May 5, 1997.
To appear in "Ringberg Workshop on Large-Scale Structure",
ed. D. Hamilton (Kluwer, Amsterdam)

arXiv:astro-ph/9708021v1 4 Aug 1997

ANALYZING REDSHIFT SURVEYS TO MEASURE THE POWER SPECTRUM ON LARGE SCALES

MAX TEGMARK[5]
Institute for Advanced Study
Princeton, NJ 08540; max@ias.edu

Max-Planck-Institut für Physik
Föhringer Ring 6, D-80805 München

AND

Max-Planck-Institut für Astrophysik
Karl-Schwarzschild-Str. 1, D-85740 Garching

Abstract. Upcoming large redshift surveys potentially allow precision measurements of the galaxy power spectrum. To accurately measure $P(k)$ on the largest scales, comparable to the depth of the survey, it is crucial that finite volume effects are accurately corrected for in the data analysis. Here we derive analytic expressions for the one such effect that has not previously been worked out exactly: that of the so-called integral constraint. We also show that for data analysis methods based on counts in cells, multiple constraints can be included via simple matrix operations, thereby rendering the results less sensitive to galactic extinction and misestimates of the shape of the radial selection function.

1. Introduction

Observational data on galaxy clustering are rapidly increasing in both quantity and quality, which brings new challenges when it comes to data analysis. As to quantity, redshifts had been published for a few thousand galaxies 15 years ago. Today the number is $\sim 10^5$, and ongoing projects such as the AAT 2dF Survey and the Sloan Digital Sky Survey (SDSS) will raise it to 10^6 within a few years. Comprehensive reviews of past redshift surveys are

[5]Hubble Fellow

given by *e.g.* Efstathiou (1994), Vogeley (1995), Strauss & Willick (1995) and Strauss (1996), the last also including a detailed description of 2dF and SDSS. As to 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, without bias. Here we will focus on how to correct for the finite volume of a survey. As is well known, this causes the measured power spectrum to be a convolution of the true power spectrum with some window function which depends on the survey geometry and the data analysis method used. Exact expressions have been derived (see *e.g.* Feldman, Kaiser & Peacock 1994, hereafter "FKP") for the window function and its normalization for the case where the number density of galaxies is assumed to be known *a priori*, but the more realistic case where the mean galaxy density is determined from the survey itself has thus far only been treated approximately (Peacock & Nicholson 1991; Park *et al.* 1994). The main purpose of this paper is to derive exact expressions for this important correction.

The methods for power spectrum estimation that have been proposed in the literature fall into two categories:

1. Direct Fourier methods
2. Pixelized methods

The direct Fourier methods make use of the exact position of each galaxy, whereas the other methods start by "pixelizing" the data set (by computing counts in cells or expansion coefficients for some set of functions), thereby reducing the problem to manipulating large vectors and matrices. In Section 2, we will derive the finite-volume correction for direct Fourier methods. The corresponding correction for pixelized methods is given in Section 3.

2. Finite Volume Correction for Direct Fourier Methods

2.1. THE POWER SPECTRUM ESTIMATION PROBLEM

It is customary (see *e.g.* FKP) to model the observed galaxy distribution as a 3D Poisson process $n(\mathbf{r}) = \sum_i \delta(\mathbf{r} - \mathbf{r}_i)$ with intensity $\lambda(\mathbf{r}) = \bar{n}(\mathbf{r})[1 + \delta_r(\mathbf{r})]$. The function $\bar{n}$ is the selection function of the galaxy survey under consideration, *i.e.*, $\bar{n}(\mathbf{r})dV$ is the expected (not the observed) number of galaxies in a volume dV about $\mathbf{r}$. The density fluctuations δ_r are modeled as a homogeneous and isotropic (but not necessarily Gaussian) random field with power spectrum $P(k)$, and the power spectrum estimation problem is to estimate $P(k)$ given a realization of $n(\mathbf{r})$.

2.2. THE DIRECT FOURIER APPROACH

Due to space limitations, the method summary below is very brief, and the interested reader is referred to FKP and Tegmark (1995, hereafter "T95") for more detailed introductions to the various methods.

All direct Fourier methods not involving random numbers[1] are specified by choosing a weight function $\psi(\mathbf{r})$ in real space and a set of weights w_i in Fourier space, as defined below. They all involve the following two steps:

1. At a grid of points $\mathbf{k}_i$ in Fourier space, fluctuation amplitudes are estimated by

$$\begin{aligned}\widehat{F}(\mathbf{k}_i) &\equiv \int \left[\frac{n(\mathbf{r})}{\bar{n}(\mathbf{r})} - 1\right] \psi(\mathbf{r}) e^{-i\mathbf{k}_i\cdot\mathbf{r}} d^3r \\ &= \sum_j \frac{\psi(\mathbf{r}_j)}{\bar{n}(\mathbf{r}_j)} e^{-i\mathbf{k}_i\cdot\mathbf{r}_j} - \widehat{\psi}(\mathbf{k}_i). \qquad (1)\end{aligned}$$

   (Here and throughout, hats denote Fourier transforms.)

2. The power P at some given k-value, say k_*, is estimated by squaring these fluctuation amplitudes, subtracting off their shot noise bias, rescaling them to correct for the integral constraint, and averaging them with some weights w_i that add up to unity:

$$\tilde{P}(k_*) \equiv \sum_i w_i \left[\frac{|\widehat{F}(\mathbf{k}_i)|^2 - \sigma_s^2(\mathbf{k}_i)}{N(\mathbf{k}_i)}\right]. \qquad (2)$$

As we will show in Section 2.6, the new and exact expressions for the shot noise and integral constraint corrections (when $\bar{n}$ is normalized so that $\widehat{F}(\mathbf{0}) = 0$) are

$$\sigma_s^2(\mathbf{k}) = \left(1 + \left|\frac{\widehat{\psi}(\mathbf{k})}{\widehat{\psi}(\mathbf{0})}\right|^2\right) c_s(0) - 2\,\mathrm{Re}\left\{\frac{\widehat{\psi}(\mathbf{k})^*}{\widehat{\psi}(\mathbf{0})^*} c_s(\mathbf{k})\right\}, \qquad (3)$$

$$N(\mathbf{k}) = \left(1 + \left|\frac{\widehat{\psi}(\mathbf{k})}{\widehat{\psi}(\mathbf{0})}\right|^2\right) f(0) - 2\,\mathrm{Re}\left\{\frac{\widehat{\psi}(\mathbf{k})^*}{\widehat{\psi}(\mathbf{0})^*} f(\mathbf{k})\right\}, \qquad (4)$$

where the functions c_s and f are defined by

$$c_s(\mathbf{k}) \equiv \int \frac{\psi(\mathbf{r})^2}{\bar{n}(\mathbf{r})} e^{-i\mathbf{k}\cdot\mathbf{r}} d^3r, \qquad (5)$$

$$f(\mathbf{k}) \equiv \int \psi(\mathbf{r})^2 e^{-i\mathbf{k}\cdot\mathbf{r}} d^3r. \qquad (6)$$

[1]Including a random mock survey as in equation (2.1.3) in FKP can never give minimal error bars, since inclusion of random numbers will always increase the variance of the estimator.

If the survey is volume limited, then $\bar{n}$ is independent of $\mathbf{r}$, $c_s(\mathbf{k}) = f(\mathbf{k})/\bar{n}$, and $\sigma_s^2(\mathbf{k})/N(\mathbf{k}) = 1/\bar{n}$.

2.3. WEIGHTING THE GALAXIES

Four different choices of the galaxy weighting function ψ have appeared in the literature:

$$\psi(\mathbf{r}) = \begin{cases} 1 \text{ inside survey volume} \\ 0 \text{ outside survey volume} \end{cases} \tag{7}$$

$$\psi(\mathbf{r}) = \bar{n}(\mathbf{r}) \tag{8}$$

$$\psi(\mathbf{r}) = \frac{\bar{n}(\mathbf{r})}{1+\bar{n}(\mathbf{r})P} \tag{9}$$

$$\psi(\mathbf{r}) = \text{eigenfunction of } \left[\nabla^2 - \frac{\gamma}{\bar{n}(\mathbf{r})}\right]. \tag{10}$$

The first choice, *i.e.*, weighing all galaxies in a survey volume equally, was employed by *e.g.* Fisher *et al.* (1993). The second choice was used for *e.g.* the APM survey (Baugh & Efstathiou 1994) — since redshifts were not measured, the radial galaxy weighting by default became the selection function (moreover, modes could of course only be computed in the directions perpendicular to the line of sight). The third choice is that advocated by Feldman, Kaiser & Peacock (1994, hereafter FKP), where P denotes an *a priori* guess as to the power in the band under consideration, and minimizes the variance in the limit when $k^{-1} \ll$ the depth of the survey. The fourth choice corresponds to the method of T95, and gives the narrowest window function for a given variance (the constant γ determines the tradeoff).

2.4. WEIGHTING THE FOURIER MODES

As to the weights in Fourier space, w_i, a common choice (*e.g.* FKP) is to perform a straight average of all modes in a spherical shell with its radius centered on k_*, although when the survey volume is anisotropic, a weighted average giving smaller error bars can generally be obtained by solving a quadratic programming problem (T95).

2.5. WINDOW FUNCTIONS

The expectation value of a power estimate $\tilde{P}$ that has been corrected for the shot noise bias and the integral constraint can always be written as

$$\langle \tilde{P} \rangle = \int W(k)P(k)dk, \tag{11}$$

where the function W, known as the *window function*, has the property that

$$\int_0^\infty W(k)dk = 1. \tag{12}$$

We can therefore think of $\tilde{P}(k_*)$ as measuring a weighted average of the true power spectrum, with W specifying the weights (for most methods, but not all, these weights are strictly non-negative as well). The window function for a general direct Fourier method is derived in Section 2.6, and is found to be

$$\mathbf{W}(k) \propto \sum_i w_i \int |\widehat{\psi}_i(\mathbf{k})|^2 k^2 d\Omega_k, \tag{13}$$

where ψ_i is given by equation (18) and the angular $\mathbf{k}$-integral is carried out over a spherical shell of radius k. In the limit where $k^{-1} \ll L$, where L is the smallest survey dimension, the 3D window function simplifies to $\mathbf{W}(k) \propto \sum_i w_i \int |\widehat{\psi}(\mathbf{k} - \mathbf{k}_i)|^2 k^2 d\Omega_k$.

2.6. DERIVATION OF THE INTEGRAL CONSTRAINT CORRECTION

If we knew the selection function $\bar{n}(\mathbf{r})$ *a priori*, before counting the galaxies in our survey, we would be able to probe the power on the largest scales. For modes of wavelength much larger than the survey volume, this would essentially correspond to counting the difference between the observed and expected number of galaxies in our sample. Of course, we do not know $\bar{n}$ *a priori*, so our most accurate way of normalizing the selection function is by using the galaxies in the survey itself. When $\bar{n}$ is normalized in this way, naive application of equation (1) will give the artifact $\widehat{F}(\mathbf{k}) \to 0$ as $\mathbf{k} \to 0$ because fluctuations on the scale of the survey are forced to zero by definition (Peacock & Nicholson 1991).

Let us assume that we know the shape of the selection function but not its normalization. To reflect this, we write

$$\bar{n}(\mathbf{r}) = \eta \bar{n}_0(\mathbf{r}), \tag{14}$$

where $\bar{n}_0$ is our guess as to the shape and η is an unknown normalization constant. If we had used $\bar{n}_0$ in place of the correct $\bar{n}$ in equation (1), we would in general not obtain the desired result $\langle \widehat{F}(\mathbf{k}_i) \rangle = 0$ but rather $\langle \widehat{F}(\mathbf{k}_i) \rangle = (\eta - 1)\widehat{\psi}(\mathbf{k})$, which would enter equation (2) as a systematic positive power bias. It is the need to eliminate this problem that forces us to impose an integral constraint. Let $\widehat{\eta}$ denote our estimate of η. We will choose $\widehat{\eta}$ so that this bias vanishes, *i.e.*, so that the integral constraint

$$\int \left[\frac{n(\mathbf{r})}{\widehat{\eta}\bar{n}_0(\mathbf{r})} - 1 \right] \psi(\mathbf{r}) d^3r = 0 \tag{15}$$

holds, or explicitly,

$$\widehat{\eta} \equiv \frac{1}{\widehat{\psi}(\mathbf{0})} \int \frac{n(\mathbf{r})}{\bar{n}_0(\mathbf{r})} \psi(\mathbf{r}) d^3r = \frac{1}{\widehat{\psi}(\mathbf{0})} \sum_j \frac{\psi(\mathbf{r}_j)}{\bar{n}_0(\mathbf{r}_j)}. \tag{16}$$

This is an unbiased estimator of the density normalization, since $\langle\widehat{\eta}\rangle = \eta$, the true value. Substituting $\bar{n}(\mathbf{r}) = \widehat{\eta}\bar{n}_0(\mathbf{r})$ and equation (16) into equation (1), we obtain

$$\begin{aligned} \widehat{F}(\mathbf{k}_i) &= \frac{1}{\widehat{\eta}} \left[\int \frac{n(\mathbf{r})}{\bar{n}_0(\mathbf{r})} e^{-i\mathbf{k}_i\cdot\mathbf{r}} \psi(\mathbf{r}) d^3r - \widehat{\psi}(\mathbf{k}_i)\widehat{\eta} \right] \\ &= \frac{1}{\widehat{\eta}} \left[\int \frac{n(\mathbf{r})}{\bar{n}_0(\mathbf{r})} e^{-i\mathbf{k}_i\cdot\mathbf{r}} \psi(\mathbf{r}) d^3r - \frac{\widehat{\psi}(\mathbf{k}_i)}{\widehat{\psi}(\mathbf{0})} \int \frac{n(\mathbf{r})}{\bar{n}_0(\mathbf{r})} \psi(\mathbf{r}) d^3r \right] \\ &= \frac{\eta}{\widehat{\eta}} \int \frac{n(\mathbf{r})}{\bar{n}(\mathbf{r})} \psi_i(\mathbf{r}) d^3r \approx \int \frac{n(\mathbf{r})}{\bar{n}(\mathbf{r})} \psi_i(\mathbf{r}) d^3r, \end{aligned} \tag{17}$$

where the function ψ_i is defined by

$$\psi_i(\mathbf{r}) \equiv \left[e^{-i\mathbf{k}_i\cdot\mathbf{r}} - \frac{\widehat{\psi}(\mathbf{k}_i)}{\widehat{\psi}(\mathbf{0})} \right] \psi(\mathbf{r}). \tag{18}$$

Since we will have $\widehat{\eta} \approx \eta$ with a relative accuracy $\Delta\widehat{\eta}/\eta$ of order $1/\sqrt{N}$, where N is the number of galaxies in the survey, we can to a good approximation treat η as a known constant from here on and take $\eta/\widehat{\eta} = 1$ on the last line of equation (17). Since $\widehat{\psi}_i(\mathbf{0}) = 0$, we now have $\langle\widehat{F}(\mathbf{k}_i)\rangle = 0$, so we see that we have succeeded in eliminating the above-mentioned power bias. The price for this is slightly more complicated equations. Let us now derive the expressions for the shot noise correction and normalization given in Equations (3) and (4).

Since $\widehat{\eta} \approx \eta$, we substitute the last expression of equation (17) into Equation (3) of T95, treating $\bar{n} = \eta\bar{n}_0$ as a known function, which gives

$$\langle|\widehat{F}(\mathbf{k}_i)|^2\rangle = \frac{1}{(2\pi)^3} \int |\widehat{\psi}_i(\mathbf{k})|^2 P(k) d^3k + \int \frac{|\psi_i(\mathbf{r})|^2}{\bar{n}(\mathbf{r})} d^3r, \tag{19}$$

Comparing this with equation (11) and equation (2), we identify the three-dimensional window function as

$$W(\mathbf{k}) \propto |\widehat{\psi}_i(\mathbf{k})|^2 \tag{20}$$

and see that the shot noise correction is

$$\begin{aligned} \sigma_s^2(\mathbf{k}_i) &= \int \frac{|\psi_i(\mathbf{r})|^2}{\bar{n}(\mathbf{r})} d^3r \\ &= \int \left| e^{-i\mathbf{k}_i\cdot\mathbf{r}} - \frac{\widehat{\psi}(\mathbf{k}_i)}{\widehat{\psi}(\mathbf{0})} \right|^2 \frac{\psi(\mathbf{r})^2}{\bar{n}(\mathbf{r})} d^3r, \end{aligned} \tag{21}$$

and expanding the square completes our derivation of equation (3). Performing an angular integral of equation (20) completes the proof of equation (13). The normalization coefficient $N(\mathbf{k}_i)$ of equation (2) is determined by the requirement that the window function integrate to unity, *i.e.*, $N(\mathbf{k}_i) = \int |\widehat{\psi}_i(\mathbf{k})|^2 d^3k/(2\pi)^3$. Using Parseval's theorem, we obtain

$$\begin{aligned} N(\mathbf{k}_i) &= \int |\psi_i(\mathbf{r})|^2 d^3r \\ &= \int \left| e^{-i\mathbf{k}_i\cdot\mathbf{r}} - \frac{\widehat{\psi}(\mathbf{k}_i)}{\widehat{\psi}(\mathbf{0})} \right|^2 \psi(\mathbf{r})^2 d^3r, \end{aligned} \tag{22}$$

and expanding the square as above completes our derivation of equation (4).

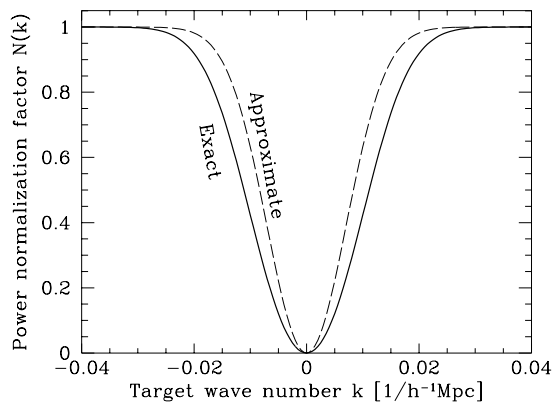

Figure 1. The exact expression for $N(k)$ is plotted together with the approximation of Park *et al* for a Gaussian weight function $\psi(\mathbf{r}) \propto \exp[-(r/R)^2/2]$, $R = 100h^{-1}$Mpc.

2.7. HOW IMPORTANT IS THIS CORRECTION?

Let us evaluate the integral constraint correction factor $N(\mathbf{k})$ for a couple of simple examples. We first note that for the special case of equation (7), we have $\psi(\mathbf{r})^2 \propto \psi(\mathbf{r})$. Hence $f(\mathbf{k}) \propto \widehat{\psi}(\mathbf{k})$, and equation (4) reduces to

$$N(\mathbf{k}) = \left(1 - \left|\frac{\widehat{\psi}(\mathbf{k})}{\widehat{\psi}(\mathbf{0})}\right|^2\right) f(0), \tag{23}$$

which we recognize as the approximation of Park *et al.* (1994). For volume-limited surveys, the prescriptions given by equations (8), (9) and (10) all coincide, so we see that this approximation becomes 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 $\psi(\mathbf{r}) = \exp[-(r/R)^2/2]/\pi^{1/4}R^{1/2}$, equation (4) gives

$$N(\mathbf{k}) = 1 + e^{-(Rk)^2} - 2e^{-\frac{3}{4}(Rk)^2}, \tag{24}$$

whereas the approximation (23) gives

$$N(\mathbf{k}) = 1 - e^{-(Rk)^2}. \tag{25}$$

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

3. Finite Volume Correction for Pixelized Methods

3.1. PIXELIZED METHODS

Pixelized data analysis methods start by reducing the galaxy survey problem to one similar to that occurring in cosmic microwave background (CMB) experiments: estimating a power spectrum given noisy fluctuation measurements in a number of discrete "pixels". After this, the remaining steps are quite analogous to the CMB case, and involve mere linear algebra (operations such as matrix inversion, diagonalization, *etc.*). Let us define the overdensity in N "pixels" $x_1, ..., x_N$ by

$$x_i \equiv \int \left[\frac{n(\mathbf{r})}{\bar{n}(\mathbf{r})} - 1\right] \psi_i(\mathbf{r}) d^3r \tag{26}$$

for some set of functions ψ_i. Although the specific choices of ψ_i are irrelevant for our present discussion, common choices are to either 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 one refers to the functions ψ_i as "modes" and to x_i as expansion coefficients). Let us group the pixels x_i into an N-dimensional vector $\mathbf{x}$. All proposed pixelized methods assume that the mean and the covariance matrix of this pixel vector are

$$\langle \mathbf{x} \rangle = \mathbf{0}, \tag{27}$$

$$\langle \mathbf{x}\mathbf{x}^t \rangle = \mathbf{C}, \tag{28}$$

where $\mathbf{C}$ depends in some known way on the power spectrum. Once the problem has been cast in this form, the power spectrum can be estimated using standard machinery, with either a brute force likelihood analysis (as in *e.g.* White & Bunn), a Karhunen-Loève eigenmode analysis (Karhunen 1947, Vogeley & Szalay 1996; Tegmark, Taylor & Heavens 1997) or a direct quadratic analysis (Hamilton 1997ab; Tegmark 1997).

3.2. DERIVATION OF THE INTEGRAL CONSTRAINT CORRECTION

For pixelized methods of power spectrum estimation, the procedure for dealing with the integral constraint is quite analogous to that for direct

Fourier methods. However, as we will now show, it is much simpler to implement. For counts in cells, for example, one simply removes the mean from all rows and columns of the covariance matrix $\mathbf{C}$ before proceeding with the analysis. Because of this simplicity, one can, at an almost negligible numerical cost, take a more ambitious approach and allow for more than one unknown parameter in the selection function. For instance, one can impose the constraints that the radial fluctuation average equals zero for a few hundred different angular bins, thereby eliminating the sensitivity to galactic extinction variations on this scale, as well as requiring that the angular fluctuation average vanish for a number of radial bins to be insensitive to errors in estimating the precise shape of $\bar{n}$.

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

$$\bar{n}(\mathbf{r}) = \sum_{j=1}^{M} \eta_j \bar{n}_j(\mathbf{r}), \tag{29}$$

where $\bar{n}_j$ are known functions and the "nuissance parameters" η_j, which we group into an M-dimensional vector η, are *a priori* unknown. Let $\bar{n}_0$ denote some *a priori* estimate of $\bar{n}$. Defining the "uncorrected" pixels as

$$x'_i \equiv \int \frac{n(\mathbf{r})}{\bar{n}_0(\mathbf{r})} \psi_i(\mathbf{r}) d^3r, \tag{30}$$

we find that

$$\langle \mathbf{x}' \rangle = \mathbf{Z}\eta, \tag{31}$$

where the $N \times M$-dimensional matrix $\mathbf{Z}$ is defined by

$$\mathbf{Z}_{ij} \equiv \int \frac{\bar{n}_j(\mathbf{r})}{\bar{n}_0(\mathbf{r})} \psi_i(\mathbf{r}) d^3r. \tag{32}$$

This means that in general, $\langle \mathbf{x}' \rangle \neq 0$, so the uncorrected data set does not satisfy equation (27). Instead, its statistical properties depend on the unknown nuissance parameters η. However, we can easily construct a new "corrected" data set whose mean is independent of η. Let us define

$$\mathbf{x} \equiv \mathbf{\Pi}\mathbf{x}', \tag{33}$$

where

$$\mathbf{\Pi} \equiv \mathbf{I} - \tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^t, \tag{34}$$

and $\tilde{\mathbf{Z}}$ is a matrix whose rows form an orthonormal basis ($\tilde{\mathbf{Z}}^t\tilde{\mathbf{Z}} = \mathbf{I}$) for the space spanned by the rows of $\mathbf{Z}$.[2] $\mathbf{\Pi}$ is a symmetric ($\mathbf{\Pi}^t = \mathbf{\Pi}$) projection

[2]Such a matrix $\tilde{\mathbf{Z}}$ is readily constructed by orthonormalizing the rows of $\mathbf{Z}$ with a Gram-Schmidt or Cholesky procedure (as in *e.g.* Tegmark & Bunn 1995).

matrix ($\mathbf{\Pi}^2 = \mathbf{\Pi}$) projecting onto the subspace orthogonal to the columns of $\mathbf{Z}$, *i.e.*, $\mathbf{\Pi Z} = \mathbf{0}$. Our corrected data set $\mathbf{x}$ satisfies equation (27), since $\langle \mathbf{x} \rangle = \mathbf{\Pi Z}\eta = \mathbf{0}$. Letting $\mathbf{C}'$ denote the covariance matrix of the uncorrected data set, the corrected data will have the covariance matrix

$$\mathbf{C} \equiv \langle \mathbf{x}\mathbf{x}^t \rangle = \mathbf{\Pi C' \Pi}. \tag{35}$$

Once $\mathbf{x}$ and $\mathbf{C}$ have been computed, the rest of the pixelized analysis proceeds just as if there had been no integral constraints. 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 is to replace all occurrences of $\mathbf{C}^{-1}$ (which is of course undefined) by the "pseudo-inverse" of $\mathbf{C}$, defined as

$$\mathbf{\Pi} \left[\mathbf{C} + \gamma \mathbf{Z}\mathbf{Z}^t \right]^{-1} \mathbf{\Pi} \tag{36}$$

for some constant $\gamma \neq 0$. The result is independent of γ, but 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 author wishes to thank Josh Frieman, Andrew Hamilton, Michael Strauss and Michael Vogeley for helpful comments on the manuscript.

References

Baugh, C. M. & Efstathiou, G. 1994, *MNRAS*, **267**, 323.
Bunn, E. F. & White, M. 1997, *ApJ*, **480**, 6.
Efstathiou, G. P. 1994, in "Les Houches Lectures 1993", eds. Shaefer & Silk (Elsevier, Netherlands).
Feldman, H. A., Kaiser, N. & Peacock, J. A. 1994, *ApJ*, **426**, 23 ("FKP").
Fisher, K. B. *et al.* 1993, *ApJ*, **402**, 42.
Hamilton, A. J. S. 1997a, preprint astro-ph/9701008, MNRAS, in press.
Hamilton, A. J. S. 1997b, preprint astro-ph/9701009, MNRAS, in press.
Karhunen, K., *Über lineare Methoden in der Wahrscheinlichkeitsrechnung* (Kirjapaino oy. sana, Helsinki, 1947).
Park, C., Vogeley, M. S, Geller, M. J. & Huchra, J. P. 1994, *ApJ*, **431**, 569.
Peacock, J. A. & Nicholson, D. 1991, *MNRAS*, **253**, 307.
Strauss, M. A. & Willick, J. A. 1995, *Phys. Rep.*, **261**, 271.
Strauss, M. A. 1996, astro-ph/9610033, to be published in "Structure Formation in the Universe", ed. A. Dekel & J. P. Ostriker.
Tegmark, M. 1995, *ApJ*, **455**, 429 ("T95").
Tegmark, M. & Bunn, E. F. 1995, *ApJ*, **455**, 1.
Tegmark, M., Taylor, A. & Heavens, A. F. 1997, *ApJ*, **480**, 22.
Tegmark, M. 1997, *Phys. Rev. D*, **55**, 5895.
Vogeley, M. S. 1995, in "Wide-Field Spectroscopy and the Distant Universe", eds. Maddox & Aragón-Salamanca (World Scientific, Singapore).
Vogeley, M. S. & Szalay, A. S. 1996, *ApJ*, **465**, 43.