To appear in *ApJ*, **499** (1998)
Preprint typeset using LaTeX style emulateapj

MEASURING THE GALAXY POWER SPECTRUM WITH FUTURE REDSHIFT SURVEYS

Max Tegmark[1]
Institute for Advanced Study, Princeton, NJ 08540; max@ias.edu
Andrew J. S. Hamilton
JILA and Dept. of Astrophysical, Planetary and Atmospheric Sciences,
Box 440, Univ. of Colorado, Boulder, CO 80309, USA; ajsh@dark.colorado.edu
Michael A. Strauss[2,3]
Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544; strauss@astro.princeton.edu
Michael S. Vogeley[1]
Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544; vogeley@astro.princeton.edu
and
Alexander S. Szalay
Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218; szalay@pha.jhu.edu

To appear in ApJ, **499** *(1998)*

ABSTRACT

Precision measurements of the galaxy power spectrum $P(k)$ require a data analysis pipeline that is both fast enough to be computationally feasible and accurate enough to take full advantage of high-quality data. We present a rigorous discussion of different methods of power spectrum estimation, with emphasis on the traditional Fourier method, and linear (Karhunen-Loève; KL), and quadratic data compression schemes, showing in what approximations they give the same result. To improve speed, we show how many of the advantages of KL data compression and power spectrum estimation may be achieved with a computationally faster quadratic method. To improve accuracy, we derive analytic expressions for handling the integral constraint, since it is crucial that finite volume effects are accurately corrected for on scales comparable to the depth of the survey. We also show that for the KL and quadratic techniques, multiple constraints can be included via simple matrix operations, thereby rendering the results less sensitive to galactic extinction and mis-estimates of the radial selection function. We present a data analysis pipeline that we argue does justice to the increases in both quality and quantity of data that upcoming redshift surveys will provide. It uses three analysis techniques in conjunction: a traditional Fourier approach on small scales, a pixelized quadratic matrix method on large scales and a pixelized KL eigenmode analysis to probe anisotropic effects such as redshift-space distortions.

1. INTRODUCTION

Observational data on galaxy clustering are rapidly increasing in both quantity and quality, which brings new challenges to data analysis. As for quantity, redshifts had been published for a few thousand galaxies 15 years ago. Today the number is $\sim 10^5$ (Huchra, private communication), and ongoing projects such as the AAT 2dF Survey and the Sloan Digital Sky Survey (hereafter SDSS; see Gunn & Weinberg 1995) will raise it to 10^6 in a few years. Comprehensive reviews of past redshift surveys are given by Efstathiou (1994), Vogeley (1995), Strauss & Willick (1995) and Strauss (1997), the last also including a detailed description of 2dF and SDSS. As for quality, more accurate and uniform photometric selection criteria (enabled by *e.g.* the well-calibrated 5-band photometry of the SDSS) reduce potential systematic errors. This increased data quality makes it desirable to avoid approximations in the data analysis process and to use methods that can constrain cosmological quantities as accurately as possible. This is especially important since a wide variety of models currently appear to be at least marginally consistent with the present data (Peacock 1997; White *et al.* 1996; Vogeley 1997), so smaller error bars on the power spectrum will be needed to discriminate between them. However, the increased data quantity makes it a real challenge to perform such an accurate analysis; as we will discuss at some length, a straightforward application of any such method is computationally unfeasible for data sets as large as those from 2dF and SDSS.

This paper addresses both the accuracy and speed issues for measuring the galaxy power spectrum $P(k)$. For the highest accuracy, we advocate the use of lossless pixelized data compression methods, both linear (Karhunen-Loève) and quadratic. To improve the speed, we present a fast implementation of a pixelized quadratic power estimation method, and show how it can reproduce Karhunen-Loève results (Vogeley & Szalay 1996; Tegmark, Taylor & Heavens 1997 — hereafter TTH) exactly, without the need to solve eigenvalue problems, although it does not have the ability to measure redshift-space distortions.

The rest of the paper is organized as follows. Section 2 is a "buyers guide", where we list the various properties that are important when deciding which data analysis method

[1]Hubble Fellow.

[2]Alfred P. Sloan Foundation Fellow.

[3]Cottrell Scholar of Research Corporation.

to use. In Section 3, we review the existing methods for power spectrum estimation in a common framework, make various extensions of them and present the fast quadratic data compression and power spectrum technique. Section 4 describes how the various methods are interrelated, and Section 5 discusses how they can be immunized from various systematic effects, such as errors in the extinction model; many of the technical details are given in Appendices A and B. Section 6 discusses the pros and cons of the different techniques. We conclude that to meet all criteria on the wish list of Section 2, it is necessary to combine three of the principal methods described. The data analysis pipeline that we propose is summarized in Figure 1, and the reader may wish to glance at this before delving into the details of Section 3, to obtain an overview of how everything fits together. We ignore redshift-space distortions for most of this paper, and assume that clustering is isotropic. In Section 5 and Appendix C, we discuss this issue further.

2. COMPARING METHODS: A WISH LIST

Since the ultimate goal of the analysis of clustering in a galaxy redshift survey is to constrain cosmological models, we want a method that minimizes the statistical error bars on cosmological parameters[4] and is robust against potential systematic errors. We will now discuss the former issue in some detail, and return to the latter in Section 5 and Appendix B.

2.1. *The problem to be solved*

The data set from a galaxy redshift survey consists of N three-dimensional vectors $\mathbf{r}_\alpha$ ($\alpha=1,...,N$) giving the measured positions of the galaxies. Following Peebles (1973, 1980), we model these positions as generated by a random Poissonian point process where the galaxy density is modulated by both selection effects and fluctuations in the underlying matter distribution. The former are described by $\bar{n}(\mathbf{r})$, the *selection function* of the galaxy survey under consideration, defined as the expected galaxy density. Thus $\bar{n}(\mathbf{r})dV$ is the expected (not the observed) number of galaxies in a volume dV about $\mathbf{r}$ in the absence of clustering. This function typically falls off at large distances, and can exhibit small angular variations due to extinction. It vanishes outside the survey volume. The fluctuations in the underlying matter density are given by the field $\delta_r(\mathbf{r})$, which is not to be confused with the Dirac delta function δ^D. This means that the observed galaxy distribution $n(\mathbf{r})=\sum_\alpha\delta^D(\mathbf{r}-\mathbf{r}_\alpha)$ is modeled as a 3D Poisson process with intensity $\lambda(\mathbf{r})=\bar{n}(\mathbf{r})[1+\delta_r(\mathbf{r})]$.

The density fluctuations δ_r are modeled as a homogeneous and isotropic (but not necessarily Gaussian) random field. This implies that the Fourier transform $\widehat{\delta}_r$ of the density fluctuation field obeys the simple relation

$$\langle\widehat{\delta}_r(\mathbf{k})^*\widehat{\delta}_r(\mathbf{k}')\rangle=(2\pi)^3\delta^D(\mathbf{k}-\mathbf{k}')P(k) \qquad (1)$$

for some function P which depends only on the magnitude of $\mathbf{k}$, not on its direction[5] P is known as the *power spectrum*. Because of equation (1), P contains all the information needed to compute any statistical quantities that depend quadratically on δ_r, for instance the variance or the correlation function on different scales. Moreover, if the random field δ_r is *Gaussian* (which means that the joint probability distribution of δ_r at any number of points is a multivariate Gaussian), then P characterizes δ_r *completely* and contains all the information needed to answer any statistical question whatsoever about δ_r. Inflationary theory (e.g., Peebles 1993) and observations of large-scale structure (e.g., Strauss & Willick 1995) imply that δ_r is Gaussian on large scales, making P an important quantity in cosmology. The power spectrum estimation problem, which is the topic of this paper, is to estimate $P(k)$ given the observed realization of $n(\mathbf{r})$.

2.2. *The traditional approach*

Let us parameterize the power spectrum $P(k)$ by some set of parameters $\theta_i, i=1,2,...,$ grouped into a vector $\mathbf{\Theta}$. These may be either the band powers in a set of narrow bands, or physically motivated parameters such as the normalization σ_8, the shape parameter Γ, the primordial spectral index n, *etc.* Let us package our data set into a vector $\mathbf{x}$; much of the distinction between different methods discussed in Section 3 lies in the way this packaging is done. The standard approach to parameter estimation is to write down the expression for the probability distribution $f(\mathbf{x};\mathbf{\Theta})$. Here we interpret f as a probability distribution over $\mathbf{x}$ for a fixed $\mathbf{\Theta}$. In a Bayesian statistical analysis with a uniform prior probability distribution for $\mathbf{\Theta}$, one reinterprets f as a probability distribution over $\mathbf{\Theta}$ for a given data set $\mathbf{x}$, and to clarify this distinction renames f the likelihood function. The final results are often presented as contour plots of this likelihood function, as at the bottom of Figure 1.

If we take $\mathbf{x}$ to be the raw data set, *i.e.*, the measured coordinates $\mathbf{r}_\alpha$ ($\alpha=1,...,N$) of the N measured galaxies, then the likelihood function f is unfortunately hopeless to compute numerically, since it involves the N-point correlation function. Even in the Gaussian approximation that f is given by a product over two-point correlation functions (*e.g.*, White 1979; Fry 1984), this requires evaluating a multivariate polynomial of degree $N/2$ in the correlations of the $N(N+1)/2$ galaxy pairs, and the CPU time required for this grows faster than exponentially with N (Dodelson, Hui, & Jaffe 1997). The traditional approach has therefore been to take $\mathbf{x}$ to be something else: band-power estimates of the power spectrum. These are essentially computed by multiplying the observed density field by some weight function, Fourier transforming it, taking the squared modulus of the result and averaging over shells in k-space (Section 3.3). In the (sometimes poor) approximation that the probability distribution for $\mathbf{x}$ is a multivariate Gaussian, its probability distribution is

$$f(\mathbf{x};\mathbf{\Theta})\propto|\mathbf{C}|^{-1/2}e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^\dagger\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})}, \qquad (2)$$

[4] This is not a mere unimportant detail, since doubling the error bars (which an inferior method can easily do) is comparable to reducing the survey volume and the number of galaxies probed by a factor of four.

[5] Note that equation (1) only holds if positions are measured in real space. Redshift-space distortions couple modes of different $\mathbf{k}$; see Appendix C for more details.

where $|\mathbf{C}|$ denotes the determinant of $\mathbf{C}$, and the mean vector $\mathbf{m} \equiv \langle \mathbf{x} \rangle$ and the covariance matrix $\mathbf{C} \equiv \langle \mathbf{x}\mathbf{x}^\dagger \rangle - \mathbf{m}\mathbf{m}^\dagger$ depend on $P(k)$ and hence on the unknown parameters $\mathbf{\Theta}$. Much of the model testing to date has been rather approximate, often little more than a "χ-by-eye" fit of theoretical power spectra to the data, which is tantamount to ignoring the correlations between the power estimates (the off-diagonal elements of $\mathbf{C}$). This approach clearly does not utilize all the information present in the data, and can also bias the results.

If we had infinite computer resources, we would improve the situation by simply performing an exact brute force likelihood analysis on the raw data set. Is there a faster way of obtaining the same result?

2.3. *The notion of a lossless method*

We will call a method for analyzing a data set *unbeatable* or *optimal* if no other method can place tighter constraints on cosmological models (as parameterized by $\mathbf{\Theta}$) using this data. In this paper, we are focussed on the power spectrum measured in real space, and so we restrict ourselves to measurements of the parameters which determine the power spectrum itself.

2.3.1. *The Fisher Information Matrix*

This can be made precise using the formalism of the *Fisher information matrix* (see TTH for a comprehensive review of this application), which offers a simple and a useful way of measuring how much information each step in the pipeline of Figure 1 destroys. Given any set of cosmological parameters of interest denoted θ_i, $i = 1, 2, ...$, their Fisher matrix $\mathbf{F}$ gives the smallest error bars with which the parameters can possibly be measured from a given data set. If the probability distribution for the data set given the parameter values is $f(\mathbf{x}; \mathbf{\Theta})$, then the Fisher matrix is defined by (Fisher 1935)

$$\mathbf{F}_{ij} \equiv -\left\langle \frac{\partial^2 \ln f}{\partial\theta_i \partial\theta_j} \right\rangle. \tag{3}$$

Crudely speaking, $\mathbf{F}^{-1}$ can be thought of as the best possible covariance matrix for the measurement errors on the parameters. Indeed, the Cramér-Rao inequality (Kenney & Keeping 1951; Kendall & Stuart 1969) states that no unbiased method whatsoever can measure the i^{th} parameter with error bars (standard deviation) less than $1/\sqrt{\mathbf{F}_{ii}}$. If the other parameters are not known but are estimated from the data as well, the minimum standard deviation rises to $(\mathbf{F}^{-1})_{ii}^{1/2}$. This formalism has recently been used to assess the accuracy with which cosmological parameters can be measured from future galaxy surveys (Tegmark 1997b; Goldberg & Strauss 1998; Hu, Eisenstein, & Tegmark 1997) and cosmic microwave background experiments (Jungman *et al.* 1996; Bond, Efstathiou, & Tegmark 1997; Zaldarriaga, Seljak, & Spergel 1997).

2.3.2. *Checking for leaks in the pipeline*

By computing the Fisher matrix separately from each of the intermediate data sets in Figure 1, we can track the flow of information down the data pipeline and check for leaks. For instance, if the Fisher matrix computed from the raw positional data is identical to that computed from the (much smaller) data set consisting of the band power estimates, then the power spectrum estimation method is *lossless* in the sense that no information about the parameters of interest has been lost in the process of compressing the data set from, say, 10^6 numbers down to 50; cf., the discussion in Section 2.4. We will use this criterion when comparing different data analysis techniques below.

2.3.3. *The power spectrum Fisher matrix*

Whether a method is lossless or not generally depends on which parameters we are interested in estimating. Fortunately, as shown by Tegmark (1997a, hereafter T97), certain methods can be shown to be lossless for *any* set of parameters in a large class. An important special case are quantities that parameterize the power spectrum $P(k)$, such as σ_8 and Γ; all the information on these parameters is retained if the power spectrum itself (parameterized by the power in many narrow bands) can be measured with the minimal error bars. This means that one can test whether a method is lossless simply by computing the Fisher matrix for the band powers. This also means that band powers have a special status compared to other parameters: if we simply measure $P(k)$ as accurately as possible, this measured function will retain all the information about all cosmological parameters. All this is of course only true within the framework of Gaussian and isotropic models, which are uniquely characterized by their power spectrum. Specifically, these methods *do* lose information about parameters that affect the data set not only via $P(k)$. Important examples to which we will return are redshift-space distortions and uncorrected galactic extinction, both of which introduce differences between the angular and radial clustering patterns. On small scales, nonlinear clustering creates non-Gaussian fluctuations where in addition to $P(k)$, higher order moments also contain cosmological information.

2.4. *Data compression, simplicity and speed*

A second and rather obvious criterion for comparing data analysis methods is their numerical feasibility. For instance, the brute-force likelihood analysis of the raw data set (the galaxy positions) described above is lossless, but too time-consuming to be numerically feasible when N, the number of galaxies, is large.

When the brute-force method is unfeasible, the general approach is to perform some form of *data compression*, whereby the data set is reduced to a smaller and simpler one which is easier to analyze. If the data compression step is lossless, a brute-force analysis on the compressed data set clearly gives just as small error bars as one on the raw data would have done.

To facilitate parameter estimation further down the pipeline, it is useful if the statistical properties of the power spectrum estimates are simple and easy to calculate. The simplest case (which often occurs with the pixelized methods described below) is that where the data set is a multivariate Gaussian, described by the likelihood function of equation (2). Then the slowest step in the likelihood calculation is computing the determinant of the $N \times N$

covariance matrix $\mathbf{C}$, for which the CPU time scales as N^3, so it is desirable to make the compressed data set as small as is possible without losing information[6]. It is also convenient if the statistical properties of the compressed data set, in particular, its covariance matrix, can be computed analytically. We will see that this is the case for the Karhunen-Loève and quadratic data compression methods describe below, but not for the the maximum-likelihood method, where it requires numerical computation of the entire likelihood surface. Finally, the simplest covariance matrix one can desire is clearly one that is diagonal, *i.e.*, where the errors on the elements of the compressed data set are uncorrelated.

2.5. *The wish list*

In summary, the ideal data analysis/data compression method would

1. be lossless, at least for the parameters of interest,
2. give easy-to-compute and uncorrelated errors,
3. be computationally feasible in practice,
4. allow one to account for redshift-space distortions and systematic effects.

The first two items can be summarized by saying that we want the method to retain the cosmological information content of the original data set, distilled into a set of mutually exclusive (2) and collectively exhaustive (1) chunks. In other words, the information chunks should be independent and together retain all the information from the original data set

3. POWER SPECTRUM ESTIMATION METHODS — AN OVERVIEW

In this section, we review the various methods for power spectrum estimation that have been proposed in the literature and present various extensions. We start in Section 3.1 by developing a formalism that is common to all our approaches. In Section 3.2, we discuss how the data might be discretized; that is, different ways of packaging the data into a convenient form $\mathbf{x}$. We then discuss various methods of power spectrum estimation: traditional methods which take the square of the amplitude of the Fourier modes (Section 3.3), the method of brute force likelihood (Section 3.4), the linear Karhunen-Loève data compression method (Section 3.5), and a new quadratic data compression method (Section 3.6). Our suggested approach to power spectrum estimation involves a combination of these methods in different regimes, as summarized in Figure 1, and described in more detail in Section 7. Throughout Section 3, we ignore redshift-space distortions and other systematic effects, and assume that clustering is isotropic. We return to this topic in Section 5.

3.1. *Density field, shot noise and window functions*

With the exception of the brute force maximum likelihood technique, all of the methods described below compute band power estimates q_i that are quadratic functions of the density field n, which means that they can be written as

$$q_i = \int\int E_i(\mathbf{r},\mathbf{r}')\frac{n(\mathbf{r})}{\bar{n}(\mathbf{r})}\frac{n(\mathbf{r}')}{\bar{n}(\mathbf{r}')}d^3r\,d^3r' = \sum_{\alpha,\beta}\frac{E_i(\mathbf{r}_\alpha,\mathbf{r}_\beta)}{\bar{n}(\mathbf{r}_\alpha)\bar{n}(\mathbf{r}_\beta)} \tag{4}$$

for some real-valued symmetric pair weighting functions E_i which are designed to isolate different ranges of wavenumber k — the methods simply differ in their choices of E_i. Taking the expectation value of $n(\mathbf{r})n(\mathbf{r}')/\bar{n}(\mathbf{r})\bar{n}(\mathbf{r}')$ produces three terms: "1" from the mean density, $\delta^D(\mathbf{r}-\mathbf{r}')/\bar{n}(\mathbf{r})$ from shot noise and $\delta_r(\mathbf{r})\delta_r(\mathbf{r}')$ from density fluctuations. By a derivation analogous to FKP and T95, one finds that these three terms give

$$\langle q_i\rangle = W_i(\mathbf{0}) + b_i + \int W_i(\mathbf{k})P(k)\frac{d^3k}{(2\pi)^3}, \tag{5}$$

where W_i, the three-dimensional *window functions*, are given by

$$W_i(\mathbf{k}) = \widehat{E}_i(\mathbf{k},\mathbf{k}) \tag{6}$$

and

$$\widehat{E}_i(\mathbf{k},\mathbf{k}') \equiv \int E_i(\mathbf{r},\mathbf{r}')e^{-i\mathbf{k}\cdot\mathbf{r}}e^{i\mathbf{k}'\cdot\mathbf{r}'}d^3r\,d^3r' \tag{7}$$

is a Fourier transform of E_i. We will often find it convenient to rewrite the last term of equation (5) as $\int_0^\infty W_i(k)P(k)dk$, where the *one-dimensional* window function is the angular average

$$W_i(k) = k^2\int W_i(\mathbf{k})d\Omega_k. \tag{8}$$

The *shot noise bias* is given by

$$b_i \equiv \int\frac{E_i(\mathbf{r},\mathbf{r})}{\bar{n}(\mathbf{r})}d^3r. \tag{9}$$

Alternatively, b_i can be made to vanish by omitting the terms with $\alpha=\beta$ from the double sum in equation (4), as described in Appendix A. The term $W_i(\mathbf{0})$ simply probes the mean density of the survey, and as described in Section 5 and Appendix B, the functions E_i should always be chosen such that this term vanishes, *i.e.*, so that $\int E_i(\mathbf{r},\mathbf{r}')d^3r d^3r'=0$, thereby immunizing the power estimates to errors in normalization of $\bar{n}$ (Section 5.1)[7]. The desirability of choosing windows with this property was first explicitly pointed out by Fisher *et al.* (1993), and this prescription was used also by *e.g.* Hamilton (1992) and Cole, Fisher & Weinberg (1994). We want to interpret q_i in equation (5) as probing a weighted average of the power spectrum, with the window function giving the weights, so E_i should be normalized so that $W_i(\mathbf{k})$ integrates to unity (throughout this paper, we will write the volume element in Fourier space as $d^3k/(2\pi)^3$ rather than

[6] Another problem with large N is the memory requirements for an $N\times N$ covariance matrix; if $N=10^5$, for example, this is currently beyond the RAM capacity of all but the very largest computers.

[7] In fact, the stronger constraint $\int E_i(\mathbf{r},\mathbf{r}')d^3r'=0$ for all $\mathbf{r}$ should be enforced, as discussed in Appendix B.

d^3k, since this minimizes the number of 2π-factors elsewhere). Using equation (6) and Parseval's theorem, this gives

$$1=\int_0^\infty W_i(k)\frac{dk}{(2\pi)^3}=\int W_i(\mathbf{k})\frac{d^3k}{(2\pi)^3}=\int E_i(\mathbf{r},\mathbf{r})d^3r, \tag{10}$$

so if we think of $E_i(\mathbf{r},\mathbf{r}')$ as a matrix with indices $\mathbf{r}$ and $\mathbf{r}'$, this normalization condition is simply tr $[E_i]=1$.

As described in Hamilton (1997a), the window function has a simple geometrical interpretation. Let us rewrite equation (8) as

$$W_i(k)=4\pi k^2\int_0^\infty w_i(r)j_0(kr)dr, \tag{11}$$

where the *separation weighting* is defined by

$$w_i(d)\equiv\int\int E_i(\mathbf{r},\mathbf{r}')\delta^D(|\mathbf{r}-\mathbf{r}'|-d)d^3r\,d^3r'. \tag{12}$$

The separation weighting is the average of E_i over all pairs of points separated by a fixed distance d, weighted by the number of such pairs. Equation (11) shows that the only aspect of E_i that affects the window function is the separation weighting, since two different E_i that give the same w_i will produce identical window functions.

An important special case, to which we return in the next subsection, is that where E_i is of rank one, *i.e.*, of the separable form

$$E_i(\mathbf{r},\mathbf{r}')=\psi_i(\mathbf{r})\psi_i(\mathbf{r}')^* \tag{13}$$

for some ψ_i. In this case, equation (4) can be written as $q_i=|x_i|^2$, where $x_i=\int\psi_i(\mathbf{r})n(\mathbf{r})/\bar{n}(\mathbf{r})$ (see equation (15) below), and the window function becomes simply

$$W_i(\mathbf{k})=|\widehat{\psi}_i(\mathbf{k})|^2. \tag{14}$$

(Here and throughout, hats denote Fourier transforms; $\widehat{\psi}(\mathbf{k})\equiv\int e^{-i\mathbf{k}\cdot\mathbf{r}}\psi(\mathbf{r})d^3r$ — see Appendix D.) We will see that both the traditional methods (Section 3.3) and the KL method (Section 3.5) are of this separable form, while the quadratic data compression method that we present in Section 3.6 is not. We will occasionally refer to the former methods as *linear* data compression, since they begin by taking linear combinations of the data with weight functions $\psi_i(\mathbf{r})$; cf., equation (15). The latter method, on the other hand, is intrinsically quadratic since q_i is not merely the square of some quantity that is linear in the data. This is because its pair weighting E is optimized to provide a minimum variance power estimator, which makes E a quadratic form of rank greater than one.

To avoid confusion, the reader should bear in mind that when we distinguish between linear and quadratic methods below, we are referring to linear versus quadratic *data compression*. The power estimator q_i is of course quadratic in both cases.

3.2. *Pixelization*

Hamilton (1997a) has recently derived the functions E_i that provide the minimum-variance power spectrum estimates for an arbitrary selection function and survey geometry. Unfortunately, this optimal weighting scheme is in general impractical to implement numerically, as it involves a computationally cumbersome infinite series expansion — only in the small-scale limit does it become simple, as will be described in Section 4.3. To proceed numerically, it is therefore convenient to discretize the problem. This reduces it to one similar to that of cosmic microwave background (CMB) experiments: estimating a power spectrum given noisy fluctuation measurements in a number of discrete "pixels". Once the pixelization is done, the remaining steps are quite analogous to the CMB case (T97), and involve mere matrix operations such as inversion and diagonalization. These operations can often be further simplified by a more suitable choice of pixelization.

Let us define the overdensity in N "pixels" $x_1,...,x_N$ by

$$x_i\equiv\int\left[\frac{n(\mathbf{r})}{\bar{n}(\mathbf{r})}-1\right]\psi_i(\mathbf{r})d^3r \tag{15}$$

for some set of functions ψ_i. We discuss specific choices of ψ_i in some detail below. One generally strives either to make these functions fairly localized in real space (in which case the pixelization is a generalized form of counts in cells) or fairly localized in Fourier space (in which case we will refer to the functions ψ_i as "modes" and to x_i as expansion coefficients). The reader should be warned that the latter case is far from what would normally be thought of as "pixels", and that we will be using this terminology nonetheless to stress that all discretization schemes can be treated in a mathematically equivalent way.

The "-1" term in equation (15) simply subtracts off the mean density. As we discuss in Section 5 and Appendix B, one can and should choose the weight functions ψ_i to have zero mean, making this term irrelevant (note indeed our use of this fact immediately preceding equation (14)). This corresponds to requiring that

$$\int\psi_i(\mathbf{r})d^3r=0. \tag{16}$$

Let us group the pixels x_i into an N-dimensional vector $\mathbf{x}$. From equations (15) and (16) and a generalization of the derivation of equation (5), T95 shows that

$$\langle\mathbf{x}\rangle = \mathbf{0}, \tag{17}$$
$$\langle\mathbf{x}\mathbf{x}^\dagger\rangle = \mathbf{C}\equiv\mathbf{N}+\mathbf{S}, \tag{18}$$

where the shot noise covariance matrix is given by

$$\mathbf{N}_{ij}=\int\frac{\psi_i(\mathbf{r})\psi_j(\mathbf{r})^*}{\bar{n}(\mathbf{r})}d^3r \tag{19}$$

and the signal covariance matrix is

$$\mathbf{S}_{ij}=\int\widehat{\psi}_i(\mathbf{k})\widehat{\psi}_j(\mathbf{k})^*P(k)\frac{d^3k}{(2\pi)^3}. \tag{20}$$

How should we choose our ψ_i's to pixelize space? For a pixelization to be useful, we clearly want the data set $\mathbf{x}$ to retain as large a fraction as possible of the cosmological information from the original data set (the galaxy positions), while simultaneously simplifying subsequent calculations. We here list several natural options, most of which have appeared in the literature.

3.2.1. *Counts in cells*

Here one partitions the survey volume into N mutually exclusive and collectively exhaustive volumes V_i, and defines $\psi_i(\mathbf{r})=\bar n(\mathbf{r})$ if $\mathbf{r}\in V_i$, $\psi_i(\mathbf{r})=0$ elsewhere. Thus x_i is simply the number of galaxies observed in V_i, minus the expected average. A set of useful approximations for computing $\mathbf{N}$ and $\mathbf{S}$ for this case is derived in VS96. To keep the number of cells from becoming intractably large, one might choose the cells to be larger in distant and poorly sampled regions of space than nearby.

With these sharp-edged cells, any linear combinations of pixels will correspond to a weight function that is discontinuous at cell boundaries. To avoid power leakage problems that this can in principle cause, one might use cells with "fuzzy" boundaries instead, for instance Gaussians as described in T95.

One can greatly simplify the computation of the covariance matrix by choosing all cells to have the same shape and to be spherically symmetric (*e.g.*, spheres or Gaussians), since the resulting $\mathbf{S}_{ij}$ will depend only on the separation of the pixel centers and this correlation function can be pre-computed and splined once and for all. For the volume-limited case ($\bar n$ constant), performing the appropriate integrals for identical spherical cells of radius R, separated by uR, gives

$$\mathbf{N}_{ij}=\begin{cases}\frac{(2-u)^2(4+u)}{16}\bar n V & \text{if } u<2,\\ 0 & \text{if } u\geq 2,\end{cases}\tag{21}$$

where $V=4\pi R^3/3$.

3.2.2. *Fourier modes*

All of the above-mentioned pixels were fairly well-localized in real space. To make pixels reasonably localized in Fourier space, one can choose modes ψ_i that are plane waves, tapered by some weight function ϕ to make them square integrable:

$$\psi_i(\mathbf{r})=\phi(\mathbf{r})e^{i\mathbf{k}_i\cdot\mathbf{r}}\tag{22}$$

for some grid points $\mathbf{k}_i$ in Fourier space. Choosing all modes that are periodic in a box containing the survey volume will ensure that this is a complete set, although some high-frequency cutoff is of course necessary to keep the number of pixels finite. Four different choices of the volume weighting function ϕ have appeared in the literature:

$$\phi(\mathbf{r})\ \propto\ \begin{cases}1 \text{ inside survey volume}\\ 0 \text{ outside survey volume}\end{cases}\tag{23}$$

$$\phi(\mathbf{r})\ \propto\ \bar n(\mathbf{r})\tag{24}$$

$$\phi(\mathbf{r})\ \propto\ \frac{\bar n(\mathbf{r})}{1+\bar n(\mathbf{r})P}\tag{25}$$

$$\phi(\mathbf{r})\ \propto\ \text{eigenfunction of }\left[\nabla^2-\frac{\gamma}{\bar n(\mathbf{r})}\right]\tag{26}$$

All are to be normalized so that the corresponding window functions integrate to unity – $\int\phi(\mathbf{r})d^3r=1$ as shown below. Note that without careful choice of the $\mathbf{k}_i$'s, equation (16) will *not* be satisfied in general for these pixelizations. The first choice, which weights all volume elements in the survey equally, was employed by *e.g.* Vogeley *et al.* (1992), Fisher *et al.* (1993), and Park *et al.* (1994). The second choice is used when determining the angular power spectrum of a sample without redshifts, *e.g.*, the APM survey (Baugh & Efstathiou 1994). Then all galaxies by default receive equal weight (moreover, modes can of course only be computed in the directions perpendicular to the line of sight). The third choice, advocated by Feldman, Kaiser & Peacock (1994, hereafter FKP), minimizes the variance in the limit when $k^{-1}\ll$ the depth of the survey and is discussed in detail in Section 4.3.1. Here P denotes an *a priori* guess of the power in the band under consideration. The fourth choice (Tegmark 1995, hereafter T95), gives the narrowest window function for a given variance, where the constant γ determines the tradeoff.

3.2.3. *Spherical harmonic modes*

The spherical wave choice

$$\psi_i(\mathbf{r})=Y_{\ell m}(\widehat{\mathbf{r}})j_\ell(k_n r),\tag{27}$$

where $Y_{\ell m}$ is a spherical harmonic and j_ℓ a spherical Bessel function, is well-suited for full-sky redshift surveys, and has the advantage (Fisher, Scharf, & Lahav 1994; Heavens & Taylor 1995) of greatly simplifying inclusion of the effect of redshift-space distortions in the analysis.

3.2.4. *Guessed eigenmodes*

In Section 3.5.1, we describe a set of smooth functions known as continuum signal-to-noise eigenmodes. These modes have a number of useful properties. In particular, there is an integer N such that the first N eigenmodes retain virtually all the cosmological information in the survey. If one has a reasonable guess as to the shape of these functions *a priori*, before computing them exactly (they depend on the survey geometry), the first N of these guessed modes are obviously a good choice for the pixelization functions ψ_i, since the amount of information destroyed by the pixelization process will then be small.

3.3. *Traditional methods*

The traditional approach has been to estimate the power by simply squaring the pixels ($q_i=|x_i|^2$), choosing the pixels to be Fourier modes as described above. This corresponds to the pair weighting of equation (13), where ψ_i is given by equation (22) and ϕ is specified by equation (23), (24), (25) or (26). Equation (5) shows that

$$\langle q_i\rangle=\widehat\phi(\mathbf{0})^2+\int\frac{\phi(\mathbf{r})^2}{\bar n(\mathbf{r})}d^3r+\frac{1}{(2\pi)^3}\int|\widehat\phi(\mathbf{k}-\mathbf{k}_i)|^2P(\mathbf{k})\frac{d^3k}{(2\pi)^3},\tag{28}$$

the last term of which is simply the true power spectrum convolved with the function $|\widehat\phi(\mathbf{k})|^2$. The 3D window function is thus $W_i(\mathbf{k})=|\widehat\phi(\mathbf{k}-\mathbf{k}_i)|^2$. Using this and Parseval's theorem (or equation (10) directly), one finds that the window function normalization constraint $\int W_i(\mathbf{k})d^3k/(2\pi)^3=1$ corresponds to simply

$$\int\phi(\mathbf{r})^2d^3r=1.\tag{29}$$

These simple expressions have frequently been used in the literature, but an annoying complication has often been neglected. As described in Section 5, the fact that the normalization of $\bar{n}$ is not known *a priori*, but is determined from the observed galaxies (the so-called integral constraint problem) can be eliminated by choosing weight functions ψ_i that are orthogonal to the monopole, *i.e.*, such that $\widehat{\psi}_i(\mathbf{0}) = 0$. Although this is easy to arrange with the pixelized methods described below, the choice of equation (22) does generally *not* have this important property. To obtain a correct answer with the traditional methods, this must be corrected for. As shown in Appendix B, the power estimator

$$\tilde{P}_i \equiv \left[\frac{|x_i|^2 - b_i}{A_i}\right] \tag{30}$$

is unbiased and incorporates the integral constraint correction (when $\bar{n}$ is normalized so that $x_i = 0$ for $\mathbf{k}_i = 0$) if the normalization factor A_i and the shot noise correction b_i are given by

$$A_i = \left(1 + \left|\frac{\widehat{\phi}(\mathbf{k}_i)}{\widehat{\phi}(\mathbf{0})}\right|^2\right) a(\mathbf{0}) - 2\,\mathrm{Re}\left\{\frac{\widehat{\phi}(\mathbf{k}_i)}{\widehat{\phi}(\mathbf{0})} a(\mathbf{k}_i)\right\} \tag{31}$$

$$b_i = \left(1 + \left|\frac{\widehat{\phi}(\mathbf{k}_i)}{\widehat{\phi}(\mathbf{0})}\right|^2\right) b(\mathbf{0}) - 2\,\mathrm{Re}\left\{\frac{\widehat{\phi}(\mathbf{k}_i)}{\widehat{\phi}(\mathbf{0})} b(\mathbf{k}_i)\right\} \tag{32}$$

where the functions a and b are defined by

$$a(\mathbf{k}) \equiv \int \phi(\mathbf{r})^2 e^{i\mathbf{k}\cdot\mathbf{r}} d^3r, \tag{33}$$

$$b(\mathbf{k}) \equiv \int \frac{\phi(\mathbf{r})^2}{\bar{n}(\mathbf{r})} e^{i\mathbf{k}\cdot\mathbf{r}} d^3r. \tag{34}$$

If the survey is volume limited, then $\bar{n}$ is independent of $\mathbf{r}$, $b(\mathbf{k}) = a(\mathbf{k})/\bar{n}$, and $b_i = A_i/\bar{n}$.

After computing power estimates q_i at a large grid of points $\mathbf{k}_i$, one finally takes some weighted averages of the q_i to obtain power estimates in some bands of (scalar) k ($= |\mathbf{k}|$). The problem of finding the weights that minimize the variance of the band power estimators is unfortunately quite a difficult one, and in general involves solving a numerically unpleasant quadratic programming problem (T95). For this reason, the customary approach has been to simply give all q_i equal weights in some spherical shells in $\mathbf{k}$-space although, as described in VS96, this is in general far from optimal.

3.4. *Brute force likelihood method*

Let us parameterize the power spectrum $P(k)$ by some parameter vector $\mathbf{\Theta}$ as in Section 2.2. In the approximation that the probability distribution for the pixel vector $\mathbf{x}$ is a multivariate Gaussian, it is given by equation (2) with mean $\mathbf{m} = \mathbf{0}$. The *maximum likelihood estimator* of $\mathbf{\Theta}$, denoted $\mathbf{\Theta}_{ml}$, is simply that $\mathbf{\Theta}$-vector that maximizes the likelihood $f(\mathbf{x};\mathbf{\Theta})$. This maximization problem can unfortunately not be solved analytically when the number of pixels exceeds one, so $\mathbf{\Theta}_{ml}$ is a complicated non-linear function of $\mathbf{x}$ that must be computed by solving the maximization problem numerically. Since one generally wants error bars on the estimate as well, one typically evaluates the likelihood function at a dense grid of points in parameter space and rescales it to integrate to unity. These final results are often illustrated in contour plots, as at the bottom of Figure 1.

3.5. *Linear (Karhunen-Loève) method*

As will be discussed in Section 6, the traditional methods generally destroy information, while the brute force method is lossless but computationally impractical. The Karhunen-Loève (KL) method (Karhunen 1947) maintains the advantage of the brute-force method (indeed, it can produce an essentially identical answer faster), and has additional useful features as well, as we detail below. It was first introduced into large-scale structure analysis by Vogeley & Szalay (see VS96), and it has also been successfully applied to Cosmic Microwave Background data (*e.g.* Bunn 1995; Bond 1995; TTH; Bunn & White 1997; Jaffe, Knox, & Bond 1997). Like any method designed to minimize error bars, the KL-technique requires an *a priori* assumption for the power spectrum. This is referred to as the *fiducial* $P(k)$. As will be described in sections 3.5.4, 4.2 and 6.3.5, this is not a problem in practice, since a bad fiducial model does not bias the result. Moreover, the measurement itself can be used as the fiducial model in an iterative procedure if desired.

We start by defining signal-to-noise eigenmodes in Section 3.5.1, before generalizing the technique in Section 3.5.2.

3.5.1. *Signal-to-noise eigenmodes*

The signal-to-noise eigenmode method consists of defining a new data vector

$$\mathbf{y} \equiv \mathbf{Bx}, \tag{35}$$

where $\mathbf{b}$, the rows of the matrix $\mathbf{B}$, are the N eigenvectors of the generalized eigenvalue problem

$$\mathbf{Sb} = \lambda\mathbf{Nb}, \tag{36}$$

sorted from highest to lowest eigenvalue λ and normalized so that $\mathbf{b}^\dagger\mathbf{Nb} = \mathbf{I}$. This implies that

$$\langle y_i y_j\rangle = \delta_{ij}(1+\lambda_i), \tag{37}$$

which means that the transformed data values $\mathbf{y}$ have the desirable property of being *statistically orthogonal*, *i.e.*, uncorrelated. In the approximation that the distribution function of $\mathbf{x}$ is a multivariate Gaussian, this also implies that they are statistically independent — then $\mathbf{y}$ is merely a vector of independent Gaussian random variables. Moreover, since the eigenmodes diagonalize both $\mathbf{S}$ and $\mathbf{N}$ simultaneously, equation (36) shows that the eigenvalues λ_i can be interpreted as a signal-to-noise ratio $\mathbf{S}/\mathbf{N}$. Since equation (37) shows that the quantity $y_i^2 - 1$ on average equals this signal-to-noise ratio, it is a useful band power estimator when normalized so that its window function integrates to unity. The window function is given by equation (14) with ψ_i replaced by the continuous KL mode defined by

$$\psi_i'(\mathbf{r}) \equiv \sum_{j=1}^{N} \mathbf{B}_{ij}\psi_j(\mathbf{r}), \tag{38}$$

since $y_i = \int[n(\mathbf{r})/\bar{n}(\mathbf{r}) - 1]\psi_i'(\mathbf{r})d^3r$. For the volume-limited case, the noise power is simply $1/\bar{n}$, so the correctly normalized and bias-corrected KL band power estimators are simply "signal = noise × signal-to-noise", *i.e.*,

$$q_i \equiv \frac{y_i^2 - 1}{\bar{n}}; \tag{39}$$

compare this with equation (30). Since the matrix $\mathbf{B}$ is invertible, the final data set $\mathbf{y}$ clearly retains all the information that was present in $\mathbf{x}$. In summary, the KL transformation partitions the information content of the original data set $\mathbf{x}$ into N chunks that are

1. mutually exclusive (independent),
2. collectively exhaustive (jointly retaining all the information), and
3. sorted from best to worst in terms of their information content.

Typically, most of the KL coefficients y_i have a signal-to-noise ratio $\lambda \ll 1$, so that the bulk of the cosmological information is retained in the first N' coefficients, $N' \ll N$, which is why the KL method is often referred to and used as *data compression*. One can thus throw away all but the first N' numbers y_i without any appreciable information loss, and this compressed data set will still satisfy properties 1 and 3 exactly, and 2 to a good approximation.

Bunn (1995) and VS96 have pointed out that the S/N-coefficients $\mathbf{y}$ are useful for power spectrum estimation since as long as the galaxy survey probes only scales smaller than the peak in the power spectrum (cf., the discussion in Section 6.3.3), the first N' window functions W_i have the following two properties:

1. They are narrow in k-space,
2. As i increases from 1 to N', they probe all the scales accurately measured by the survey, from largest to smallest.

Since they are also uncorrelated, these power spectrum estimators therefore have all desirable properties that one may wish for: they distill the cosmological information content of the data set into a set of mutually exclusive and collectively exhaustive chunks, which correspond to the band powers in a set of narrow bands. In the approximation that $\mathbf{y}$ has a Gaussian distribution, the probability distributions of the power estimates q_i are simple: they are independent χ^2 distributions with one degree of freedom.[8]

3.5.2. *General KL modes*

Let us write $\mathbf{C} = \theta_i\mathbf{S} + \mathbf{N}$, where the power spectrum normalization parameter $\theta_i = 1$ in the fiducial model. Since $\mathbf{S} = \mathbf{C}_{,i} \equiv \partial\mathbf{C}/\partial\theta_i$, equation (36) can be rewritten as (TTH)

$$\mathbf{C}_{,i}\,\mathbf{b} = \lambda'\mathbf{C}\mathbf{b}, \tag{40}$$

where $\lambda' = \lambda/(1+\lambda)$ or, equivalently, $\lambda = \lambda'/(1-\lambda')$. From now on, we will normalize the eigenvectors so that $\mathbf{b}^\dagger\mathbf{C}\mathbf{b} = 1$ instead of $\mathbf{b}^\dagger\mathbf{N}\mathbf{b} = 1$, since this is more convenient throughout the rest of the paper. The matrix elements $\mathbf{B}_{ij}$ are thus a factor $(1+\lambda_i)^{1/2}$ smaller than in the previous subsection. As shown in TTH, solving the eigenvalue equation (40) is useful for *any* parameter θ_i on which $\mathbf{C}$ depends, even those which do not affect only the power spectrum (e.g., redshift-space distortions, cf., Appendix C), and the signal-to-noise eigenmode method discussed above is just the special case where θ_i is the power normalization. The three properties listed above continue to hold in the general case, and the "information content" in item 3 above now refers to the information about the parameter θ_i.

The KL method is a very general data analysis tool. Note that the eigenmodes continue to be mutually exclusive and collectively exhaustive if we replace $\mathbf{C}_{,i}$ by *any* symmetric matrix $\mathbf{M}$ in equation (40). To ensure that the KL modes give narrow window functions ranging from small to large k, one can therefore choose $\mathbf{M}$ to be the signal covariance matrix $\mathbf{S}$ that would arise from some monotonically decreasing fiducial power spectrum, for instance $P(k) \propto k^{-3}$. In that case, the modes become sorted by the ratio of "signal" in the *fiducial* power spectrum to noise; as the former is monotonic, the modes are sorted by scale. We discuss this point further in Section 6.3.3.

3.5.3. *KL modes are asymptotically pixelization-independent*

The pixelizations listed in Section 3.2 were all in some sense arbitrary, and generally somewhat redundant. The KL modes can eliminate this arbitrariness, as mentioned in Section 3.2.4. If the pixelization functions ψ_i formed a complete set, spanning the space of all square-integrable functions over the survey volume, the continuum KL modes defined by equation (38) would be some smooth functions $\psi_i'(\mathbf{r})$ that were independent of the pixelization used to compute them and depended only on $P(k)$, $\bar{n}(\mathbf{r})$ and the geometry of the survey. In practice, the functions ψ_i do not of course form a complete set, since one is limited to a finite number of pixels, but the continuum KL modes $\psi_i'(\mathbf{r})$ can nonetheless be computed numerically. By choosing a pixelization that can resolve all features down to some scale R, equation (38) will accurately approximate all continuum modes whose window functions $W_i(k)$ probe only scales $k^{-1} \gg R$. If the number of pixels are increased further to probe smaller scales, the first N' modes remain stable against this perturbation: the first N' eigenvalues and eigenvectors would only change by a small amount (Szalay & Vogeley 1997). The new modes, on the other hand, would represent the small scale noise probed by the new pixel scale. By using the functions $\psi_i'(\mathbf{r})$ to pixelize the data, $i = 1, ..., N$, one can thus make sure that all of the cosmological signal down to the scale probed by ψ_N' is retained.

[8]To prevent power spectrum plots from becoming too cluttered with points with large error bars, it is convenient to combine neighboring band power estimates with a weighted average — these broader band powers will of course still be uncorrelated since all the q_i are.

3.5.4. *Using KL modes for trouble spotting*

If the assumed power spectrum model is correct, the KL coefficients y_i defined by equation (35) will be independent Gaussian random variables with zero mean and unit variance. This offers an efficient way of testing whether the data are inconsistent with this model. The detection of, say, a $6-\sigma$ outlier ($|y_i| > 6$) would provide strong evidence that there is either more variance on the scale probed by the i^{th} mode than the fiducial power spectrum assumed, or that the probability distribution is strongly non-Gaussian on that scale. Even if it goes undetected, an incorrect assumed model does not bias the estimate of the power spectrum, as discussed below in Section 6.3.5.

3.5.5. *Using KL modes for linear filtering*

The KL eigenmodes have an additional use. The process of throwing away the eigenmodes with low signal-to-noise ratios splits the space of all possible density fields given the data into two — one subspace that mostly contains noise, and one that is dominated by our generalized signal. The two are statistically orthogonal to one another. The expansion of a given dataset over the signal subspace will substantially reduce the noise, thus representing a useful linear filtering of the data (cf., Seljak 1997). Since it maximizes the signal for a given number of included modes, the KL transform is sometimes referred to as "optimal subspace filtering". More generally, let us define a filtered data set

$$x'_i \equiv \sum_{j=1}^{N} (\mathbf{C}\mathbf{B}^\dagger)_{ij} w_j y_j \tag{41}$$

for some weights w_j. The generalized eigenvector orthogonality relation gives $\mathbf{B}\mathbf{C}\mathbf{B}^\dagger = \mathbf{I}$ (cf., VS96, TTH), which implies that $\mathbf{C}^{-1} = \mathbf{B}^\dagger\mathbf{B}$ and $\mathbf{C}\mathbf{B}^\dagger\mathbf{B} = \mathbf{I}$. Hence equation (35) gives $\mathbf{C}\mathbf{B}^\dagger\mathbf{y} = \mathbf{x}$, which implies that we recover the original data set ($\mathbf{x}' = \mathbf{x}$) in equation (41) if we choose the weights $w_j = 1$. Another simple example is the optimal subspace filtering mentioned above, which corresponds to the choice $w_j = 1$ for $j \leq N'$, $w_j = 0$ otherwise. Finally, it is easy to show that the choice $w_j = \lambda'_j$ gives *Wiener filtering*, which is defined by $\mathbf{x}' = \mathbf{S}\mathbf{C}^{-1}\mathbf{x}$. In other words, Wiener filtering becomes diagonal in the KL basis, since it diagonalizes $\mathbf{S}$ and $\mathbf{C}$ simultaneously. Indeed, Wiener filtering is but one of many linear filters that are straightforward to implement in the KL basis.

3.6. *Quadratic method*

In both traditional methods and the pixelized KL technique, the power spectrum estimates q_i are some quadratic functions of the observed density field $n(\mathbf{r})$, *i.e.*, of the form given by equation (4). Hamilton (1997a) adopted a more ambitious approach, and computed the unbiased quadratic power estimators that have minimal variance, using a series expansion. T97 subsequently showed that these estimators are unbeatable; their Fisher information matrix is identical to that of the raw data, so no non-quadratic unbiased estimators can give smaller variance. Moreover, they can be computed without recourse to the numerically cumbersome series expansion of Hamilton (1997a) (c.f. T97 and Knox, Bond, & Jaffe 1997 for applications to CMB observations). Here we show how this method can be applied to galaxy surveys, and its relation to the KL method. Just as for the KL technique, a fiducial power spectrum is assumed.

We will parameterize the power spectrum as a piecewise constant function with N' "steps" of height p_i, which we term the **band powers**. Thus $P(k) = p_i$ for $k_i \leq k < k_{i+1}$, where

$$0 = k_1 < k_2 < ... < k_{N'+1} = \infty, \tag{42}$$

and group them into an N'-dimensional vector $\mathbf{p}$. For the method to be strictly lossless, these bands should be chosen to be quite narrow compared to the scales on which the power spectrum varies. One first computes a compressed data vector $\mathbf{q}$ whose N' elements are quadratic functions of the data set $\mathbf{x}$. These are defined by

$$q_i \equiv \frac{1}{2}\mathbf{z}^\dagger \mathbf{C}_{,i}\mathbf{z}, \tag{43}$$

where the vector $\mathbf{z}$ is given by

$$\mathbf{z} \equiv \mathbf{C}^{-1}\mathbf{x}, \tag{44}$$

and the matrix $\mathbf{C}_{,i}$ is defined by

$$(\mathbf{C}_{,i})_{ab} \equiv \int_{k_i<|\mathbf{k}|<k_{i+1}} \widehat{\psi}_a(\mathbf{k})\widehat{\psi}_b(\mathbf{k})^* \frac{d^3k}{(2\pi)^3}. \tag{45}$$

That is, $\mathbf{C}_{,i}$ is the derivative of the covariance matrix $\mathbf{C}$ (equation (18)) with respect to the normalization of the i^{th} band, in the limit of narrow bands. Rewriting this as

$$q_i \equiv \frac{1}{2}\mathbf{x}^\dagger \mathbf{E}_i\mathbf{x}, \tag{46}$$

where

$$\mathbf{E}_i \equiv \mathbf{C}^{-1}\mathbf{C}_{,i}\,\mathbf{C}^{-1}, \tag{47}$$

we see that the matrix $\mathbf{E}_i$ is simply a discrete version of the pair weighting function $E_i(\mathbf{r},\mathbf{r}')$ in equation (4). Note that it is not separable, *i.e.*, it has rank greater than one.

For the Gaussian case, the Fisher information matrix for $\mathbf{p}$ defined by equation (3) reduces to (VS96, TTH)

$$\mathbf{F}_{ij} = \frac{1}{2}\mathrm{tr}\,\left[\mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{C}_{,j}\right], \tag{48}$$

and T97 shows that both the mean and the covariance of $\mathbf{q}$ are given in terms of $\mathbf{F}$:

$$\begin{aligned} \langle\mathbf{q}\rangle &= \mathbf{F}\mathbf{p}, & (49)\\ \langle\mathbf{q}\mathbf{q}^t\rangle - \langle\mathbf{q}\rangle\langle\mathbf{q}\rangle^t &= \mathbf{F}. & (50) \end{aligned}$$

This means that $\mathbf{F}^{-1}\mathbf{q}$ is an optimal estimator of $\mathbf{p}$, since its covariance matrix is precisely the inverse of the Fisher matrix. Moreover, as shown in T97, compressing the data set $\mathbf{x}$ into the coefficients $\mathbf{q}$ for some sufficiently narrow bands is a strictly lossless procedure, retaining all the information about those cosmological parameters that affect galaxy clustering through the power spectrum alone. Equation (49) shows that, in terms of the band powers, the window functions W of equation (8) for the coefficients q_i are simply proportional to the rows of the ubiquitous Fisher matrix.

The coefficients $\mathbf{F}^{-1}\mathbf{q}$ tend to be both correlated and noisy, and therefore it is better to work with the transformed coefficients defined by

$$\tilde{\mathbf{p}} \equiv \mathbf{F}^{-1/2}\mathbf{q}, \tag{51}$$

renormalized so that their window functions integrate to unity. Here $\mathbf{F}^{1/2}$ denotes the symmetric matrix whose square is $\mathbf{F}$ – it is readily computed by diagonalization. As shown by Tegmark & Hamilton (1997), these coefficients are all uncorrelated (multiply equation (50) on both the right and the left by $\mathbf{F}^{-1/2}$ to see this), and moreover tend to be very well-behaved numerically, with narrow non-negative window functions (the rows of $\mathbf{F}^{1/2}$) spanning the entire range of scales probed by the survey.

4. RELATIONS BETWEEN THE METHODS

4.1. *Relation between KL and quadratic method*

As we will now show, the linear and quadratic pixelized methods are closely related — the latter is simply a faster way of computing the same band power estimates.

Let us optimize our KL modes to estimate not the overall power normalization, but the band power in band i. As we showed in Section 3.5.5,

$$\mathbf{C}^{-1} = \mathbf{B}^{\dagger}\mathbf{B}. \tag{52}$$

Introducing the diagonal matrix $\mathbf{\Lambda} \equiv \mathrm{diag}\{\lambda'_j\}$, equation (40) implies that

$$\mathbf{B}\mathbf{C}_{,i}\mathbf{B}^{\dagger} = \mathbf{\Lambda}. \tag{53}$$

We are partially suppressing the index i here for simplicity, although the eigenvectors in $\mathbf{B}$ and the eigenvalues in $\mathbf{\Lambda}$ of course depend on which power band i we optimize for. Equations (35), (52) and (53) allow us to rewrite equation (43) as

$$\begin{aligned} 2q_i &= \mathbf{x}^{\dagger}\mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{x} = \mathbf{x}^{\dagger}\mathbf{B}^{\dagger}\mathbf{B}\mathbf{C}_{,i}\mathbf{B}^{\dagger}\mathbf{B}\mathbf{x} \\ &= \mathbf{y}^{\dagger}\mathbf{\Lambda}\mathbf{y} = \sum_j \lambda'_j y_j^2, \end{aligned} \tag{54}$$

i.e., as a weighted average of the squared KL coefficients y_j, with weights given by the KL eigenvalues λ'_j. Thus the coefficients are weighted by their inverse variance, which means that this is the minimum-variance band power estimator based on the squared KL coefficients, so after subtracting off shot noise and normalizing correctly, the linear (KL) and quadratic methods give exactly the same estimator for the band power. As described in Section 6.4, the quadratic estimator is is simply faster to compute. A more intuitive way of understanding why the two methods give the same result is to note that the quadratic method was derived explicitly to be the minimum variance estimator (T97; cf., equation (49)), and that VS96 showed that the KL approach is guaranteed to minimize the variance as well.

4.2. *Relation between quadratic and ML methods: iteration*

If the quadratic method is repeated using the output (measured) power spectrum as the input (fiducial) power spectrum, then this iteration will eventually converge to the power spectrum that would be obtained with the brute force maximum-likelihood method (Bond, Jaffe, & Knox 1997). This is because the quadratic method can be derived by expanding the logarithmic likelihood $\ln f$ to second order around the fiducial point and maximizing it. The iteration thus seeks the maximum of $\ln f$ by repeatedly approximating it with a parabola. This is essentially the maximum-gradient method for maximization, which is known to have excellent convergence properties. In the one-dimensional case, this is equivalent to finding the zero of $(\ln f)'$ by repeated linear approximations. This is simply the Newton-Raphson root-finding method, known to converge exponentially fast and asymptotically double the number of correct decimals in each iteration.

Although the ML-result can be computed by iterating the quadratic method, it should be stressed that this is not necessarily a good thing to do. If the measured power spectrum is noisy, so that the band power uncertainties do not satisfy $\Delta P \ll P$, then iteration can produce misleadingly small error bars, as the following example illustrates. Suppose sample variance or a shot noise fluctuation makes a band power measurement ten times smaller than the true value and this is used as the new fiducial band power P (prior). Then equation (62) shows that the nominal ΔP from sample variance will be ten times too small as well. If iteration is nonetheless desired, it is crucial not to let the prior over-fit the data. A better approach than simple iteration is therefore to use as prior a fit to the measured $P(k)$ with a smooth parameterized fitting function, and keep adding fitting parameters until the the value of χ^2 per degree of freedom drops below unity (Tegmark 1997a; Seljak 1997).

4.3. *Relation between quadratic and FKP methods: the small scale limit*

For a traditional method with a volume weighting function $\phi(\mathbf{r})$, we define a quantity L implicitly by

$$\int_{kL<1} |\widehat{\phi}(\mathbf{k})|^2 \frac{d^3k}{(2\pi)^3} = \frac{1}{2}, \tag{55}$$

i.e., L^{-1} is the radius of a sphere in Fourier space containing half of the $\mathbf{k}=\mathbf{0}$ window function. For simple volumes such as a pencil beams, slices or spheres, L is of the order of the width of the survey in its *narrowest* direction. We will now show that in the small-scale limit where $k^{-1} \ll L$, the quadratic method reduces to the FKP method, which implies that the latter is lossless and hence unbeatable for measuring the power spectrum on the smallest scales, as was pointed out by Hamilton (1997a).

4.3.1. *Derivation of the FKP method*

[9] An intuitive way to understand why all directions in k-space should receive equal weight when $\Delta k \gg L^{-1}$ is to note that in this limit, the number of coherence volumes that fit into a given solid angle in the shell is independent of its shape (and hence independent of direction in k-space), being merely the ratio of the shell subvolume to the coherence volume.

The FKP method (Feldman, Kaiser & Peacock 1994) is the traditional method described in Section 3.3, using the volume weighting of equation (25). The rank one power estimates $|x_a^2|$ are averaged with equal weights[9] for all modes in a spherical shell $|\mathbf{k}| = k_a$, so integrating equation (13) over this shell using equation (22) for the pixels, we see that this corresponds to the pair weighting

$$E_i(\mathbf{r},\mathbf{r}') = \phi(\mathbf{r})\phi(\mathbf{r}')j_0(k_i|\mathbf{r}-\mathbf{r}'|), \tag{56}$$

where $j_0(x) \equiv \sin(x)/x$. This holds for any choice of volume weighting function ϕ. The specific FKP choice given by equation (25) is derived by minimizing the variance of the corresponding power estimators q_i. This involves a number of approximations. We summarize the derivation here, to clarify why the FKP weighting is optimal only on small scales.

Substituting equations (15) and (22) into equations (19) and (20), we obtain the pixel covariance matrix

$$\begin{aligned} \mathbf{C}_{ab} &= \int \widehat{\phi}(\mathbf{k}-\mathbf{k}_a)\widehat{\phi}(\mathbf{k}-\mathbf{k}_b)^* P(k)\frac{d^3k}{(2\pi)^3} \\ &+ \int e^{i(\mathbf{k}_a-\mathbf{k}_b)\cdot\mathbf{r}}\frac{\phi(\mathbf{r})^2}{\bar{n}(\mathbf{r})}d^3r. \end{aligned} \tag{57}$$

The function $\widehat{\phi}(\mathbf{k})$ roughly falls off on a scale L^{-1}. As long as $L^{-1} \ll k$ and the power spectrum $P(k)$ is a smooth function, $P(k)$ will therefore be almost constant where the first integrand is non-negligible (*i.e.*, where $|\mathbf{k}-\mathbf{k}_a| \lesssim L^{-1}$ and $|\mathbf{k}-\mathbf{k}_b| \lesssim L^{-1}$) and can be approximately factored out of the k-integral. Using the convolution theorem, this yields

$$\mathbf{C}_{ab} \approx \int e^{i(\mathbf{k}_a-\mathbf{k}_b)\cdot\mathbf{r}}\phi(\mathbf{r})^2\left[P + \frac{1}{\bar{n}(\mathbf{r})}\right]d^3r, \tag{58}$$

where $P \equiv P([k_a+k_b]/2)$. This shows that $\mathbf{C}_{ab} \approx 0$ if $|\mathbf{k}_a-\mathbf{k}_b| \gg L^{-1}$, so power estimates separated by much more than $\Delta k = L^{-1}$ are essentially uncorrelated. Conversely, power estimates from nearby shells with $|\mathbf{k}_a-\mathbf{k}_b| \ll L^{-1}$ are almost perfectly correlated and therefore redundant. The FKP method therefore averages the power estimates $|x_a|^2$ in thicker shells $|\mathbf{k}_a| \in [k_i, k_{i+1}]$ as in equation (42), whose widths satisfy $L^{-1} \ll |k_{i+1}-k_i| \ll k$, giving approximately uncorrelated power estimates q_i. This approximation clearly only holds for small scales, $k^{-1} \ll L$. We denote the volume of the i^{th} shell

$$V_s \equiv \frac{4}{3}\pi\left(k_{i+1}^3 - k_i^3\right)/(2\pi)^3. \tag{59}$$

Assuming that the pixelized data $\mathbf{x}$ is Gaussian distributed, the variance of the power estimator q_i is simply given by averaging $2|C_{ab}|^2$ over the shell, *i.e.*,

$$(\Delta q_i)^2 = \frac{2}{V_s^2}\int\int|\mathbf{C}_{ab}|^2\frac{d^3k_a}{(2\pi)^3}\frac{d^3k_b}{(2\pi)^3}, \tag{60}$$

where both integrals are to be taken over the i^{th} shell. Substituting equation (58), and using the fact that $L^{-1} \ll |\mathbf{k}_a-\mathbf{k}_b| \ll k$, one of these integrals simply produces a factor V_s. Applying Parseval's theorem to the result and using equation (29) to normalize[10] finally leaves us with the approximation

$$\left(\frac{\Delta q_i}{\langle q_i\rangle}\right)^2 \approx 2\frac{\int\phi(\mathbf{r})^4\left[1+\frac{1}{\bar{n}(\mathbf{r})P}\right]^2d^3r}{V_s\left[\int\phi(\mathbf{r})^2d^3r\right]^2}. \tag{61}$$

This holds for any weighting function ϕ. The FKP choice of ϕ is derived by minimizing this approximate expression for the variance. Since it is left unchanged if we multiply ϕ by a constant, we can for simplicity impose the normalization constraint $\int\phi^2d^3r = 1$. Minimizing the numerator of equation (61) with a Lagrange multiplier for this constraint now gives the FKP weighting of equation (25). Substituting this back into equation (61) and omitting the power band index i for simplicity finally yields

$$\left(\frac{\Delta P}{P}\right)^2 \equiv \left(\frac{\Delta q}{\langle q\rangle}\right)^2 \approx \frac{2}{V_s V^{eff}}, \tag{62}$$

where

$$V^{eff}(k) \equiv \int\left[\frac{\bar{n}(\mathbf{r})P(k)}{1+\bar{n}(\mathbf{r})P(k)}\right]^2d^3r \tag{63}$$

can be interpreted as the *effective volume* probed, since the integrand is of order unity where one is signal dominated ($P \gg 1/\bar{n}$) and ~ 0 where one is noise dominated. For a volume-limited survey with spatial volume V and constant $\bar{n}$, the FKP prescription weights all galaxies equally, and we simply have $V^{eff} = [1+1/\bar{n}P]^{-2}V$. An intuitive way to understand equation (62) is to note that the nearby Fourier amplitudes x_i are correlated over a *coherence volume* $V_c \sim 1/V$. Thus as long as shot noise is not dominant, $\Delta P/P \sim \sqrt{2/\mathcal{N}}$, where $\mathcal{N}$ is the number of approximately uncorrelated volumes V_c that fit into the shell volume V_s. In conclusion, we have shown that for traditional methods, the FKP volume weighting of equation (25) is optimal if and only if we limit ourselves to small scales, $k^{-1} \ll L$.

4.3.2. *The small-scale limit of the quadratic method*

It is often convenient to work in the continuum limit where a discrete index i is replaced by a continuous variable such as $\mathbf{r}$ or $\mathbf{k}$. Vectors a_i and matrices $\mathbf{A}_{ij}$ then correspond to functions of one and two variables, such as $a(\mathbf{r})$ and $A(\mathbf{r},\mathbf{r}')$. Since all sums get replaced by integrals in this limit, units frequently differ from the discrete case [11].

Let us take the pixelization functions to be Dirac delta functions $\psi_i(\mathbf{r}) \equiv \delta^D(\mathbf{r}-\mathbf{r}_i)$, corresponding to a continuum of pixels

$$x(\mathbf{r}) = \frac{n(\mathbf{r})}{\bar{n}(\mathbf{r})} - 1, \tag{64}$$

[10] Since $kL \gg 1$, equation (29) does not need the integral constraint correction.

[11] Since $\mathbf{b} = \mathbf{A}\mathbf{x}$ means $b_i = \sum_j \mathbf{A}_{ij}x_j$ in the discrete case and $b(\mathbf{r}) = \int\mathbf{A}(\mathbf{r},\mathbf{r}')x(\mathbf{r}')d^3r'$ in the continuous case, we see that matrix multiplication introduces units of the volume element, in this example d^3r. Thus the continuous analog of the identity matrix, $\mathbf{I}(\mathbf{r},\mathbf{r}') \equiv \delta^D(\mathbf{r}-\mathbf{r}')$, is its own inverse, since $(\mathbf{I}^2)(\mathbf{r},\mathbf{r}') \equiv \int\mathbf{I}(\mathbf{r},\mathbf{r}'')\mathbf{I}(\mathbf{r}'',\mathbf{r}')d^3r'' = \mathbf{I}(\mathbf{r},\mathbf{r}')$, even though it is not dimensionless — the delta function δ^D has units of inverse volume.

and choose our parameters to be the values of the 3D power spectrum:

$$\theta_i = P(\mathbf{k}_i). \tag{65}$$

The matrices $\mathbf{C}$ and $\mathbf{C}_{,i}$ then reduce to

$$\mathbf{C}(\mathbf{r},\mathbf{r}') = \int e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}P(\mathbf{k})\frac{d^3k}{(2\pi)^3} + \frac{\delta^D(\mathbf{r}-\mathbf{r}')}{\bar{n}(\mathbf{r})}, \tag{66}$$

$$\mathbf{C}_{,i}(\mathbf{r},\mathbf{r}') = \frac{\partial\mathbf{C}(\mathbf{r},\mathbf{r}')}{\partial P(\mathbf{k}_i)} = \frac{1}{(2\pi)^3}e^{-i\mathbf{k}_i\cdot(\mathbf{r}-\mathbf{r}')}, \tag{67}$$

since $P(\mathbf{k}) = \int P(\mathbf{k}_i)\delta^D(\mathbf{k}-\mathbf{k}_i)d^3\mathbf{k}_i$ implies that $\partial P(\mathbf{k})/\partial P(\mathbf{k}_i) = (2\pi)^3\delta^D(\mathbf{k}-\mathbf{k}_i)$. As we saw in Section 4.3.1, the small-scale limit corresponds to neglecting the k-dependence of P, which gives $\mathbf{C}(\mathbf{r},\mathbf{r}') \approx [P + 1/\bar{n}(\mathbf{r})]\delta^D(\mathbf{r}-\mathbf{r}')$ and

$$\mathbf{C}^{-1}(\mathbf{r},\mathbf{r}') = \left[P + \frac{1}{\bar{n}(\mathbf{r})}\right]^{-1}\delta^D(\mathbf{r}-\mathbf{r}') = \phi(\mathbf{r})\delta^D(\mathbf{r}-\mathbf{r}'), \tag{68}$$

where ϕ is the FKP weighting function of equation (25) normalized so that $\phi = \bar{n}/(1+\bar{n}P)$.

We will now rederive the FKP results with much less effort than in the previous section, by simply using the quadratic method formulas. Substituting equations (67) and (68) into equation (43) shows that the quadratic estimators $\mathbf{q}$ are given by

$$q_i \equiv \frac{1}{2}|\widehat{z}(\mathbf{k}_i)|^2, \tag{69}$$

where the function $z(\mathbf{r}) \equiv \phi(\mathbf{r})x(\mathbf{r})$, which apart from the factor of $1/2$ are exactly the FKP estimators of $P(\mathbf{k})$. Calculating the Fisher matrix by substituting equations (67) and (68) into equation (48) gives

$$\begin{aligned}\mathbf{F}_{ij} &= \frac{1}{2}\mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{C}_{,j}\right]\\ &\approx \frac{1}{2}\int\phi(\mathbf{r})e^{-i\mathbf{k}_i\cdot(\mathbf{r}-\mathbf{r}')}\phi(\mathbf{r}')e^{-i\mathbf{k}_j\cdot(\mathbf{r}'-\mathbf{r})}d^3r\,d^3r'\\ &= \frac{1}{2}|\widehat{\phi}(\mathbf{k}_i-\mathbf{k}_j)|^2.\end{aligned} \tag{70}$$

Equation (49) now gives

$$\langle q_i\rangle = \int\mathbf{F}_{ij}P(\mathbf{k}_j)\frac{d^3k_j}{(2\pi)^3}, \tag{71}$$

which shows that the Fisher matrix is simply the 3D window function. To make the window function of the power estimate q_i integrate to unity, we need to divide it by the quantity

$$\int\mathbf{F}_{ij}\frac{d^3k_j}{(2\pi)^3} = \frac{V^{eff}}{2P^2}, \tag{72}$$

where we have used equation (70), Parseval's theorem, equation (25) and equation (63) in the last step. Using equation (50), this shows that

$$\frac{\Delta q_i}{\langle q_i\rangle} \approx \sqrt{2}\frac{\int\phi(\mathbf{r})d^3r}{\int\phi(\mathbf{r})^2d^3r} = \sqrt{2}\left[1+\frac{1}{\bar{n}P}\right], \tag{73}$$

where the last equal sign only holds for the volume limited case where $\bar{n}$ is constant. Alternatively, averaging the power estimates q_i over a shell in $\mathbf{k}$-space as in the previous section, equation (50) reproduces the FKP error formula of equation (62). In summary, we have shown that the FKP method becomes lossless for measuring the power on small scales, since it equals the optimal quadratic method in this limit.

5. SYSTEMATIC PROBLEMS — THE SELECTION FUNCTION AND EXTINCTION

We have discussed the importance of the integral constraint in Section 3.3. Here we discuss this issue in the context of pixelized methods, and generalize it to a whole host of systematic problems, including errors in our assumed selection function $\bar{n}(\mathbf{r})$ and our assumed extinction map.

5.1. *What is the integral constraint?*

If we knew the selection function $\bar{n}(\mathbf{r})$ *a priori*, before counting the galaxies in our survey, we would be able to measure the power on the scale of the survey. Our power spectrum estimate would essentially be the square of the ratio of the observed and expected number of galaxies in our sample. Of course, we do not know $\bar{n}$ *a priori*, so we use the galaxies themselves to normalize the selection function. Thus the measured density fluctuation automatically vanishes on the scale of the survey, and naive application of any of the power spectrum estimation methods we have described will falsely indicate that $P(k)\to 0$ as $k\to 0$, regardless of the behavior of the true power spectrum on large scales (Peacock & Nicholson 1991). See also Tadros & Efstathiou (1996). Equation (5) tells us that apart from the noise bias b_i, there is an additional term $W_i(\mathbf{0})$ that must be subtracted off to make q_i an unbiased power spectrum estimator. Let us assume that we know the shape of the selection function but not its normalization. To reflect this, we write the true selection function as

$$\bar{n}(\mathbf{r}) = a\bar{n}_0(\mathbf{r}), \tag{74}$$

where $\bar{n}_0$ is our guessed selection function, and a is an unknown normalization constant, and find that our noise bias corrected power estimator will have a residual bias $(a-1)^2W_i(\mathbf{0})$. Since we do not know the exact value of a, our only way to eliminate this bias is to require that $W_i(\mathbf{0}) = 0$, *i.e.*, by requiring the window function to vanish at $\mathbf{k} = \mathbf{0}$. This was first explicitly pointed out by Fisher *et al.* (1993).

5.2. *Its relation to extinction and other systematic problems*

Since requiring $W_i(\mathbf{0}) = 0$ eliminates the integral constraint problem, the trouble is confined to the $\mathbf{k} = \mathbf{0}$ mode. If $W_i(\mathbf{0}) = 0$, then the power estimate q_i is clearly completely independent of $P(0)$, the fluctuations in this untrustworthy mode. The essence of our approach is therefore the following:

- *We have a systematic problem with a certain mode, and can immunize our results from this problem by making them independent of this untrustworthy mode.*

When phrased in this way, it is clear that this approach can be applied to a variety of other systematic problems as well. For instance, incorrectly modeled extinction adds excess power in the form of purely angular modes (density fluctuations that have no radial component, *i.e.*, being perpendicular to the line of sight). It might therefore be desirable to make the results independent of all purely angular modes on the relevant scales or, in a less ambitious approach, at least independent of those modes whose shape coincide with known dust templates. Similarly, misestimating the *shape* of the radial selection function (of $\bar{n}_0$ in equation (74)) pollutes certain purely radial modes which one may wish to discard.

5.3. *How to eliminate untrustworthy modes with pixelized methods*

In this section, we show how the power spectrum estimate from a pixelized method can be made insensitive to the type of systematic errors described above.

Let us parameterize the true selection function $\bar{n}$ as

$$\bar{n}(\mathbf{r}) = \sum_{j=1}^{M} a_j \bar{n}_j(\mathbf{r}), \tag{75}$$

where $\bar{n}_j$ are known functions and the parameters a_j, which we group into an M-dimensional vector $\mathbf{a}$, are *a priori* unknown. Imagine for example a simple case where $M = 3$, $\bar{n}_1$ is our best guess for a purely radial selection function based on a Schechter luminosity function, $\bar{n}_2$ is a (purely angular) dust template, and $\bar{n}_3$ gives the effect of an infinitesimal error in the estimate of the characteristic luminosity L_*: $\bar{n}_3(\mathbf{r}) = \partial\bar{n}_1(\mathbf{r})/\partial L_*$. Let $\bar{n}_0$ denote some *a priori* estimate of $\bar{n}$. Defining the "uncorrected" pixels as

$$x_i' \equiv \int \frac{n(\mathbf{r})}{\bar{n}_0(\mathbf{r})}\psi_i(\mathbf{r})d^3r, \tag{76}$$

we find that

$$\langle \mathbf{x}' \rangle = \mathbf{U}\mathbf{a}, \tag{77}$$

where the $N \times M$ matrix $\mathbf{U}$ of untrustworthy modes is defined by

$$\mathbf{U}_{ij} \equiv \int \frac{\bar{n}_j(\mathbf{r})}{\bar{n}_0(\mathbf{r})}\psi_i(\mathbf{r})d^3r. \tag{78}$$

This means that in general, the data set $\langle \mathbf{x}' \rangle \neq 0$, so the uncorrected data set does not satisfy equation (17).

We show how to solve this problem in Appendix B. In short, one replaces $\mathbf{x}$ by a cleaned data set $\mathbf{\Pi x}$, where $\mathbf{\Pi}$ is a matrix that satisfies $\mathbf{\Pi U} = \mathbf{0}$ and thus projects out the untrustworthy modes. We find that the best choice is $\mathbf{\Pi} = \mathbf{I} - \mathbf{U}(\mathbf{U}^\dagger\mathbf{C}^{-1}\mathbf{U})^{-1}\mathbf{U}^\dagger\mathbf{C}^{-1}$. Although the covariance matrix of the cleaned data set is not invertible, we find that with this choice of $\mathbf{\Pi}$, the quadratic method remains strictly optimal if we simply use $\mathbf{C}$ from the uncleaned data in equation (44). We also show that the integral constraint correction given by equations (31) and (32) corresponds to this optimal method in the small scale limit.

6. PROS AND CONS OF THE METHODS

Above we have presented all methods for galaxy power spectrum estimation that have proposed in the literature, as well as the new pixelized quadratic technique and various extensions, and showed how they are related to one another. We will now discuss their relative merits at some length. It will become clear that they are highly complementary, and we summarize the pros and cons of each method in Table 1. We will return to this in the discussion section, where we describe how the traditional, KL and quadratic methods can be used in concert to produce a data analysis pipeline having all the properties on our wish list from Section 2.

	TRAD	LIN	QUAD
Optimal on largest scales	$-$	$+$	$+$
Optimal on smallest scales	$+$	$-$	$-$
Simple & uncorrelated errors	$-/+$	$+/-$	$+$
Measures z-space distortions	$-$	$+$	$-$
Accomodates systematic effects	$-$	$+$	$+$

Table 1: *Pros and cons of the traditional, linear (KL) and quadratic power spectrum estimation methods.*

6.1. *Pros and cons of the direct Fourier method*

Unlike the pixelized methods, the traditional Fourier method uses the exact galaxy positions and thus retains all the small-scale information about $P(k)$.

As discussed in Section 4.3, this method becomes lossless in the limit $kL \to \infty$ when the FKP choice of ϕ, equation (25), is used. In addition, choosing the shell widths $\Delta k \gg L^{-1}$ guarantees that the errors in the band power estimates $\tilde{P}_i$ will be approximately uncorrelated. This means that on small scales, say $kL \ll 10\%$, this method satisfies the first three criteria on our wish list in Section 2.

However, these advantages no longer hold when measuring power on scales comparable to the size of the survey. VS96 review problems with the direct Fourier approach that occur unless $kL \ll 1$. They fall into the two categories described below.

6.1.1. *The direct Fourier method destroys information*

Hamilton (1997b) has shown that a strictly optimal direct summation method can be derived *in principle*, in terms of a series expansion, but this is unfortunately extremely burdensome numerically except for scales much smaller than the survey size, away from the boundaries, where it approximates equations (25) and (26). The optimal galaxy pair weighting is not separable (in the sense of equation (13)) and thus cannot be expressed in terms of a volume weighting function ϕ as above. Consequently, no direct Fourier methods are lossless except on small scales.

6.1.2. *The method is complicated and computationally slow for large scales*

It is important to note that the Fourier transformation calculation takes only a negligible amount of time in a power spectrum analysis — the lion's share of the work involves computing the mean corrections $W_i(\mathbf{0})$, the shot noise correction b_i, the normalization factor A_i and the covariance matrix of the power estimates. This becomes numerically cumbersome on the largest scales, when $kL \sim 1$.

This is because both smearing from the window function and the effect of the integral constraint become important and must be accurately computed in this regime. Even though the relevant integrals can be greatly accelerated with FFT's, the calculations are not only much more complicated and obscure than the simple linear algebra of the pixelized methods, but generally substantially slower as well. The most time-consuming step is the computation of the error bars ΔP and the covariance between power estimates using an integral constraint corrected and appropriately generalized version of equation (60), since the integral in equation (57) must be done separately for each *pair* of grid points $(\mathbf{k}_i, \mathbf{k}_j)$ in Fourier space (cf., Goldberg & Strauss 1998). The property of uncorrelated errors is clearly lost, making it quite difficult to compute the optimal weights for averaging the power estimates in $\mathbf{k}$-space into power bands in k-space (T95, VS96). In short, when comparing the direct Fourier method to the KL and quadratic methods on the largest scales, it is more complicated, uses more CPU time and produces an inferior result.

6.2. *Pros and cons of the brute force method*

The pixelized brute force method has been applied to galaxy survey analysis using expansions in spherical harmonics (Fisher *et al.* 1994; Heavens & Taylor 1995; Ballinger, Heavens & Taylor 1996).

It is arguably the simplest of all methods, both conceptually and in implementation. Moreover, it can be shown to be lossless in the limit of large data sets, thus giving minimal error bars. An important disadvantage is that it is slow. The slowest step in evaluating the likelihood function f is to compute det $\mathbf{C}$, for which the CPU time required scales as N^3, and f is evaluated at a large number of grid points in the multi-dimensional parameter space. This is why it is so useful if N, the number of pixels, can be reduced by a lossless data compression scheme before performing the likelihood analysis, throwing away noise and keeping the signal. As we described above, this is exactly what the Karhunen-Loève and quadratic methods do. In the KL scheme, the compressed data set $\mathbf{y}$ was a linear function of $\mathbf{x}$ (of the form $\mathbf{y} = \mathbf{Bx}$ for some matrix $\mathbf{B}$), and in the quadratic scheme $\mathbf{q}$ was a quadratic function of $\mathbf{x}$, of the form $q_i = \mathbf{x}^\dagger \mathbf{E}_i \mathbf{x}/2$ for some matrices $\mathbf{E}_i$.

A second disadvantage is that the maximum likelihood parameter estimate $\mathbf{\Theta}_{ml}$ (defined in Section 3.4) is such a complicated function of $\mathbf{x}$ that we cannot calculate its statistical properties analytically. As we saw above, both the linear and quadratic schemes allow us to write down power spectrum estimators in closed form (equations (39) and (43), respectively) This allows one to compute their probability distributions exactly, (the y_i are Gaussian and the q_i are are generalized χ^2 distributions, often close to Gaussian), which makes these power estimates easy to use for parameter fitting further down the data analysis pipeline.

6.3. *Pros and cons of the KL method*

6.3.1. *It retains the phase information*

An important advantage of the KL method over the others in the table is that the KL coefficients retain the spatial information about the data (not only the Fourier amplitudes, but the corresponding phases as well). This means that information on processes which affect the radial and angular clustering patterns differently can be optimally probed with the KL method. Such effects include

1. Redshift space distortions (cf., Appendix C),
2. Galactic extinction (which affects only angular modes),
3. Evolution and mis-estimates of $\bar{n}(r)$ (which affect only radial modes).

We have not discussed in any detail how one might include these effects in a KL analysis. Suffice it to say that for any physical effect that affects the covariance matrix of the pixels, one can determine KL modes which are optimized for parameters which describe this effect; cf., Section 3.5.2. The way to do this for redshift-space distortions (see Appendix C) is described in more detail in TTH. We argue below in Section 6.3.4 that the simple signal-to-noise eigenmodes are appropriate general purpose modes for a KL analysis of any parameter which affects the covariance matrix only via the power spectrum. On the other hand, the KL modes should be custom tailored (using equation (40)) for parameters not in this category, such as ones causing anisotropic clustering.

Another approach is that discussed in Section 5 and Appendix B: rather than *measuring* these systematic effects, one can project out those modes that are sensitive to them, making the resulting dataset immune from them. This can be done for any pixelized method, but is especially simple for the quadratic method.

6.3.2. *Pixelization is not lossless*

The "$-$" on row 2 of Table 1 refers to the fact that computational constraints place an upper limit on the number of pixels N used for the eigenvalue problem, since the storage required scales as N^2 and the CPU time as N^3. $N = 10^4$ is readily handled on a high-end 1997 workstation (TTH), and new methods under development (Szalay & Vogeley 1997) may well be able to increase this by an order of magnitude or more, but since the dynamic range is $\sim N^{1/3} \propto (\text{CPU time})^{1/9}$, some information will always be lost on the very smallest scales. Fortunately, this is not a problem in practice, since the complementary traditional methods work best precisely on the smallest scales.

6.3.3. *The KL power peak problem*

For very deep surveys such as SDSS and 2dF, which probe scales substantially beyond the expected peak in the power spectrum at $\sim 200h^{-1}$Mpc, the KL window functions will generally not be narrow but double-peaked, since fluctuations longward of the peak have the same signal-to-noise ratio as certain fluctuations shortward of the peak and will get mixed in the corresponding KL modes. One readily circumvents this degeneracy problem by performing a likelihood analysis on the KL modes, with the band powers being the parameters to be estimated. However, the statistical errors on the resulting band power estimates will no longer be uncorrelated, and since this is a nonlinear operation, they will also not have the simple χ^2 distribution in general. This is why Table 1 indicates "+/−" in row 4: the "+" applies when we use the direct approach on a volume smaller than $\sim 200h^{-1}$ Mpc in size, and the "−" applies when we use the indirect approach on a deep data set such as SDSS or 2dF. As was described in Section 3.5.2, one can circumvent this power peak problem by using choosing a monotonically decreasing fiducial power spectrum such as $P(k) \propto k^{-3}$. This can hardly be said to make the KL-modes less "optimal", since they will still partition the information into mutually exclusive and collectively exhaustive chunks. They simply become sorted according to a different criterion: not by their information content regarding the power normalization, but by the physical scale they probe.

6.3.4. *The KL multi-parameter complication*

The derivation of the signal-to-noise eigenmode method in *e.g.* VS96 or TTH does *not* prove that the compressed data set retains the bulk of the information about all parameters of cosmological interest, but merely that it is lossless with respect to the overall power spectrum normalization. In fact, this is not a problem in practice. TTH describe a method where one carries out a series of KL transforms, optimizing for each parameter of interest in turn, then pools all the resulting eigenmodes and uses singular value decomposition to eliminate redundancy from the pool of modes. In addition, there is good reason to believe that the information about all cosmological parameters is nonetheless preserved even with the signal-to-noise eigenmodes alone. We now give a hand-waving argument to this effect, and describe some detailed numerical experiments that support it. TTH computed three separate sets of KL modes for the CMB data set of the COBE satellite ($N = 4016$), optimized for measuring three different parameters: the power spectrum normalization Q, the slope n and the reionization parameter τ. The 3×3 Fisher matrix was then computed from the three compressed data sets separately for $N' = 500$ modes retained. Each set of modes retained virtually all the information about their corresponding parameter, but the error bars on Q with the n or τ modes were substantially larger than their Cramér-Rao minimum. That is, the n-modes and the τ-modes lost information about Q. On the other hand, the Q-modes were found to retain virtually all the information about n and τ. Examination of their window functions revealed why. To obtain a large lever arm for determining the slope, the n-modes were probing mainly the largest and the smallest available scales, ignoring those in the middle near the "pivot point". The τ-modes were ignoring the very largest scales, since these are unaffected by reionization and therefore carry no information about τ. The Q-modes, on the other hand, were faithfully probing the power on all available scales, and therefore automatically retained all of the information about n and τ as well, as a side effect. Thus as long as the galaxy power spectrum has no sharp features (which the signal-to-noise eigenmodes might potentially ignore) we expect the standard KL modes optimized for the normalization to be close to lossless with respect to all cosmological parameters affecting only the power spectrum.

6.3.5. *Does KL bias the results towards our theoretical prejudice?*

It has been argued that the KL method biases the results by guessing an *a priori* fiducial power spectrum when computing the pixel covariance matrix. This claim has been extensively tested numerically (Bunn 1995, TTH), and found to be completely unfounded. In general, the effect of guessing an incorrect prior model is to leave the estimates unbiased, but with slightly larger error bars than what is optimal. If desired, any dependence on the initial data can clearly be eliminated by iterating the KL procedure as described in Section 4.2, while at the same time reducing the error bars close to the minimum allowed by the Fisher matrix.

6.4. *Pros and cons of the quadratic method*

One advantage of this method is that its simple and uncorrelated quadratic estimators have narrow window functions at all k, even beyond the peak in the power spectrum. Thus it is a useful complement to the KL method on the very largest scales, as indicated by row 4 in Table 1.

A second advantage is evident from equation (43): Since the CPU time for multiplication of a vector by a matrix scales as N^2, the time for computing $\mathbf{z}$ scales as N^2 as well if equation (44) is solved by an iterative technique such as the conjugate gradient method (Press *et al.* 1992). This is much faster than the KL method, which scales as N^3. This speed increase may allow the quadratic method to be used over a somewhat larger dynamical range than the KL method, extending down to smaller scales. The quadratic method can also be used as a faster way to obtain the same results as the brute force ML method, using the iteration scheme described in Section 4.2.

Furthermore, we saw in Section 5 and Appendix B that the quadratic method can be made immune to various sorts of systematic errors which might plague the data.

An important disadvantage is that, unlike the KL method, it does not retain any phase information. This is a drawback when estimating the underlying real-space power spectrum, since although it can measure this directly by computing the appropriate $\mathbf{C}$ once the distortion parameter $\beta \equiv \Omega^{0.6}/b$ is known, the KL method or another linear approach must be used first, to measure β. Once cannot simply immunize the data from mis-estimates of β using the formalism of Appendix B, since β affects virtually *all* the modes. It does, however, appear possible to generalize the quadratic method to overcome this limitation (Hamilton 1997, private communication).

7. DISCUSSION & CONCLUSIONS

In this section, we summarize our discussion of the pros and cons of the various methods, and conclude by describing an approach combining the strengths of all of them, illustrated in Figure 1.

We found that although the direct Fourier approach is both simple to implement and virtually lossless on scales much smaller than the smallest dimension of the sample in question, it has several drawbacks on larger scales:

1. It loses information, giving unnecessarily noisy measurements.

2. It is quite tedious to implement numerically if one uses the exact expressions we have derived for the integral constraint correction, especially for computing the covariance.

In contrast, the two pixelized methods are lossless on *large* scales, but lose small-scale information because numerical constraints on the number of pixels limit the dynamical range. Since they are simpler to implement as well, they allow a more ambitious approach incorporating complications such as redshift-space distortions, residual extinction and radial selection function errors (Section 5, Appendices B and C). The quadratic method can compute exactly the same band powers as the KL method, and do so faster (the number of operations scaling as the square rather than the cube of the number of pixels), allowing more pixels and a larger dynamical range. The KL method, on the other hand, is the only one which retains the phase information in which clustering anisotropies (differences between the angular and radial clustering patterns) is encoded. Since redshift distortions and various systematic problems manifest themselves in this way, the KL method is therefore a powerful complement to the quadratic method, since the former can quantify and subtract these systematic effects and pass the appropriate redshift-distortion parameter β along to the latter, which can then measure the power spectrum directly in real space as described in Appendix C. Alternatively, the quadratic method can be immunized from such systematic effects, as described in Appendix B. The KL method is also useful for cosmographic purposes, where spatial information is everything. Finally, the KL method also has the advantage of greatly simplifying trouble spotting such as search for outliers and non-Gaussian behavior.

In conclusion, we have found that although none of the methods can be made both feasible and lossless on its own, we can obtain a feasible and virtually lossless data analysis pipeline satisfying our entire wish list by combining three of them, as outlined in Figure 1.

1. The power spectrum on scales $k^{-1} \lesssim L/10$ is estimated directly from the raw redshift data with the traditional direct Fourier approach.
2. The raw data is binned into spatial pixels substantially smaller than $L/10$, so that this pixelization process retains all the information except that which was already captured by the traditional method.
3. The linear (KL) method is used to measure anisotropy parameters such as $\Omega^{0.6}/b$ (from redshift-space distortions), a residual extinction template and corrections to the radial selection function, as well as large-scale band powers.
4. Uncorrelated estimates of the power spectrum on scales $k^{-1} \gtrsim L/10$ are computed with the quadratic method, extending down to even smaller scales if $N \sim 10^4 - 10^5$ pixels are feasible. This can be done both by incorporating the systematic effects found in Step 3 with the KL method, and by using quadratic estimators which are insensitive to these systematic effects. The comparison of the results for the band powers allows us to quantify how successful we are in eliminating these effects.
5. The entire process may be iterated, using the (smoothed) measured power spectrum as the fiducial one.
6. Remaining cosmological parameters are estimated with a likelihood or χ^2 analysis from the power spectrum.

This approach should allow future redshift surveys to realize their full potential to constrain cosmological models. In the meantime, it appears worthwhile to reanalyze various existing surveys with the same pipeline, to eliminate any method-induced artifacts and allow more accurate cross-comparisons of results.

We thank the Aspen Center for Physics where this paper neared its final form, David Weinberg and Josh Frieman for useful conversations, and an anonymous referee for a very close reading of the manuscript and many helpful suggestions. Support for this work was provided by NASA through grant NAG5-6034 and Hubble Fellowships HF-01078.01-94A and HF-01084.01-96A, awarded by the Space Telescope Science Institute, which is operated by AURA, Inc. under NASA contract NAS5-26555. M.A.S. acknowledges the support of the Alfred P. Sloan Foundation, Research Corporation, and NSF Grant AST96-16901.

REFERENCES

Ballinger, B., Heavens, A., & Taylor, A. 1996, MNRAS, 276, L59
Baugh, C. M. & Efstathiou, G. 1994, MNRAS, 267, 323
Bond, J. R. 1995, Phys. Rev. Lett., 74, 4369
Bond, J. R., Efstathiou, G. & Tegmark, M. 1997, MNRAS, 291, L33
Bond, J. R., Jaffe, A. H., & Knox, L. 1997, preprint astro-ph/9708203
Bunn, E. F. 1995, Ph.D. Thesis, U.C. Berkeley, *ftp pac2.berkeley.edu/pub/bunn*
Bunn, E. F., & White, M 1997, ApJ, 480, 6
Cole, S., Fisher, K. B. & Weinberg, D. H. 1994, MNRAS, 267, 785
Dodelson, S., Hui, L. & Jaffe, A. 1997, preprint astro-ph/9712074
Efstathiou, G. P. 1996, in *Cosmology and Large Scale Structure*, eds. R. Schaeffer, J. Silk, M. Spiro, & J. Zinn-Justin (Elsevier, Amsterdam)
Feldman, H. A., Kaiser, N. & Peacock, J. A. 1994, ApJ, 426, 23 (FKP)
Fisher, K. B., Davis, M., Strauss, M. A., Yahil, A., & Huchra, J. P. 1993, ApJ, 402, 42
Fisher, K. B., Scharf, C. A. & Lahav, O. 1994, MNRAS, 266, 219
Fisher, R. A. 1935, J. Roy. Stat. Soc., 98, 39
Fry, J. N. 1984, ApJ, 277, L5
Goldberg, D. M. & Strauss, M. A. 1998, ApJ, 495, in press
Gunn, J. E., & Weinberg, D. H. 1995, in *Wide-Field Spectroscopy and the Distant Universe*, eds. S. J. Maddox and A. Aragón-Salamanca (Singapore: World Scientific), 3
Hamilton, A. J. S. 1992, ApJ, 385, L5
Hamilton, A. J. S. 1997a, MNRAS, 289, 285
Hamilton, A. J. S. 1997b, MNRAS, 289, 295
Hamilton, A. J. S. 1997c, preprint astro-ph/9708102
Hamilton, A. J. S. & Culhane, M. 1996, MNRAS, 278, 73
Heavens, A. F. & Taylor, A. N., 1995, MNRAS, 483, 497
Jaffe, A. H., Knox, L. & Bond, J. R. 1997, preprint astro-ph/9702109
Jungman, G.. Kamionkowski, M., Kosowsky, A & Spergel, D. N. 1996, Phys. Rev. D, 54, 1332
Kaiser, N. 1987, MNRAS, 227, 1
Karhunen, K., *Über lineare Methoden in der Wahrscheinlichkeitsrechnung* (Kirjapaino oy. sana, Helsinki, 1947)
Kendall M. G. & Stuart, A. 1969, *The Advanced Theory of Statistics*, Vol. II (Griffin, London)
Kenney, J. F. & Keeping, E. S. 1951, *Mathematics of Statistics, Part*

II, 2nd ed. (Van Nostrand, New York)
Knox, L. 1997, in preparation
Knox, L., Bond, J. R. & Jaffe, A. H. 1997, preprint astro-ph/9702110
Park, C., Vogeley, M. S, Geller, M. J. & Huchra, J. P. 1994, ApJ, 431, 569
Peacock, J. A. 1997, MNRAS, 284, 885
Peacock, J. A. & Dodds, S. J. 1994, MNRAS, 267, 1020
Peacock, J. A. & Nicholson, D. 1991, MNRAS, 253, 307
Peebles, P. J. E 1973, ApJ, 185, 413
Peebles, P. J. E 1980, *The Large-Scale Structure of the Universe* (Princeton U. P., Princeton)
Peebles, P. J. E 1993, *Principles of Physical Cosmology* (Princeton U. P., Princeton)
Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. 1992, *Numerical Recipes* (Cambridge Univ. Press)
Seljak, U. 1997, preprint astro-ph/9710269
Strauss, M. A. 1997, preprint astro-ph/9610033, to appear in *Structure Formation in the Universe*, eds. A. Dekel & J. P. Ostriker
Strauss, M. A. & Willick, J. A. 1995, Phys. Rep., 261, 271
Szalay, A. S., Matsubara, T., & Landy, S. D. 1997, preprint astro-ph/9712007
Szalay, A. & Vogeley, M. S. 1997, in preparation
Tadros, H. & Efstathiou, G. 1996, MNRAS, 282, 1381
Tegmark, M. 1995, ApJ, 455, 429 (T95)
Tegmark, M. 1997a, Phys. Rev. D, 55, 5895 (T97)
Tegmark, M. 1997b, Phys. Rev. Lett., 79, 3806
Tegmark, M. 1997c, in *Ringberg Workshop on Large-Scale Structure*, ed. D. Hamilton (Kluwer, Amsterdam)
Tegmark, M. & Bromley, B. C. 1995, ApJ, 453, 533
Tegmark, M. & Hamilton, A. J. S. 1997, preprint astro-ph/9702019
Tegmark, M., Taylor, A. & Heavens, A. F. 1997, ApJ, 480, 22 (TTH)
Vogeley, M. S. 1995, in *Wide-Field Spectroscopy and the Distant Universe*, eds. S.J. Maddox & A. Aragón-Salamanca (World Scientific, Singapore), 142
Vogeley, M. S. 1997, in *Ringberg Workshop on Large-Scale Structure*, ed. D. Hamilton (Kluwer, Amsterdam)
Vogeley, M. S, Park, C., Geller, M. J. & Huchra, J. P. 1992, ApJ, 391, L5
Vogeley, M. S. & Szalay, A. S. 1996, ApJ, 465, 34 (VS96)
White, M., Viana, T. P., Liddle, A. R. & Scott, D. 1996, MNRAS, 283, 107
White, S. D. M. 1979, MNRAS, 186, 145
Zaldarriaga, M., Seljak, U. & Spergel, D. N. 1997, ApJ, 488, 1
Zaroubi, S. & Hoffman, Y. 1996, ApJ, 462, 25

APPENDIX

APPENDIX A: SHOT NOISE REMOVAL

There are two basic approaches in the literature to removing shot noise. In this Appendix, we show that the two give essentially identical results.

Using the Gaussian approximation

Let us define write our shot noise corrected band power estimator as $(q_i - \tilde{b}_i)$, where $\tilde{b}_i$ denotes our *bias correction*. Equation (5) shows that this estimator will only be unbiased if $\langle\tilde{b}_i\rangle = b_i$, where b_i is given by equation (9). A convenient way of removing the noise bias in pixelized methods is to choose simply $\tilde{b} = b_i$, since it can be done after pixelizing, without ever going back to the individual galaxy positions; for a general quadratic combination of pixels q_i as in equation (46), using equation (18) shows that the shot noise bias can be written in terms of pixelized quantities alone, as

$$b_i = \mathrm{tr}\,[\mathbf{E}_i\mathbf{N}], \tag{A1}$$

where $\mathbf{N}$ is given by equation (19).

The strict minimum variance method

In the approximation that the shot noise fluctuations in the pixels has a Gaussian probability distribution, the above-mentioned method of choosing $\tilde{b}_i = b_i$ is readily shown to give the unbiased power estimator with the smallest variance. This is an excellent approximation when the number of galaxies is large, as we now show. The strict minimum-variance method is (Peebles 1980) to simply omit self-pairs in equation (4):

$$q_i - \tilde{b}_i = \sum_{\alpha\neq\beta}\frac{E_i(\mathbf{r}_\alpha,\mathbf{r}_\beta)}{\bar{n}(\mathbf{r}_\alpha)\bar{n}(\mathbf{r}_\beta)}. \tag{A2}$$

This corresponds to the shot noise correction

$$\tilde{b}_i = \int E_i(\mathbf{r},\mathbf{r})\frac{n(\mathbf{r})}{\bar{n}(\mathbf{r})^2}d^3r = \sum_\alpha\frac{E_i(\mathbf{r}_\alpha,\mathbf{r}_\alpha)}{\bar{n}(\mathbf{r}_\alpha)^2}, \tag{A3}$$

which is to be compared with equation (9). This is an unbiased method since $\langle\tilde{b}_i\rangle = b_i$. How much smaller is the variance with this approach? We illustrate this with the toy problem of estimating the power at $k = 0$ in a volume-limited survey with N galaxies, which is simply proportional to

$$q \equiv (N-\bar{N})^2 - \tilde{b}, \tag{A4}$$

where N is a Poisson-distributed random variable with mean $\bar{N}$. Since $\langle(N-\bar{N})^2\rangle = \bar{N}$, the shot noise correction $\tilde{b}_i = b_i$ corresponds to the choice $\tilde{b} = \bar{N}$. The strict minimum variance method of equation (A3) corresponds to $\tilde{b} = N$. Both methods are unbiased, giving $\langle q\rangle = 0$. The higher order moments differ, however. The variances $\langle q^2\rangle$ are $2\bar{N}^2$ and $2\bar{N}^2+\bar{N}$, respectively, so whereas the strictly optimal method gives the same variance that Gaussian noise would, the variance of the other method is a factor $(1+1/2\bar{N})$ larger. [12]

Which method is preferable?

This means that the optimal method only reduces the standard deviation by a negligible 0.01% for $N = 10^4$ galaxies. Moreover, it is incorrect to claim that the strictly optimal method in some sense removes the "exact" shot noise: since the higher order moments depart from the Gaussian values even for this method, there is clearly Poissonian noise left in q. The difference in variance between the two methods remain equally negligible for more realistic examples, generally being of order the inverse of the number of galaxies in the survey. The choice of which method to use should therefore be dictated by practical convenience. Whereas the strict minimum variance method is of course trivial to implement in techniques involving an explicit sum over galaxy pairs (such as the FKP method), the other method is generally simpler to use for pixelized techniques, since it can be implemented using the pixelized data alone.[13]

APPENDIX B: DERIVATION OF INTEGRAL CONSTRAINT AND RELATED EXPRESSIONS

In this appendix, we derive some of the results described in the text: the integral constraint correction for the traditional Fourier method (Section 3.3), and the optimal way to immunize the data from the untrustworthy modes of Section 5.

The Integral Constraint in the Direct Fourier Method

How should one deal with the integral constraint when using a traditional method as in Section 3.3? From the discussion in Section 5, it is clear that we should modify the weighting functions of equation (22) so that they become orthogonal to the mean density, *i.e.*, so that $\int\psi_i(\mathbf{r})d^3r = 0$. There are infinitely many ways of doing this, and some are clearly better than others if we want the power estimators $q_i = |x_i|^2$ to retain as much cosmological information as possible. We here derive one such correction method which is both simple and intuitive, following Tegmark (1997c). We will adopt a more ambitious approach in the following subsection, deriving the optimal correction method for the pixelized case. At the end of this Appendix, we show that in the small-scale limit, the two methods are in fact identical.

If we use our guess $\bar{n}_0$ (equation 74) in place of the unknown true selection function $\bar{n}$ in equation (15), we will have $\langle x_i\rangle = (a-1)\widehat{\psi}_i(0) \neq 0$. When using a traditional power estimator $q_i = |x_i|^2$, this causes a systematic positive power bias $(a-1)^2|\widehat{\psi}_i(0)|^2$ that we cannot subtract off, as a is unknown. We must therefore modify ψ_i so that $\widehat{\psi}_i(0)$ vanishes. Let $\widehat{a}$ denote our estimate of a. We will choose $\widehat{a}$ so that this bias vanishes, *i.e.*, so that the integral

[12] The differences between the two methods get larger for higher moments, but always remain of the order $1/\bar{N}$. Compared with the Gaussian approximation $\langle q^3\rangle = 8\bar{N}^3$, the skewness of the two methods is up by factors of $(1+2/\bar{N})$ and $(1+11/4\bar{N}+1/8\bar{N}^2)$, and compared with the Gaussian approximation $\langle q^4\rangle = 60\bar{N}^4$, the kurtosis is up by factors of $(1+12/5\bar{N}+2/15\bar{N}^2)$ and $(1+33/5\bar{N}+23/12\bar{N}^2+1/60\bar{N}^3)$.

[13] There is one useful exception: one can use the strict minimum variance method based on pixelized data alone in the special case where the pixels are counts in (sharp-edged) cells, in which case the terms x_i^2 get replaced by $x_i(x_i-1)$.

constraint

$$\int\left[\frac{n(\mathbf{r})}{\widehat{a}\bar{n}_0(\mathbf{r})}-1\right]\phi(\mathbf{r})d^3r=0 \tag{B1}$$

holds, or explicitly,

$$\widehat{a}\equiv\frac{1}{\widehat{\phi}(\mathbf{0})}\int\frac{n(\mathbf{r})}{\bar{n}_0(\mathbf{r})}\phi(\mathbf{r})d^3r. \tag{B2}$$

This is an unbiased estimator of the density normalization, since $\langle\widehat{a}\rangle=a$, the true value. Substituting $\bar{n}(\mathbf{r})=\widehat{a}\bar{n}_0(\mathbf{r})$ and equations (22) and (B2) into equation (15), we obtain

$$\begin{aligned}
x_i &= \frac{1}{\widehat{a}}\left[\int\frac{n(\mathbf{r})}{\bar{n}_0(\mathbf{r})}e^{i\mathbf{k}_i\cdot\mathbf{r}}\phi(\mathbf{r})d^3r-\widehat{\phi}(\mathbf{k}_i)\widehat{a}\right]\\
&= \frac{1}{\widehat{a}}\left[\int\frac{n(\mathbf{r})}{\bar{n}_0(\mathbf{r})}e^{i\mathbf{k}_i\cdot\mathbf{r}}\phi(\mathbf{r})d^3r-\frac{\widehat{\phi}(\mathbf{k}_i)}{\widehat{\phi}(\mathbf{0})}\int\frac{n(\mathbf{r})}{\bar{n}_0(\mathbf{r})}\phi(\mathbf{r})d^3r\right]\\
&= \frac{a}{\widehat{a}}\int\frac{n(\mathbf{r})}{\bar{n}(\mathbf{r})}\psi_i(\mathbf{r})d^3r\approx\int\frac{n(\mathbf{r})}{\bar{n}(\mathbf{r})}\psi_i(\mathbf{r})d^3r,
\end{aligned} \tag{B3}$$

where the function ψ_i is defined by

$$\psi_i(\mathbf{r})\equiv\left[e^{i\mathbf{k}_i\cdot\mathbf{r}}-\frac{\widehat{\phi}(\mathbf{k}_i)}{\widehat{\phi}(\mathbf{0})}\right]\phi(\mathbf{r}). \tag{B4}$$

Hence its Fourier transform is

$$\widehat{\psi}_i(\mathbf{k})=\widehat{\phi}(\mathbf{k}-\mathbf{k}_i)-\frac{\widehat{\phi}(\mathbf{k}_i)}{\widehat{\phi}(\mathbf{0})}\widehat{\phi}(\mathbf{k}). \tag{B5}$$

The relative error in $\widehat{a}$ is of order the inverse square root of the number of galaxies in the survey, so we can to a good approximation treat a as a known constant from here on and take $a/\widehat{a}=1$ on the last line of equation (B3). We see that the volume weighting ψ_i given by equation (B4) is better than the traditional choice of equation (22) since it is orthogonal to the mean, *i.e.*, it satisfies $\widehat{\psi}_i(\mathbf{0})=0$, which guarantees that $\langle x_i\rangle=0$.

In practice, we need never use equation (B4) to compute x_i with equation (15), since this is implicitly done if we first correct $\bar{n}$ by estimating its normalization with equation (B2) and then apply the simple weight function of equation (22). (This is mathematically equivalent to applying the weight function equation (B4) directly to the data, using an arbitrary $\bar{n}$-normalization.) However, we do need equation (B4) to derive the expressions for the shot noise correction and normalization given in Equations (32) and (31). Substituting equation (B4) into equation (9) gives the shot noise correction

$$\begin{aligned}
b_i &= \int\frac{|\psi_i(\mathbf{r})|^2}{\bar{n}(\mathbf{r})}d^3r\\
&= \int\left|e^{i\mathbf{k}_i\cdot\mathbf{r}}-\frac{\widehat{\phi}(\mathbf{k}_i)}{\widehat{\phi}(\mathbf{0})}\right|^2\frac{\phi(\mathbf{r})^2}{\bar{n}(\mathbf{r})}d^3r,
\end{aligned} \tag{B6}$$

and expanding the square completes our derivation of equation (32). The normalization coefficient A_i of equation (30) is determined by the requirement that the window function integrate to unity, *i.e.*, $A_i=\int|\widehat{\psi}_i(\mathbf{k})|^2d^3k/(2\pi)^3$. Using Parseval's theorem, we obtain

$$\begin{aligned}
A_i &= \int|\psi_i(\mathbf{r})|^2d^3r\\
&= \int\left|e^{i\mathbf{k}_i\cdot\mathbf{r}}-\frac{\widehat{\phi}(\mathbf{k}_i)}{\widehat{\phi}(\mathbf{0})}\right|^2\phi(\mathbf{r})^2d^3r,
\end{aligned} \tag{B7}$$

and expanding the square as above completes our derivation of equation (31).

How important is this correction?

Let us evaluate the integral constraint correction factor A_i for a couple of simple examples. We first note that for the special case of equation (23), we have $\phi(\mathbf{r})^2\propto\phi(\mathbf{r})$. Hence $a(\mathbf{k})\propto\widehat{\phi}(\mathbf{k})$, and equation (31) reduces to

$$A_i=\left(1-\left|\frac{\widehat{\phi}(\mathbf{k}_i)}{\widehat{\phi}(\mathbf{0})}\right|^2\right)a(\mathbf{0}), \tag{B8}$$

which we recognize as the result of Park *et al.* (1994). For volume-limited surveys, the prescriptions given by equations (23), (24) and (25) all coincide, so this expression is exact for the volume-limited case with these galaxy weighting schemes. For flux-limited surveys, on the other hand, these schemes all give a decreasing weight function ψ, since $\bar{n}$ decreases with distance. For the simple Gaussian case $\phi(\mathbf{r})=\exp[-(r/R)^2/2]/\pi^{1/4}R^{1/2}$, equation (31) gives

$$A_i=1+e^{-(Rk_i)^2}-2e^{-\frac{3}{4}(Rk_i)^2}, \tag{B9}$$

whereas the approximation (B8) gives

$$A_i=1-e^{-(Rk_i)^2}. \tag{B10}$$

A Taylor expansion shows that for $kR\ll 1$, the latter overestimates A_i by a factor of two as illustrated in Figure 2.

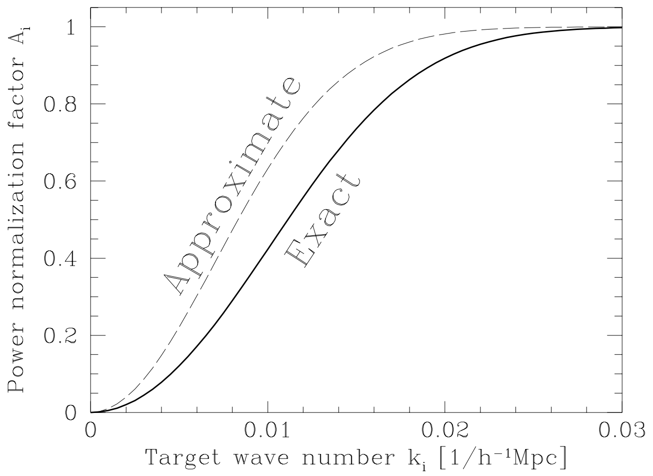

FIG. 2 — The exact expression for the integral constraint correction A_i is plotted together with the approximation of Park *et al* (1994) for a Gaussian weight function $\psi(\mathbf{r})\propto\exp[-(r/R)^2/2]$, $R=100h^{-1}$Mpc.

Eliminating contaminated modes with pixelized methods: A Simple Solution

In this and the next subsection, we continue the discussion of Section 5, and present a method to project out modes of the density field that we believe might be contaminated by, *e.g.*, errors in the selection function or in the assumed extinction map. Our starting point is equation (78), the $N \times M$ matrix $\mathbf{U}$ of untrustworthy modes.

To remedy the problem, we construct a new "cleaned" data set that is independent of $\mathbf{a}$, the coefficients of these modes. Let us define

$$\mathbf{x} \equiv \mathbf{\Pi x}', \tag{B11}$$

where $\mathbf{\Pi}$ is an $N \times N$ matrix satisfying

$$\mathbf{\Pi U} = \mathbf{0}, \tag{B12}$$

i.e., having the columns of $\mathbf{U}$ in its null space. This implies that $\mathbf{\Pi}$ has at most rank $N-M$. We will choose it to have exactly this rank, since otherwise $\mathbf{\Pi}$ will destroy more information than necessary (for instance, null matrix choice $\mathbf{\Pi}=\mathbf{0}$ satisfies equation (B12), but destroys all our information). It is easy to construct such matrices, a simple choice being

$$\mathbf{\Pi} = \mathbf{I} - \mathbf{U}(\mathbf{U}^\dagger\mathbf{U})^{-1}\mathbf{U}^\dagger. \tag{B13}$$

This is the Hermitean ($\mathbf{\Pi}^\dagger = \mathbf{\Pi}$) projection matrix ($\mathbf{\Pi}^2 = \mathbf{\Pi}$) projecting onto the subspace orthogonal to the columns of $\mathbf{U}$. Our corrected data set $\mathbf{x}$ satisfies equation (17), since $\langle\mathbf{x}\rangle = \mathbf{\Pi U a} = \mathbf{0}$. Letting $\mathbf{C}'$ denote the covariance matrix of the uncorrected data set, the corrected data will have the covariance matrix

$$\mathbf{C} \equiv \langle\mathbf{x}\mathbf{x}^\dagger\rangle = \mathbf{\Pi C}'\mathbf{\Pi}^\dagger. \tag{B14}$$

Once $\mathbf{x}$ and $\mathbf{C}$ have been computed, the rest of the pixelized analysis proceeds just as described in Sections 3.4, 3.5 or 3.6. The only complication is that $\mathbf{C}$ is now singular, having rank $N-M$ instead of N. As shown in the Appendix of T97, the correct way to deal with this in the quadratic method is to replace all occurrences of $\mathbf{C}^{-1}$ (which is of course undefined) by the "pseudo-inverse" of $\mathbf{C}$, defined as

$$\mathbf{\Pi}\left[\mathbf{C}+\gamma\mathbf{U}\mathbf{U}^\dagger\right]^{-1}\mathbf{\Pi} \tag{B15}$$

for some constant $\gamma \neq 0$. T97 shows that the result is independent of γ, and that a good choice for numerical stability is $\gamma \sim c/N$, where c is the order of magnitude of a typical matrix element of $\mathbf{C}$. The same trick can be used for the KL method, in the step where equation (40) is reduced to an ordinary eigenvalue problem by Cholesky decomposing $\mathbf{C}$ as described in TTH.

The optimal solution

Equation (B13) does not give the only rank $N-M$ projection matrix satisfying $\mathbf{\Pi U} = \mathbf{0}$ — there are in fact infinitely many such matrices of the form $\mathbf{\Pi} = \mathbf{I} - \mathbf{U}(\mathbf{U}^\dagger\mathbf{M}\mathbf{U})^{-1}\mathbf{U}^\dagger\mathbf{M}$, where $\mathbf{M}$ is an arbitrary non-singular matrix, and equation (B13) simply corresponds to the case $\mathbf{M} \propto \mathbf{I}$. Since they all have the same null space $\mathbf{U}$, it is clear that they all destroy the same information (all the information about the untrustworthy modes, no more and no less). The power spectrum Fisher matrix for the cleaned data set is therefore independent of $\mathbf{M}$, so our choice is purely one of numerical convenience. For the quadratic method in particular, there turns out to be a much more appropriate choice than that of equation (B13), which altogether eliminates the above-mentioned problem of $\mathbf{C}$ being singular by allowing the quadratic pair weighting $\mathbf{E}$ to be computed analytically. We will derive this choice of $\mathbf{\Pi}$ by generalizing the derivation of the quadratic method (T97) to include our constraint that the results be independent of the corrupted modes.

The most general quadratic estimator can clearly be written as in equation (46) for some symmetric matrix $\mathbf{E}_i$. As shown in T97, this implies that the variance of q_i is given by

$$V(q_i) \equiv \langle q_i^2\rangle - \langle q_i\rangle^2 = \frac{1}{2}\mathrm{tr}\,[\mathbf{C}\mathbf{E}_i\mathbf{C}\mathbf{E}_i] \tag{B16}$$

and that the signal, the expected contribution to q_i from the power band of interest, is $\mathrm{tr}\,[\mathbf{C}_{,i}\,\mathbf{E}_i]/2$, where $\mathbf{C}_{,i}$ is defined by equation (45). To maximize the signal-to-noise ratio, we want to minimize the variance given a fixed signal, *i.e.*, subject to the constraint that $\mathrm{tr}\,[\mathbf{C}_{,i}\,\mathbf{E}_i]$ is held constant. If we write

$$q_i = \frac{1}{2}(\mathbf{x}+\mathbf{U}\mathbf{a})^\dagger\mathbf{E}_i(\mathbf{x}+\mathbf{U}\mathbf{a}), \tag{B17}$$

it is clear that we can phrase our constraint on $\mathbf{E}_i$ as $\mathbf{E}_i\mathbf{U} = \mathbf{0}$. This is in fact $N \times M$ separate constraint equations, so using the fact that $\mathbf{E}_i$ is symmetric, our constrained minimization problem involves minimizing

$$L \equiv \mathrm{tr}\,\left\{\frac{1}{2}\mathbf{C}\mathbf{E}_i\mathbf{C}\mathbf{E}_i - \lambda\mathbf{C}_{,i}\,\mathbf{E}_i + \lambda[\mathbf{U}\mathbf{A}^\dagger + \mathbf{A}\mathbf{U}^\dagger]\mathbf{E}_i\right\}, \tag{B18}$$

where $\mathbf{A}$ is some arbitrary $N \times M$ matrix of Lagrange multipliers. Requiring the derivatives with respect to the components of $\mathbf{E}_i$ to vanish, we obtain

$$\mathbf{E}_i \propto \mathbf{C}^{-1}\left[\mathbf{C}_{,i} - \mathbf{U}\mathbf{A}^\dagger - \mathbf{A}\mathbf{U}^\dagger\right]\mathbf{C}^{-1}, \tag{B19}$$

where $\mathbf{A}$ is determined by the constraint $\mathbf{E}_i\mathbf{U} = \mathbf{0}$. Defining $\tilde{\mathbf{U}} \equiv \mathbf{C}^{-1}\mathbf{U}$, the solution is

$$\mathbf{A} = \left[\mathbf{I} - \frac{1}{2}\mathbf{U}(\tilde{\mathbf{U}}^\dagger\mathbf{U})^{-1}\tilde{\mathbf{U}}^\dagger\right]\mathbf{C}_{,i}\,\tilde{\mathbf{U}}(\tilde{\mathbf{U}}^\dagger\mathbf{U})^{-1}, \tag{B20}$$

which can be verified by direct substitution. Substituting this back into equation (B19) finally yields

$$\mathbf{E}_i \propto \mathbf{\Pi}^\dagger\mathbf{C}^{-1}\mathbf{C}_{,i}\,\mathbf{C}^{-1}\mathbf{\Pi}, \tag{B21}$$

where

$$\mathbf{\Pi} = \mathbf{I} - \mathbf{U}(\mathbf{U}^\dagger\tilde{\mathbf{U}})^{-1}\tilde{\mathbf{U}}^\dagger = \mathbf{I} - \mathbf{U}(\mathbf{U}^\dagger\mathbf{C}^{-1}\mathbf{U})^{-1}\mathbf{U}^\dagger\mathbf{C}^{-1} \tag{B22}$$

is a projection matrix satisfying $\mathbf{\Pi U} = \mathbf{0}$, $\mathbf{\Pi}^\dagger\tilde{\mathbf{U}} = \mathbf{0}$ and $\mathbf{C}^{-1}\mathbf{\Pi} = \mathbf{\Pi}^\dagger\mathbf{C}^{-1}$. Inserting equation (B21) into equation (46), we see that our quadratic estimator retains the simple form of equation (43) if we generalize equation (44) to

$$\mathbf{z} \equiv \mathbf{C}^{-1}\mathbf{\Pi x}, \tag{B23}$$

so after cleaning the data set (replacing $\mathbf{x}$ by $\mathbf{\Pi x}$), the quadratic method proceeds exactly as before. This generalization of the quadratic method clearly reduces to the prescription in Section 3.6 ($\mathbf{z} = \mathbf{C}^{-1}\mathbf{x}$) if there are no untrustworthy modes, in which case $M = 0$ and $\mathbf{\Pi} = \mathbf{I}$. Note that this technique is quite useful for estimating the power spectrum from cosmic microwave background experiments as well, in which case obvious candidates for corrupted modes are the monopole and the three dipole components.

Relation between the pixelized and continuous cleaning schemes

In this section, we will show that the integral correction procedure for traditional methods that we derived above is lossless for the FKP volume weighting in the small-scale limit, corresponding to the quadratic method.

When we have merely one untrustworthy mode ($M = 1$) corresponding to the normalization of $\bar{n}$, the matrix defined by equation (78) consists of a single column vector; $\mathbf{U} = \mathbf{u}$. Using the continuum pixels $x(\mathbf{r})$ of equation (64), this vector is simply the $\mathbf{k} = \mathbf{0}$ (constant) mode, *i.e.*,

$$u(\mathbf{r}) = 1. \tag{B24}$$

Let us now evaluate the optimal power estimate q_i that we derived above, given by Equations (43) and (B23). Since $\mathbf{U} = \mathbf{u}$, we have

$$\mathbf{\Pi x} = \mathbf{x} - \left(\frac{\mathbf{u}^\dagger \mathbf{C}^{-1}\mathbf{x}}{\mathbf{u}^\dagger \mathbf{C}^{-1}\mathbf{u}}\right)\mathbf{u}, \tag{B25}$$

where Equations (68) and (B24) show that $\mathbf{u}^\dagger \mathbf{C}^{-1}\mathbf{x} = \int \phi(\mathbf{r})x(\mathbf{r})d^3r$ and $\mathbf{u}^\dagger \mathbf{C}^{-1}\mathbf{u} = \int \phi(\mathbf{r})d^3r = \widehat{\phi}(0)$. We can thus write

$$z(\mathbf{r}) = (\mathbf{C}^{-1}\mathbf{\Pi x})(\mathbf{r}) = \int \left[\delta^D(\mathbf{r}-\mathbf{r}') - \frac{\phi(\mathbf{r})}{\widehat{\phi}(0)}\right]\phi(\mathbf{r}')x(\mathbf{r}')d^3r'. \tag{B26}$$

Fourier transforming this with respect to $\mathbf{r}'$, Equation (69) now shows that $q_i = |x_i|^2/2$, where

$$x_i = \widehat{z}(\mathbf{k}_i) = \int \psi_i(\mathbf{r})x(\mathbf{r})d^3r, \tag{B27}$$

and the volume weighting function ψ_i is exactly the one that we derived in Section B.1, given by equation (B4). In other words, we have shown that the FKP choice of the function ϕ together with the volume weighting of equation (B4) is identical to the lossless quadratic method in the small-scale limit.

APPENDIX C: REDSHIFT DISTORTIONS AND CLUSTERING EVOLUTION

In Section 5 and Appendix B, we showed that the pixelized methods allow a more ambitious approach than is feasible with the direct Fourier methods, incorporating multiple integral constraints, since all the complications simply became buried in the appropriate matrices. In this Appendix, we show how two additional complications can be incorporated with pixelized methods in the same vein: redshift distortions and clustering evolution.

Redshift distortions

A ubiquitous problem with power spectrum estimation is that of "redshift distortions". When estimating the distance to a galaxy by its redshift, galaxies receding faster than the Hubble flow due to local gravitational interactions appear to be further away than they really are, and vice versa. This was first discussed by Kaiser (1987) in the context of $P(k)$, and a recent review is given by Hamilton (1997c). Denoting the apparent density field in redshift space $\delta_s(\mathbf{r})$, Hamilton & Culhane (1996) use Kaiser's formalism to show in linear perturbation theory that

$$\delta_s = \left[1 + \beta\left(\frac{\partial^2}{\partial r^2} + \frac{\alpha(\mathbf{r})}{r}\frac{\partial}{\partial r}\right)\nabla^{-2}\right]\delta_r, \tag{C1}$$

where $\beta \equiv \Omega^{0.6}/b$, the constant b is the so-called linear bias factor, and

$$\alpha(\mathbf{r}) \equiv 2 + \frac{\partial \ln \bar{n}(\mathbf{r})}{\partial \ln r} \tag{C2}$$

is two plus the logarithmic slope of the radial selection function. Fourier transforming this gives (Hamilton & Culhane 1996; Hamilton 1997c)

$$\widehat{\delta}_s(\mathbf{k}) = \widehat{\delta}_r(\mathbf{k}) + \beta \int f(\mathbf{k},\mathbf{k}')\widehat{\delta}_r(\mathbf{k}')\frac{d^3k'}{(2\pi)^3}, \tag{C3}$$

where the function f is defined by

$$f(\mathbf{k},\mathbf{k}') \equiv \int e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{r}}\left[(\widehat{\mathbf{k}}'\cdot\widehat{\mathbf{r}})^2 - \frac{\alpha(\mathbf{r})i}{k'r}(\widehat{\mathbf{k}}'\cdot\widehat{\mathbf{r}})\right]d^3r. \tag{C4}$$

Thus we obtain

$$\langle\widehat{\delta}_s(\mathbf{k})^*\widehat{\delta}_s(\mathbf{k}')\rangle = (2\pi)^3\int g(\mathbf{k},\mathbf{k}',\mathbf{k}'')P(\mathbf{k}'')d^3k'', \tag{C5}$$

where

$$\begin{aligned} g(\mathbf{k},\mathbf{k}',\mathbf{k}'') &\equiv \delta^D(\mathbf{k}-\mathbf{k}'')\delta^D(\mathbf{k}'-\mathbf{k}'') + \\ &+ \beta[\delta^D(\mathbf{k}-\mathbf{k}'')f(\mathbf{k}',\mathbf{k}) + \delta^D(\mathbf{k}'-\mathbf{k}'')f(\mathbf{k},\mathbf{k}')^*] + \\ &+ \beta^2 f(\mathbf{k},\mathbf{k}'')^* f(\mathbf{k}',\mathbf{k}''). \end{aligned} \tag{C6}$$

The above expressions are derived and discussed in detail by Zaroubi & Hoffman (1996), and also in Tegmark & Bromley (1995) and T95 for the volume limited case; see Szalay, Matsubara, & Landy (1997) for further discussion. The key point here is that although $\langle\widehat{\delta}_s(\mathbf{k})^*\widehat{\delta}_s(\mathbf{k}')\rangle$ is no longer diagonal, and rather messy, it is still linear in the power spectrum. Thus the pixel covariance matrix $\mathbf{C}$ will still be some shot noise term plus a term *linear* in $P(\mathbf{k})$. In other words, by letting $\mathbf{C}_{,i}$ refer to the derivative of $\mathbf{C}$ with respect to the band powers in *real* space instead of redshift space, the quadratic method will measure the real space power spectrum directly (given *a priori* knowledge of β), and the corresponding window functions (the rows of $\mathbf{F}$, say) will show the contributions to the measurements q_i from the various real space power bands.

Clustering evolution

The density fluctuation field δ_r maintains its shape in linear perturbation theory, simply increasing in amplitude by a position-independent growth factor D. Since we are seeing distant galaxies at an earlier time, we see the apparent density fluctuations

$$\delta_a(\mathbf{r}) \equiv D(r)\delta_r(\mathbf{r}), \qquad \text{(C7)}$$

where $D(r) = 1/(1+z)$ for $\Omega = 1$. This effect is straightforward to include in a pixelized analysis. Equation (19) remains unchanged and equation (20) simply gets replaced by

$$\mathbf{S}_{ij} = \int \widehat{\psi}_i'(\mathbf{k})\widehat{\psi}_j'(\mathbf{k})^* P(k)\frac{d^3k}{(2\pi)^3}, \qquad \text{(C8)}$$

where we have defined the functions $\psi_i'(\mathbf{r}) \equiv D(r)\psi_i(\mathbf{r})$. This correction is quite small for shallow galaxy surveys, where $\bar{n}$ typically varies dramatically between $z = 0$ and $z = 0.2$, a range over which D changes by at most about 20%, less for small Ω. If this effect is incorporated into the analysis, Ω can be made a free parameter to be fit for in the pipeline.

This clustering evolution should not be confused with galaxy evolution, which we do not discuss here, and which affects only $\bar{n}$, not δ_r.

APPENDIX D: NORMALIZATION CONVENTIONS

Unfortunately, the power spectrum $P(k)$ is defined in many different ways in the literature, differing by normalization factors such as $(2\pi)^3$ and a fiducial box volume V. In this paper, we normalize Fourier transforms as

$$\hat{f}(\mathbf{k}) \equiv \int f(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}d^3r, \qquad \text{(D1)}$$

and normalize $P(k)$ so that

$$\langle\widehat{\delta}_r(\mathbf{k})^*\widehat{\delta}_r(\mathbf{k}')\rangle = (2\pi)^3\delta^D(\mathbf{k}-\mathbf{k}')P(k). \qquad \text{(D2)}$$

The units of $P(k)$ are volume. With this normalization, the dimensionless power Δ^2 of Peacock & Dodds (1994) is given by

$$\Delta^2(k) = \frac{4\pi}{(2\pi)^3}k^3P(k), \qquad \text{(D3)}$$

the r.m.s. fluctuations σ_8 in a sphere of radius $R = 8h^{-1}$ Mpc are

$$\sigma_8^2 = 4\pi\int_0^\infty\left[\frac{\sin x - x\cos x}{x^3/3}\right]^2 P(k)\frac{k^2dk}{(2\pi)^3}, \qquad \text{(D4)}$$

where $x \equiv kR$, and the Sachs-Wolfe quadrupole Q in the cosmic microwave background is given by

$$Q^2 \equiv \frac{5}{4\pi}C_2 = \frac{10}{\pi^2}\int_0^\infty\frac{j_2(x)^2}{x^4}P(k)k^2dk, \qquad \text{(D5)}$$

where $x \approx 2kc/H_0 \approx k\times 6000h^{-1}$ Mpc and

$$j_2(x) = \frac{3\sin x - 3x\cos x - x^2\sin x}{x^3}. \qquad \text{(D6)}$$

T95 uses a convention where the $(2\pi)^3$ factor in equation (D2) is replaced by $(2\pi)^6$.

Tegmark et al 1997

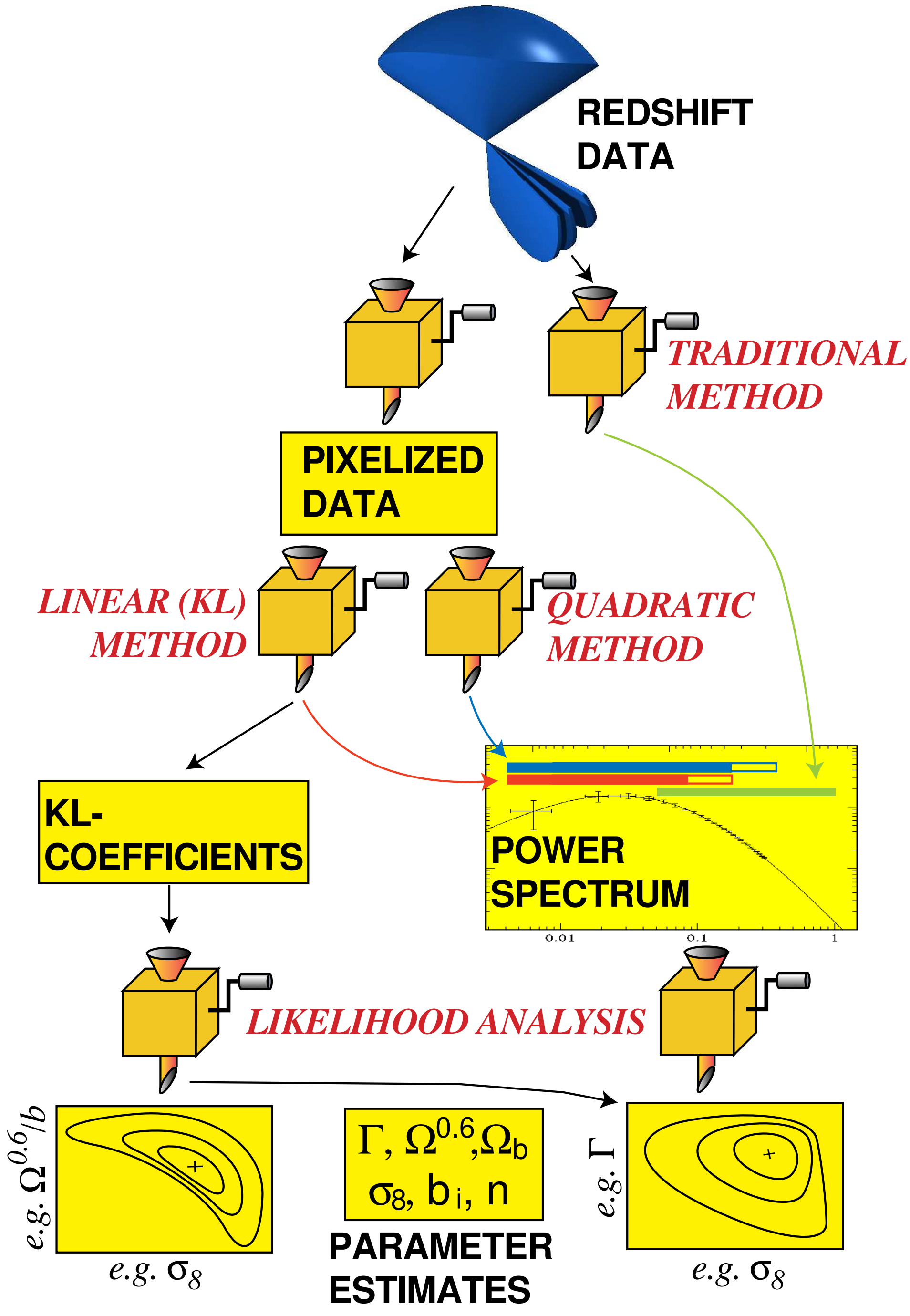

Fig. D1.— We propose analyzing large future galaxy redshift surveys such as the SDSS, using three techniques in conjunction: a traditional Fourier approach on small scales, a pixelized quadratic matrix method on large scales and a pixelized Karhunen-Loève eigenmode analysis to probe anisotropic effects such as redshift-space distortions and residual extinction. The horizontal bars in the power spectrum box indicate that the quadratic method has a larger dynamic range than the KL method. The bottom of the figure indicates that numbers such as the redshift distortion parameter $\Omega^{0.6}/b$ which reflect anisotropic clustering can only be optimally constrained using the KL modes, which retain not merely the overall power, but the phase information as well.