$e_{j,l} = \sum_{i=1}^{N_g} w_i \psi_j(x_i) .$$

(46)

The errors on $e_{j,l}$ can also be calculated from the errors on $x_i$.

Since we used the periodized distribution $\delta(x)$ in equation (6), the discontinuity between the data at the two boundaries may introduce false coefficients. However, this possible false signal is only related to the boundaries. One can expect that these false coefficients will not be important for detecting the power spectrum on scales much less than $L$. This boundary effect has been tested numerically by using simulated samples over a finite length divided into 512 bins with two different boundary conditions: (1) periodic boundary conditions, and (2) zero padding. The results show that the spectrum can be correctly reconstructed by the DWT regardless of the boundary conditions on scales to or less than 64 bins (Pando & Fang 1998a).

We should note the difference between the usual mass assignment and the DWT projection (eq. [46]). In the former, the mass assignment is given by partitioning the mass on the grids according to the binning function, $W(x)$, and the binning data are the mesh-defined densities. For the DWT projection, on the other hand, the binning data, i.e., the WFCs $e_{j,l}$, are not the mesh-defined densities, but the fluctuations on scale $j$ at position $l$, which is obviously not positive-definite.

### 4.2. Binning with Scaling Functions

In the DWT analysis, the mass assignment is realized by the scaling function $\phi_j(x)$ (eq. [A30]). In addition to the orthogonality equations (A33) and (A34), the basic scaling function $\phi(x)$ (which is not yet periodized) satisfies the so-called "partition of unity" (Daubechies 1992)

$$\sum_{l=-\infty}^{\infty} \phi(\eta - l) = 1 .$$

(47)

One can also define the periodized scaling function as

$$\phi_p(x) = \left( \frac{2}{L} \right)^{1/2} \sum_{n=-\infty}^{\infty} \phi(\frac{2}{L} + n - l) .$$

(48)

Thus, equation (47) can be rewritten as

$$\sum_{l=0}^{2^l-1} \frac{L}{2} \phi_p(x) = 1 .$$

(49)

We will only use the periodized scaling function below, and we drop the superscript $P$.

With the periodized scaling function, equations (A39)–(A41) give

$$\rho(x) = \rho^l(x) + \sum_{j=0}^{\infty} \sum_{l=0}^{2^j-1} e_{j,l} \psi_j(x) ,$$

(50)

where

$$\rho^l(x) = \sum_{l=0}^{2^j-1} e_{j,l} \phi_j(x) .$$

(51)

The scaling function coefficients (SFCs) $e_{j,l}$ are given by

$$e_{j,l} = \frac{L}{\rho(x)} \rho(x) \phi_j(x) dx .$$

(52)

Subjecting the distribution equation (44) to the transform equation (50), we have

$$\rho(x) = \sum_{l=0}^{2^j-1} e_{j,l} \phi_j(x) + \sum_{j=0}^{\infty} \sum_{l=0}^{2^j-1} e_{j,l} \psi_j(x) ,$$

(53)

where

$$e_{j,l} = \sum_{l=1}^{N_g} w_i \phi_j(x_i) .$$

(54)

Using equations (44) and (54), equation (49) yields

$$\sum_{l=0}^{2^j-1} \frac{L}{2} e_{j,l} = \sum_{l=1}^{N_g} w_i .$$

(55)

This shows that the $i$th galaxy is assigned onto grid $l$ by the number $(L/2^l)w_i \phi_j(x_i)$. Therefore, the SFC $(L/2^l)e_{j,l}$ is the mass assignment of $\rho^l(x)$.
4.3. The DWT Binning and FFT

Given a galaxy distribution as in equation (44), its Fourier transform is evaluated by the trigonometric summation

\[ \hat{\rho}(n) = \sum_{i=1}^{N_g} w_i e^{i 2 \pi n x_i / L}, \]

and the power spectrum is \(|\hat{\rho}(n)|^2\). However, the power spectrum given by the FFT of \(\hat{\rho}(n)\) (eq. [45]) is

\[ |\hat{\rho}(n)|^2 = \sum_{n'=0}^{\infty} |\hat{W}(n+2'n')|^2 |\hat{\rho}(n+2'n')|^2, \tag{57} \]

where \(\hat{W}(n)\) is the Fourier transform of the binning function, \(W(x)\). The power spectrum given by equation (57) is obviously not equal to the power spectrum \(|\hat{\rho}(n)|^2\). The power spectrum of equation (57) is given by superpositions of the power spectrum \(|\hat{\rho}(n+2'n')|^2\) on all scales \(n+2'n'\). This is the “aliasing” effect (Hockney & Eastwood 1988; Hoyle et al. 1999).

In the DWT representation, the Fourier transform of equation (53) yields

\[ \hat{\rho}(n) = \sum_{l=0}^{2^{l-1}} \sum_{i=1}^{N_g} \epsilon_{j,l,i} \hat{\phi}_{j,l}(n) + \sum_{j=1}^{\infty} \sum_{l=0}^{2^{j-1}} \sum_{i=1}^{N_g} \epsilon_{j,l,i} \hat{\psi}_{j,l}(n), \tag{58} \]

where the function \(\hat{\phi}_{j,l}(n)\) is the Fourier transform of \(\phi(x)\), i.e.,

\[ \hat{\phi}_{j,l}(n) = \int_{-\infty}^{\infty} \phi(x) e^{-i 2 \pi n x} dx. \tag{59} \]

Using the definition of \(\phi(x)\) (eq. [A30]), equation (59) becomes

\[ \hat{\phi}_{j,l}(n) = \left( \frac{2}{L} \right)^{-1/2} \frac{1}{n/2^{j}} e^{-i 2 \pi n x} \phi(2^{j}n/2^{j}), \tag{60} \]

where \(\hat{\phi}(n)\) is the Fourier transform of the basic scaling function \(\phi(\eta)\),

\[ \hat{\phi}(n) = \int_{-\infty}^{\infty} \phi(\eta) e^{-i 2 \pi n \eta} d\eta. \tag{61} \]

Equation (58) then gives

\[ \hat{\rho}(n) = \left( \frac{2}{L} \right)^{-1/2} \frac{1}{n/2^{j}} \sum_{l=0}^{2^{j-1}} \sum_{i=1}^{N_g} \epsilon_{j,l,i} e^{-i 2 \pi n x} \phi(2^{j}n/2^{j}), \]

\[ + \sum_{j=1}^{\infty} \sum_{l=0}^{2^{j-1}} \frac{1}{n/2^{j}} \hat{\phi}(2^{j}n/2^{j}) \sum_{i=1}^{N_g} \epsilon_{j,l,i} e^{-i 2 \pi n x}. \tag{62} \]

Since \(\hat{\psi}(n/2)\) is localized in \(n/2^{j} \sim n_{p}\), the second term in the right-hand side of equation (62) is important only for \(n \geq 2^{j}n_{p}\). Thus, the Fourier transform \(\hat{\rho}(n)\) can be evaluated by

\[ \hat{\rho}(n) = \hat{\phi}(2^{j}n/2^{j}), \quad n \leq 2^{j}n_{p}, \tag{63} \]

where

\[ \hat{\rho}(n) = \left( \frac{2}{L} \right)^{-1/2} \frac{1}{n/2^{j}} \sum_{l=0}^{2^{j-1}} \sum_{i=1}^{N_g} \epsilon_{j,l,i} e^{-i 2 \pi n x} \phi(2^{j}n/2^{j}). \tag{64} \]

Here \(\hat{\rho}\) can be calculated by the standard FFT technique. Therefore, the Fourier transform of the galaxy distribution \(\rho(x)\) can be evaluated directly by the FFT of its SFC mass assignment, \(\epsilon_{j,l,i}\). Equations (63) and (64) are actually a scale-adaptive FFT for estimating the power spectrum of an irregular data set. This algorithm computes \(\rho(n)\) up to scales \(n \leq 2^{j}n_{p}\), where the adapted scale \(J\) can be chosen up to as high as the scales to be studied.

5. THE DWT ALGORITHM ON THE POISSON SAMPLING

The observed or mock galaxy distributions, \(\rho(x)\), are considered to be a Poisson sampling with an intensity \(\rho(x) = \bar{\rho}(x)[1 + \delta(x)]\), where \(\bar{\rho}(x)\) is the galaxy distribution if galaxy clustering is absent, and are given by the selection function (Peebles 1980). A proper power-spectrum estimator should be effective in obtaining the power spectrum debiased from the Poisson sampling. It has been realized that, in order to handle the Poisson sampling with a non-uniform selection function, the decomposition basis \(\psi(x)\) (eq. [1]) is required to have zero average (e.g., Tegmark et al. 1998), i.e.,

\[ \int \psi(x) dx = 0. \tag{65} \]

This is where we can take advantage of the DWT analysis, since for the wavelets, \(\psi(x)\), equation (65) always holds due to admissibility (eq. [7]).

5.1. Algorithm for the DWT Covariance Affected by Poisson Sampling

Considering the Poisson sampling, the characteristic function of the galaxy distribution, \(\rho(x)\), is

\[ Z[e^{i \int \rho(x) dx}] = \exp \left\{ \int dx \rho^M(x) [e^{\rho(x)} - 1] \right\}. \tag{66} \]

and the correlation functions of \(\rho(x)\) are given by

\[ \langle \rho(x_1), \ldots, \rho(x_n) \rangle_p = \frac{1}{p^n} \int_0^1 \ldots \int_0^1 Z[e^{i \int \rho(x) dx}] \delta^n \left\| \rho(x) - \rho(x) \right\|_p \left( x_1 \ldots x_n \right) \left( x_1 \ldots x_n \right), \tag{67} \]

where \(\langle \ldots \rangle_p\) is the average for the Poisson sampling. We then have

\[ \langle \rho(x) \rangle_p = \rho^M(x) , \tag{68} \]

and

\[ \langle \rho(x) \rho(x') \rangle_p = \rho^M(x) \rho^M(x') + \delta_D(x - x') \rho^M(x'). \tag{69} \]

This equation yields

\[ \delta_D(x - x') = 1 + \frac{\langle \rho(x) \rho(x') \rangle_p}{\rho(x) \rho(x')} - \delta_D(x - x') \frac{1}{\rho(x)}. \tag{70} \]

Since \(\rho(x)\) is not subject to a Poisson process, the second term of the right-hand side of equation (70) can be rewritten as \(\langle [\rho(x) / \rho(x)] \rho(x) / \rho(x) \rangle_p\). Using equation (44), we have

\[ \frac{\rho(x)}{\bar{\rho}(x)} = \sum_{i=1}^{N_g} \frac{1}{\bar{\rho}(x)} w_i \delta_D(x - x_i), \tag{71} \]

in which the factor \(\bar{\rho}(x)\) can be absorbed into the weight factors, \(w_i\). The WFC covariance is given by

\[ \langle \epsilon_{j,l,i} \epsilon_{j',l',i} \rangle = \langle \epsilon_{j,l,i}^2 \rangle - \frac{\langle \psi_{j,l,i}(x) \psi_{j',l',i}(x) \rangle}{\bar{\rho}(x)} dx. \tag{72} \]
The first term in the right-hand side of equation (70) disappears, since all the basis functions $\psi_{J,l}(x)$ are admissible (eq. [7]).

5.2. The Estimators for the DWT Band Power Spectra

If the selection function varies slowly on a scale $j$, i.e.,

$$\frac{d \ln \rho(x)}{dx} \leq \frac{2^j}{L},$$  

(73)

we have approximately

$$\int \frac{\psi^{*}_{J,l}(x)\psi^{*}_{J,l'}(x)}{\tilde{\rho}(x)} dx = \frac{1}{\tilde{\rho}(x)} \delta_{j,l} \delta_{l,l'},$$  

(74)

where $\tilde{\rho}(x)$ is the number density of galaxies averaged over a volume of $L/2^j$ at $l$. In this case, the band power spectrum can be simplified as

$$P_j = \frac{1}{2} \sum_{i=0}^{2^j-1} \langle \langle \tilde{e}_{J,l}^{\theta} \tilde{e}_{J,l'}^{\theta} \rangle \rangle_p - \frac{1}{2} \sum_{l=0}^{2^{j-1}} \frac{1}{\tilde{\rho}(x)}.$$  

(75)

The second term in the right-hand side is the variance from the Poisson process. Since the Poisson process does not change the ergodicity, the average over $l$ in equation (75) is already a fair estimation for the ensemble average. Therefore, one can drop $\langle \langle \ldots \rangle \rangle_p$ in equation (75), and the estimation of the DWT band power spectrum is given by

$$P_j = \frac{1}{2} \sum_{i=0}^{2^j-1} \langle \langle \tilde{e}_{J,l}^{\theta} \tilde{e}_{J,l'}^{\theta} \rangle \rangle_p - \frac{1}{2} \sum_{l=0}^{2^{j-1}} \frac{1}{\tilde{\rho}(x)}.$$  

(76)

The second term is to subtract the contribution of the discreteness effect (or shot noise) in the Poisson sampling from the power spectrum. Here $P_j$ is debiased from the Poisson process.

5.3. The Estimators for the Scale-Scale Corrections

Similarly, one can calculate the debiased scale-scale correlations from a galaxy sample, $\rho^0(x)$. From equation (70), the term of the Poisson process is free from scale-scale correlation, and the second-order scale-scale correlation can be calculated from the WFCs of the galaxy distribution without the correction for the shot noise,

$$C_{j,l}(\Delta l) = \frac{1}{2} \sum_{l=0}^{2^{j-1}} \langle \langle \tilde{e}_{J,l}^{\theta} \tilde{e}_{J,l+\Delta l}^{\theta} \rangle \rangle_p, \quad j > l.$$  

(77)

However, the Poisson process is not free from higher order scale-scale correlations. For instance, to estimate the band-band correlations (eq. [42]), we use equation (67) with $n = 4$. This gives

$$C^2_{J,j} = \frac{1}{2} \left[ \sum_{l=0}^{2^{j-1}} \langle \langle \tilde{e}_{J,l}^{\theta} \rangle \tilde{e}_{J,l'}^{\theta} \rangle_p^2 \right]$$

$$- 2 \sum_{l=0}^{2^{j-1}} \int \frac{\psi_{J,l}(x)\psi_{J,l'}(x)}{\tilde{\rho}(x)} dx \int \frac{\psi_{J,l}(x)\psi_{J,l'}(x)}{\tilde{\rho}(x')} dx'$$

$$- \sum_{l=0}^{2^{j-1}} \int \frac{\psi_{J,l}(x)}{\tilde{\rho}(x)} dx \int \frac{\psi_{J,l}(x)}{\tilde{\rho}(x')} dx'$$

$$- \sum_{l=0}^{2^{j-1}} \int \frac{\psi_{J,l}(x)}{\tilde{\rho}(x')} dx \int \frac{\psi_{J,l}(x)}{\tilde{\rho}(x)} dx.$$

(78)

where $j > j'$ and $l' = \text{mod}(l/2^{j-l}) + \Delta l$. The last three terms are the scale-scale correlations $C^2_{J,j}$ from the Poisson sampling. Precisely, the factor $\rho(x)$ in the Poisson terms should be $\rho^0(x) = \rho(x)[1 + \delta(x)]$, but we ignore the contributions of $\delta(x)$ for the moment.

If the selection function is slowly varying on scales $j$ and $j'$ (eq. [73]), we have

$$C^2_{j,j'} = \frac{1}{2} \left[ \sum_{l=0}^{2^{j-1}} \langle \langle \tilde{e}_{J,l}^{\theta} \tilde{e}_{J,l'}^{\theta} \rangle \rangle_p^2 \right]$$

$$- \sum_{l=0}^{2^{j-1}} \int \frac{\psi_{J,l}(x)\psi^{*}_{J,l'}(x)}{\tilde{\rho}(x)} dx \int \frac{\psi_{J,l}(x)\psi^{*}_{J,l'}(x)}{\tilde{\rho}(x')} dx'$$

(79)

The second and third terms correct for the shot noise on the fourth order. Numerical results show that for typical samples of galaxy surveys, the local ($l' = l$) scale-scale correlation of the Poisson sampling is significant on small scales (Feng et al. 2000).

6. DISCUSSIONS AND CONCLUSIONS

We have presented a method of extracting the band power spectrum from observed data and simulation samples via a DWT multiresolution decomposition. The DWT scale-by-scale approach provides a physical insight into the covariance matrix of the cosmic mass field.

A key indicator of the DWT power spectrum estimator is the scale-scale and/or the band-band correlations, which can be calculated directly from the DWT covariance and the WFCs. In the scale range over which the scale-scale correlations are negligible, the DWT covariance is $j$ (scale) diagonal, and it is already an information-lossless estimation of a banded power spectrum, $P_j$. This DWT band power spectrum is optimized in the sense that the spatial resolution is automatically adaptive to the scales of the density perturbations.

In the scale range over which the scale-scale (or band-band) correlations are significant, the diagonalization of the covariance may not yield an accurate power spectrum, but may be seriously contaminated by errors. In this case, an effective confrontation between the observed sample and model predictions may be given not by a full diagonalized covariance, but by both the DWT power spectrum and scale-scale correlations. With the DWT representation, one can calculate the scale-scale correlation as well as the DWT power spectrum. Therefore, the DWT covariance is also useful when scale-scale correlations are strong.

In summary, the basic DWT algorithm proceeds in the following steps:

1. Calculate the WFCs $\tilde{e}_{J,l}$ and/or the SFCs $e_{J,l}$ from the data $\rho^0(x)$, where $J$ corresponds to the highest resolution of the samples.
2. Calculate the WFCs $\tilde{e}_{J,l}$ for various scales $j$.
3. Calculate the band power spectrum, $P_j$ and scale-scale correlations, $C_{j,j'}$.
4. In the $j$ range of $C_{j,j'} \approx 0$, test models or constrain parameters by comparing the model-predicted DWT band power spectrum, $P_j$, with observed results.
5. In the $j$ range of $C_{j,j'} \neq 0$, test models or constrain parameters by comparing the model-predicted DWT band power spectrum and scale-scale correlations with observed results.
Since the DWT is computationally powerful, the algorithm given above is found to be numerically efficient and flexible. Moreover, the method developed is open in the sense that, based on the WFCs and SFCs, one can add subsequent items to realize further goals related to power-spectrum measurement and model discrimination. Some of these problems are discussed below.

6.1. Higher Dimensions and Complex Geometry

The DWT analysis in a two- and/or three-dimensional space \( x \) can be performed on the bases of the one-dimensional direct product, i.e.,

\[
\psi_{(j_1,j_2,j_3),l_1,l_2,l_3}(x_1, x_2, x_3) = \psi_{j_1,l_1}(x_1)\psi_{j_2,l_2}(x_2)\psi_{j_3,l_3}(x_3) .
\]

(80)

In this case, the three scales \((j_1, j_2, j_3)\) of the WFCs can be different for different directions. One can define radial scales by

\[
k = 2\pi \left[ \left( \frac{2^{j_1}}{L_1} \right)^2 + \left( \frac{2^{j_2}}{L_2} \right)^2 + \left( \frac{2^{j_3}}{L_3} \right)^2 \right]^{1/2},
\]

(81)

where \(L_1 \times L_2 \times L_3\) is the three-dimensional box.

For two- and three-dimensional samples, one can also decompose by the mixed direct product of one-dimensional wavelets and scaling functions. For instance, a three-dimensional sample can be decomposed by the basis

\[
\psi_{(1,2),l_1,l_2,l_3}(x_1, x_2, x_3) = \phi_{j_1,l_1}(x_1)\psi_{j_2,l_2}(x_2)\psi_{j_3,l_3}(x_3) ,
\]

(82)

where the scaling functions \(\phi_{j,l}\) actually play the role of chopping a three-dimensional sample into \(2^j\)-dimensional slices in the \(x_1\) direction, \(l = 0, \ldots, 2^j - 1\). Like binning by the scaling function (§ 4.2), the chopping equation (82) will not cause spurious features.

The problem of samples with complex geometry can be treated by using the locality of the \(\psi_{j,l}\) (Pando & Fang 1998a). The locality property allows the WFCs to be independent of the data outside an “influence” cone. The WFCs \(\tilde{\epsilon}_{j,l}\) are only determined by data in the interval \([IL_1/2^{j+1} - \Delta x, IL_1/2^{j+1} + \Delta x]\) \([IL_2/2^{j+1} + \Delta x, IL_2/2^{j+1} + \Delta x]\), where \(\Delta x\) is the width of the basic wavelet, \(\psi\). With this property, any complex geometry of samples can be regularized into a two- or three-dimensional box by zero padding in the field between the sample geometry and the box. Since all WFCs at the zero-padding zone are zero, one can use the DWT to analyze the regular box, but not to treat the WFCs related to the zero padding as the variables of valid degrees of freedom.²

6.2. Non-Gaussianity and Power Spectrum Detection

We have emphasized that information on the non-Gaussian features is important for a precise detection of the power spectrum, or band power spectrum. This is because from the covariance, one can only find statistically uncorrelated (or statistical orthogonal) bases or modes to second order. For non-Gaussian fields, the modes statistically uncorrelated at second order might be statistically correlated at the third and fourth orders. On the other hand, the power spectrum is of second order, and therefore the power spectrum estimates at different scales might not be statistically uncorrelated if there are third- and fourth-order correlations. The accuracy of a power-spectrum estimation is affected by the higher order statistical correlations.

For instance, one popular bias model for galaxy formation employs the selection probability functions as (Cole et al. 1998)

\[
P[\delta(r)] \propto \exp \left[ \frac{\alpha \delta(r)}{\sigma} \right],
\]

(83)

where \(\alpha\) is constant, and \(\delta(r)\) and \(\sigma\) are the smoothed density field and variance, respectively. Therefore, if the density field is Gaussian, the galaxy distribution given by the Poisson sampling with the intensity equation (83) will be lognormal. The baryonic distribution is sometimes also modeled by a lognormal relation with the underlying Gaussian mass field (Bi, Ge, & Fang 1995; Bi & Davidsen 1997). As is well known, for a lognormal distribution, the most likely value can be significantly different from their mean value. In this case, to estimate the accuracy of a power spectrum detection, a higher order cumulant statistics is needed.

In the DWT analysis, the \(2^j\) WFCs give the one-point distribution of the fluctuations on scale \(j\). Therefore, the third and forth cumulants can be calculated by

\[
S_j = \frac{1}{P_j^{1/2}} \frac{1}{2^j} \left[ \sum_{l=0}^{2j-1} (\tilde{\epsilon}_{j,l} - \bar{\tilde{\epsilon}}_{j,l})^3 \right] ,
\]

(84)

\[
K_j = \frac{1}{P_j^{3/2}} \frac{1}{2^j} \left[ \sum_{l=0}^{2j-1} (\tilde{\epsilon}_{j,l} - \bar{\tilde{\epsilon}}_{j,l})^4 - 3 \right] .
\]

(85)

These are, respectively, the skewness and kurtosis spectra. It is not difficult to generalize equations (84) and (85) to higher orders.

6.3. Selection of the Basis of the Multiresolution Analysis

In computing the samples of redshift surveys, two coordinate systems have been widely used: (1) the parallel-plane system and (2) the spherical-shell system. For system 1, the volume of the survey can be approximated as a box, and therefore the wavelets of equations (80) and (82) are suitable for the decomposition. For system 2, we should use the wavelets on a two-dimensional spherical surface. With the development of the DWT analysis, the bank of DWT analyses has stored more and more sets of the orthogonal and complete basis for the multiresolution decomposition of different geometries. Multiscale analysis on geometry beyond the two simple cases given above is becoming feasible.

6.4. Systematic Effects

The influence of various systematic effects on the power spectrum detection has only been studied very preliminarily. The linear effect of redshift distortion on the power spectrum detection has been well studied (e.g., Hamilton 1998). It is not difficult to incorporate the linear theory of the redshift distortion with the DWT analysis. A key operator for mapping a real-space distribution into redshift space is \([1 - a(\partial^2/\partial z^2)V^{-2}]\), where the coefficient \(a\) is constant. To diagonalize this differential-integral operator, the

² Regarding DWT on manifolds, see also W. Sweldens, http://cm.bell-labs.com/who/wim or http://www.wavelet.org.
Fourier representation is certainly the best. However, it has been shown that this operator is quasi-diagonal in the DWT representation (Farge et al. 1996).

Moreover, it would be straightforward to include a scale-dependent bias in the DWT representation. The redshift distortion is usually calculated under the assumption that the galaxy distribution, $\rho(x)$, is linearly related to the underlying mass field, $\rho(x)$, i.e., $\rho(r) = b\rho(r)$, where $b$ is the bias parameter. However, observations have indicated that the bias parameters are probably scale-dependent (Fang, Deng, & Xia 1998). It is easy to introduce scale-dependent bias in the DWT representation. For instance, one can define a bias parameter on a given scale by $\hat{\epsilon}_{j,l} = b_j \hat{\epsilon}_{j,l}$.

L. L. F. acknowledges support from the National Science Foundation of China (NSFC) and World Laboratory Scholarship. This project was done during L. L. F.’s visiting to the Department of Physics, University of Arizona. This work was supported in part by the LWL foundation. We thank the anonymous referee for helpful comments.

APPENDIX

THE DISCRETE WAVELET TRANSFORM (DWT) OF DENSITY FIELDS

Let us briefly introduce the DWT analysis of the cosmic mass density fields; for mathematical details, we refer the reader to the classic papers by Mallat (1989a, 1989b, 1989c), Meyer (1992), Daubechies (1992), and references therein, and for physical applications, to Fang & Thews (1998) and references therein. Some other cosmological applications of wavelets can also be found in, e.g., Pando, Vills-Gabaud, & Fang (1998b), Hobson, Jones, & Lasenby (1999), Sanz et al. (1999), Tenorio et al. (1999), Xu, Fang, & Wu (2000), and Cayon et al (2000).

A1. EXPANSION BY SCALING FUNCTIONS

We consider here a one-dimensional mass-density distribution, $\rho(x)$, or contrast, $\delta(x) = [\rho(x) - \bar{\rho}] / \bar{\rho}$, which are mathematically random fields over a spatial range $0 \leq x \leq L$. It is not difficult to extend all results developed in this section into two and three dimensions, because the DWT bases for higher dimension can be constructed from a direct product of one-dimensional bases.

First, we introduce the scaling functions for the Haar wavelets. These are top-hat window functions defined by

$$\phi_{j,l}^H(x) = \begin{cases} 1, & \text{for } L2^{-j} \leq x \leq L(l + 1)2^{-j}, \\ 0, & \text{otherwise}, \end{cases}$$

(A1)

where the superscript “$H$” denotes Haar. The scaling function, $\phi_{j,l}^H(x)$, actually gives a window at resolution scale $L/2^l$ and position $L2^{-j} \leq x \leq L(l + 1)2^{-j}$. With the scaling function, the mean of the density contrast distribution in the spatial range $L2^{-j} \leq x \leq L(l + 1)2^{-j}$ can be expressed as

$$\epsilon_{j,l} = \frac{2^j}{L} \int_0^L \delta(x) \phi_{j,l}^H(x) \, dx.$$  

(A2)

The number $\epsilon_{j,l}$ is called the scaling function coefficient (SFC). Using SFCs, one can construct a density contrast field as

$$\delta(x) = \sum_{l=0}^{2^j-1} \epsilon_{j,l} \phi_{j,l}^H(x).$$

(A3)

This is the density contrast, $\delta(x)$, smoothed on scales $L/2^l$, or simply on a $j$ scale.

The scaling function $\phi_{j,l}^H(x)$ can be rewritten as

$$\phi_{j,l}^H(x) = \phi^H\left(\frac{2^j x}{L} - l\right),$$

(A4)

where

$$\phi^H(\eta) = \begin{cases} 1, & \text{for } 0 \leq \eta \leq 1, \\ 0, & \text{otherwise}, \end{cases}$$

(A5)

where $j$ and $l$ are integers, with $j \geq 0$ and $0 \leq l \leq 2^j - 1$. Here $\phi^H(\eta)$ is called the basic scaling function. The scaling function $\phi_{j,l}^H(x)$ is thus a translation and dilation of the basic scaling function.

The functions $\phi_{j,l}^H(x)$ are orthogonal with respect to $l$, i.e.,

$$\int_0^L \phi_{j,l}^H(x) \phi_{j,l'}^H(x) dx = \frac{L}{2^j} \delta_{l,l'},$$

(A6)

where $\delta_{l,l'}$ is the Kronecker delta function. Thus, equation (A3) gives functions in the function space $V_j$ spanned by bases.
The Haar wavelets are orthogonal with respect to
\[ \phi_{j,t}^H(x) = \phi_{j+1,2t}^H(x) + \phi_{j+1,2t+1}^H(x) , \quad (A7) \]
\[ \epsilon_{j,t} = \frac{1}{2}(\epsilon_{j+1,2t} + \epsilon_{j+1,2t+1}) . \quad (A8) \]

Therefore, \( V_j \subset V_{j+1} \) for all \( j \). Thus, the orthogonal projectors \( P_j \) onto \( V_j \), i.e., \( P_j f \in V_j \), satisfy
\[ \lim_{j \to -\infty} P_j f = f \quad (A9) \]
for all \( f \in L_2(R) \). A multiresolution analysis is then defined by the sequence of subspaces \( V_j \).

### A2. Expansion by Wavelets

Equations (A7) and (A8) show that \( \delta_j(x) \) contains less information than \( \delta^{j+1}_j(x) \), because information on the scale \( j + 1 \) has been smoothed out by equation (A8). It would be nice not to lose any information during the smoothing from \( j + 1 \) to \( j \) (eq. [A8]). This can be accomplished if the differences, \( \delta^{j+1}_j(x) - \delta_j(x) \), between the smoothed distributions on succeeding scales are somehow retained. That is, if we are able to retain these differences, this scheme will then make it possible to smooth the distribution and yet not lose any information as a result of the smoothing.

To calculate the differences, we define the difference function, or wavelet, as
\[ \psi^H(\eta) = \begin{cases} 1, & \text{for } 0 \leq \eta < \frac{1}{2}, \\ -1, & \text{for } \frac{1}{2} \leq \eta \leq 1, \\ 0, & \text{otherwise}. \end{cases} \quad (A10) \]

This is the basic Haar wavelet. As with the scaling functions, one can construct a set of wavelets \( \psi_{j,t}^H(x) \) by dilating and translating equation (A10) as
\[ \psi_{j,t}^H(x) = \psi^H\left(\frac{2^j x}{L} - t\right). \quad (A11) \]

The Haar wavelets are orthogonal with respect to both indexes \( j \) and \( l \), i.e.,
\[ \int_0^L \phi_{j,t}^H(x) \psi_{j',l}^H(x) dx = \left(\frac{L}{2}\right) \delta_{j,j'} \delta_{l,l'} . \quad (A12) \]

For a given \( j \), \( \psi_{j,t}^H(x) \) is also orthogonal to the scaling functions \( \phi_{j',l}^H(x) \) with \( j' \leq j \), i.e.,
\[ \int_0^L \phi_{j,t}^H(x) \psi_{j',l}^H(x) dx = 0, \quad \text{if } j' \leq j . \quad (A13) \]

From equations (A4) and (A11), we have
\[ \phi_{j,2t}^H(x) = \frac{1}{2}\left[\psi_{j-1,t}^H(x) + \psi_{j-1,t}^H(x)\right], \quad (A14) \]

\[ \phi_{j,2t+1}^H(x) = \frac{1}{2}\left[\phi_{j-1,t}^H(x) - \psi_{j-1,t}^H(x)\right]. \]

Thus, the difference \( \delta^{j+1}_j(x) - \delta^j(x) \) is given by
\[ \delta^{j+1}_j(x) - \delta^j(x) = \sum_{i=0}^{j-1} \tilde{\epsilon}_{j,i} \psi_{j-1,i}^H(x) , \quad (A15) \]

where \( \tilde{\epsilon}_{j-1,i} \) are called the wavelet function coefficients (WFCs), which are given by
\[ \tilde{\epsilon}_{j,i} = \frac{2^j}{L} \int_0^L \delta(x) \psi_{j,i}^H(x) dx . \quad (A16) \]

Using equation (A15) repeatedly, we have
\[ \delta^j(x) = \delta^0(x) + \sum_{j=0}^{j-1} \tilde{\epsilon}_{j,i} \psi_{j-1,i}^H(x) . \quad (A17) \]

This is an expansion of the function \( \delta^j(x) \) with respect to the basis \( \psi_{j,t}^H(x) \), and \( \delta^0(x) \) is the mean of \( \delta(x) \) in the range \( L \). We have \( \delta^0(x) = 0 \) if \( \delta(x) \) is the density contrast. Considering equation (A9), for any \( f(x) \in L^2(R) \) in \( L \) with mean \( f = 0 \), we have
\[ f(x) = \sum_{j=0}^{\infty} \sum_{i=0}^{2^j-1} \tilde{\epsilon}_{j,i} \psi_{j,i}^H(x) \quad (A18) \]

and
\[ \tilde{\epsilon}_{j,i} = \frac{2^j}{L} \int_0^L f(x) \psi_{j,i}^H(x) dx . \quad (A19) \]
For a given \( j \), the wavelets \( \psi_{j,l}(x) \) form a space \( W_j \) that is the orthogonal complements of \( V_j \) in \( V_{j+1} \), i.e., \( V_{j+1} = V_j \oplus W_j \). Thus, every \( f^j \in V_j \) has a unique decomposition \( f^j = f^{j-1} + d^{j-1} \), with \( f^{j-1} \in V_{j-1} \) and \( d^{j-1} \in W_{j-1} \). Since \( W_j \subset V_{j+1} \), and \( W_j \) is orthogonal to \( V_j \), \( W_j \) is also orthogonal to \( W_{j-1} \) and \( W_{j+1} \). Thus, all the spaces \( W_j \) are mutually orthogonal. Since \( V_j \) contains only \( W_j \) with \( j < j \), \( V_j \) is orthogonal to all \( W_j \) with \( j \geq j \).

A3. COMPACTLY SUPPORTED ORTHOGONAL BASIS

In terms of the subspace \( V_j \), the basic scaling function \( \phi(\eta) \) and basic \( \psi(\eta) \) belong to \( V_0 \) and \( W_0 \), respectively, and they can be expressed by the basis of \( V_1 \), \( \phi(2\eta - l) \), i.e.,

\[
\phi(\eta) = \sum_{l=-\infty}^{\infty} a_l \phi(2\eta - l), \tag{A20}
\]

\[
\psi(\eta) = \sum_{l=-\infty}^{\infty} b_l \phi(2\eta - l), \tag{A21}
\]

where \( a_l \) and \( b_l \) are called the filter coefficients.

If we require that the scaling function \( \phi(\eta) \) is normalized, equation (A21) yields

\[
\sum_l a_l = 2. \tag{A22}
\]

Requiring orthogonality for \( \phi(x) \) with respect to discrete integer translations, i.e.,

\[
\int_{-\infty}^{\infty} \phi(\eta - m)\phi(\eta) d\eta = \delta_{m,0}, \tag{A23}
\]

we have

\[
\sum_l a_l a_{l+2m} = 2\delta_{0,m}. \tag{A24}
\]

The orthogonality between \( \phi \) and \( \psi \) means

\[
\int_{-\infty}^{\infty} \psi(\eta)\phi(\eta - l) d\eta = 0. \tag{A25}
\]

Therefore, one has

\[
b_l = (-1)^l a_{l-1}. \tag{A26}
\]

Furthermore, the wavelet \( \psi(\eta) \) must be admissible,

\[
\int_{-\infty}^{+\infty} \psi(\eta) d\eta = 0, \tag{A27}
\]

so we need

\[
\sum_l b_l = 0. \tag{A28}
\]

Equations (A22), (A24), (A26), and (A28) for the filter coefficients were employed to construct families of scaling functions and wavelets. The simplest solution of the filter coefficients is \( a_0 = a_1 = b_0 = -b_1 = 1 \), and all others 0. This solution gives the Haar wavelet. After the Haar wavelet, the simplest solution for the filter coefficients is

\[
a_0 = 1 + \frac{\sqrt{3}}{4}, \quad a_1 = \frac{3 + \sqrt{3}}{4}, \tag{A29}
\]

\[
a_2 = \frac{3 - \sqrt{3}}{4}, \quad a_3 = \frac{1 - \sqrt{3}}{4}.
\]

This is the Daubechies 4-wavelet (D4). It is compactly supported and continuous.

With these wavelets, the multiresolution analysis can be performed in a similar way as developed in last two sections for the Haar wavelets. The scaling functions and wavelets for spanning the subspace \( V_j \) and \( W_j \) are given, respectively, by a translation and dilation of the basic scaling function and basic wavelet,

\[
\phi_{j,l}(x) = \left( \frac{2^j}{L} \right)^{1/2} \phi \left( \frac{2^j}{L} x - l \right) \tag{A30}
\]

and

\[
\psi_{j,l}(x) = \left( \frac{2^j}{L} \right)^{1/2} \psi \left( \frac{2^j}{L} x - l \right). \tag{A31}
\]
The wavelets are orthonormal, i.e.,
\[
\int \psi_{j,l,x}(x) \psi_{j',l'}(x) \, dx = \delta_{j,j'} \delta_{l,l'} .
\]  
(A32)

Equations (A23) and (A25) also yield
\[
\int \phi_{j,l,x}(x) \phi_{j',l'}(x) \, dx = \delta_{l,l'}
\]  
(A33)

and
\[
\int \phi_{j,l,x}(x) \psi_{j',l'}(x) \, dx = 0 , \quad j' \geq j .
\]  
(A34)

The set of \( \psi_{j,l} \) and \( \phi_{j,m}(x) \) with \( 0 \leq j < \infty \) and \( -\infty < l, m < \infty \) form a complete, orthonormal basis in the space of functions with period length \( L \).

Thus, a density field \( \rho(x) \) with period length \( L \) can be expanded as (Fang & Thews 1998)
\[
\rho(x) = \bar{\rho} + \tilde{\rho} \sum_{j=0}^{\infty} \sum_{l=-\infty}^{\infty} \tilde{c}_{j,l} \psi_{j,l}(x) ,
\]  
(A35)

or the density contrast, \( \delta(x) = [\rho(x) - \bar{\rho}] / \bar{\rho} \), is
\[
\delta(x) = \sum_{j=0}^{\infty} \sum_{l=-\infty}^{\infty} \tilde{c}_{j,l} \psi_{j,l}(x) ,
\]  
(A36)

where
\[
\bar{\rho} = L^{-1} \int_{0}^{L} \rho(x) \, dx
\]  
(A37)

and
\[
\tilde{c}_{j,l} = \int_{-\infty}^{\infty} \delta(x) \psi_{j,l}(x) \, dx .
\]  
(A38)

More generally, we have
\[
\rho(x) = \rho'(x) + \tilde{\rho} \sum_{j=0}^{\infty} \sum_{l=-\infty}^{\infty} \tilde{c}_{j,l} \psi_{j,l}(x) ,
\]  
(A39)

where \( \rho'(x) \) is the density field smoothed on a scale \( J \),
\[
\rho'(x) = \sum_{l=-\infty}^{+\infty} \epsilon_{j,l} \phi_{j,l}(x) ,
\]  
(A40)

and the SFC \( \epsilon_{j,l} \) is given by
\[
\epsilon_{j,l} = \int_{-\infty}^{+\infty} \rho(x) \phi_{j,l}(x) \, dx .
\]  
(A41)
Pando, J., & Fang, L. Z. 1998a, Phys. Rev. E, 57, 3593
———. 1998b, A&A, 340, 335
Pando, J., Lipa, P., Greiner, M., & Fang, L. Z. 1998a, ApJ, 496, 9
Pando, J., Valls-Gabaud, D., & Fang, L. Z. 1998b, Phys. Rev. Lett., 81, 4568
Peebles, P. J. E. 1980, The Large-Scale Structure of the Universe (Princeton: Princeton Univ. Press)
Percival, D. B., & Walden, A. T. 1993, Spectral Analysis for Physical Applications (Cambridge: Cambridge Univ. Press)
Sanz, J. L., et al. 1999, MNRAS, 309, 672
Scoccimarro, R., Zaldarriaga, M., & Hui, L. 1999, ApJ, 527, 1
Tegmark, M., Hamilton, A. J. S., Strauss, M. A., Vogeley, M. S., & Szalay, A. S. 1998, ApJ, 499, 555
Tenorio, L., Jaffe, A. H., Hanany, S., & Lineweaver, C. H. 1999, MNRAS, 310, 823
Weinberg, D. H. 1992, MNRAS, 254, 315
Xu, W., Fang, L. Z., & Wu, X. P. 2000, ApJ, 532, 728