Unbiased contaminant removal for 3D galaxy power spectrum measurements

B. Kalus\textsuperscript{1,*}, W. J. Percival\textsuperscript{1}, D. J. Bacon\textsuperscript{1} and L. Samushia\textsuperscript{2,1,3}

\textsuperscript{1}Institute of Cosmology & Gravitation, Dennis Sciama Building, University of Portsmouth, Portsmouth, PO1 3FX, UK
\textsuperscript{2}Department of Physics, Kansas State University, L16, Cardwell Hall, Manhattan, KS, 66506, USA
\textsuperscript{3}National Abastumani Astrophysical Observatory, Ilia State University, 2A Kazbegi Ave., GE-1060 Tbilisi, Georgia

Abstract
We assess and develop techniques to remove contaminants when calculating the 3D galaxy power spectrum. We separate the process into three separate stages: (i) removing the contaminated cosmological power spectrum, (ii) estimating the uncontaminated cosmological power spectrum, (iii) debiasing the resulting estimates. For (i), we show that removing the best-fit contaminant (mode subtraction) and setting the contaminated components of the covariance to be infinite (mode deprojection) are mathematically equivalent. For (ii), performing a Quadratic Maximum Likelihood (QML) estimate after mode deprojection gives an optimal unbiased solution, although it requires the manipulation of large $N_{\text{mode}}$ matrices ($N_{\text{mode}}$ being the total number of modes), which is unfeasible for recent 3D galaxy surveys. Measuring a binned average of the modes for (ii) as proposed by Feldman, Kaiser & Peacock (1994, FKP) is faster and simpler, but is sub-optimal and gives rise to a biased solution. We present a method to debias the resulting FKP measurements that does not require any large matrix calculations. We argue that the sub-optimality of the FKP estimator compared with the QML estimator, caused by contaminants is less severe than that commonly ignored due to the survey window.

Key words: methods: statistical -- cosmology: large-scale structure of Universe.

1 INTRODUCTION

Galaxy surveys provide a rich store of information about the nature of the Universe, allowing us to constrain cosmological models with baryon acoustic oscillations (BAO), gravitational models with redshift space distortions (RSD) and inflationary models with primordial non-Gaussianity. A basic statistic containing large-scale structure information is the galaxy power spectrum $P(k)$, which is the 2-point function of the Fourier transformed density field. Future large-scale structure surveys, such as the Dark Energy Spectroscopic Instrument survey (Schlegel et al. 2011; Levi et al. 2013, DESI), Euclid (Laureijs et al. 2011)\textsuperscript{1} and the Square Kilometre Array (SKA)\textsuperscript{2}, will probe larger volumes, therefore allowing us to measure more Fourier modes of the galaxy density field.

The observed galaxy field can be contaminated with fluctuations of non-cosmological origin, such as variations due to the galactic extinction and the stellar density. Often the contaminants are not known exactly (e.g. we may know the shape of the spurious mode but may not know its exact amplitude) which makes their exact removal impossible. These modes have the potential to strongly bias cosmological constraints derived from the clustering measurements, so we need to correct or suppress these misleading modes in a responsible way.

We now introduce the basic mathematical problem that we wish to solve and introduce the main methods of removing contaminants discussed in literature. We assume that we have measured the galaxy density field as real numbers in configuration space, which we (fast) Fourier transform to obtain a Hermitian density field $F(k)$. Furthermore, we assume that the contamination can be described by another Hermitian field $f(k)$, such that the true density field is given by

$$D(k) = F(k) - \epsilon_{\text{true}} f(k),$$

with $\epsilon_{\text{true}}$ unknown. In cases with multiple contaminants (which we label with capital Latin indices), we extend Eq. (1) to

$$D(k) = F(k) - \sum A \epsilon_A^{\text{true}} f_A(k).$$

Furthermore, we assume that $F(k)$ and $f(k)$ are uncorrelated, which is a valid assumption for most sources of systematics since they originate from our Galaxy or due to telescope effects. Large scale surveys will reduce the current sample variance limitation on the power spectrum on scales where the systematic errors have a significant impact. As a consequence, having control of these systematics is a key requirement to provide accurate cosmological measurements.

\footnotesize
\textsuperscript{*} E-mail: benedict.kalus@port.ac.uk
\textsuperscript{1} www.euclid-ec.org
\textsuperscript{2} www.skatelescope.org
In order to investigate techniques for estimating the power spectrum in the presence of contaminants, we separate the process into three separate stages: (i) removing the contaminant signal, (ii) estimating the uncontaminated cosmological power spectrum, (iii) debiasing the resulting estimates. Two techniques are in common usage for removing the contaminant signal (i): The first is mode subtraction (cf. Sec. 4 and Sec. 5), where contaminants are removed by fitting the amplitude of the contaminant field $f(k)$ to the data and simply subtracting off from $F(k)$. The second is mode deprojection (Rybicki & Press 1992), which is based on assigning infinitely large covariances to contaminated modes, thus removing them from any analysis. In our nomenclature, a mode is a linear combination of Fourier modes rather than a single $k$-mode. This is reflected in the naming of mode subtraction and mode deprojection. This choice of names shall distinguish the mode subtraction technique from a third technique for removing the contaminant signal, called template subtraction, where the observed power spectra are corrected using best-fit amplitudes derived via cross-correlations between the data and the templates. Elsner, Leistedt & Peiris (2016) have shown that this method provides a biased estimate of the power and we will not consider it further in this article. For (ii), the power spectrum $P(k)$ is commonly estimated by the FKP estimator (Feldman et al. 1994), which is an approximation to the Quadratic Maximum Likelihood (QML) estimator (Tegmark et al. 1998). As well as being optimal, the QML estimator has the advantage of producing unbiased power spectrum estimates. However, when applying this methodology to data with $N$ modes, one has to calculate, for each bin, a $N_{\text{mode}} \times N_{\text{mode}}$ matrix, and then, after binning the data into $N_{\text{bin}}$ bins, an overall $N_{\text{bin}} \times N_{\text{bin}}$ normalisation matrix, which makes the application of this methodology unfeasible for future surveys with increased number of modes $N_{\text{mode}}$. In this work, we suggest a modified FKP-style mode subtraction approach. We show that this technique can be made unbiased and, on a mode-by-mode basis, is mathematically identical to mode deprojection. The FKP estimator with debiased mode subtraction is not optimal in that it discards more information than the full QML estimator, but we expect that, in realistic cases, this loss of information will be small.

The outline of this paper is as follows: We provide an introduction to power spectrum estimation in Sec. 2, introducing the QML and FKP estimators. We introduce the systematic removal techniques, mode deprojection and mode subtraction, in sections 3 and 4, respectively, and we show that before normalisation their resulting power spectra are the same. These are extended to multiple contaminants in Appendices A & B, respectively. We introduce a new normalisation factor in Sec. 5 for a single contaminant and compare it to the normalisation of the quadratic maximum likelihood (QML) estimator of Tegmark et al. (1998). This derivation is extended to allow for a non-diagonal covariance in Appendix C1. We show that we can apply our methodology also to multiple contaminants in Sec. 6 and test the different methods on simulations in Sec. 7. We conclude in Sec. 8.

## 2 POWER SPECTRUM ESTIMATORS

In this section, we review two basic power spectrum estimators: the quadratic maximum likelihood (QML) estimator (Tegmark et al. 1998) and the simplified FKP estimator (Feldman et al. 1994), to which QML reduces in the limit of uncorrelated modes with equal noise per mode in each bin. Even without considering any contaminants, the FKP estimator is easier to implement and is used in most recent analyses of large-scale structure, while the QML estimator is optimal but difficult to implement especially on smaller scales.

The quadratic maximum likelihood (QML) estimator (Tegmark et al. 1998) is given by

$$\hat{P}(k) = \sum_j N_{ij}^{-1} p_j,$$

where the power is a convolution of the inverse of a normalisation matrix $N_{ij}$ and a weighted two-point function

$$p_j = \sum_{\alpha,\beta} F^* (k_{\alpha}) E_{\alpha\beta} (k_j) F(k_{\beta}).$$

The weight is given by the estimator matrix

$$E(k_j) = - \frac{\partial C^{-1}}{\partial P(k_i)} C^{-1} \frac{\partial C}{\partial P(k_j)},$$

which describes how the inverse of the density field covariance matrix $C$ changes with respect to the prior of the power spectrum of the respective bin. If the QML normalisation is proportional to the Fisher information, i.e.

$$N_{ij} = \text{tr} \left\{ C^{-1} \frac{\partial C}{\partial P(k_i)} C^{-1} \frac{\partial C}{\partial P(k_j)} \right\},$$

the QML estimator is the optimal maximum likelihood estimator of the variance of a field that obeys a multivariate Gaussian distribution (Tegmark et al. 1998). Assuming a Gaussian density field $D(k)$, the QML estimator therefore provides an estimate of the power spectrum with minimal errors.

Under the assumption that all modes are independent, the covariance of the density field is given by the power spectrum (and the Kronecker delta $\delta_{\mu\nu}$):

$$C_{\mu\nu} = \delta_{\mu\nu} P(k_{\mu}).$$

We assume for the derivative of $C$ with respect to $P(k_i)$ that it is unity if the modes $k_\alpha$ and $k_\beta$ are equal and contained in the bin $k_i$, and zero otherwise, which we write using the Heaviside function $\Theta$ as:

$$\frac{\partial C_{\alpha\beta}}{\partial P(k_i)} = \delta_{\alpha\beta} \Theta(k_{\alpha} \in k_i) \equiv \delta_{\alpha\beta} \Theta_{\alpha i}.$$ 

Given Eq. (7) and (8), we find

$$E_{\alpha\beta} (k_j) = \frac{\delta_{\alpha\beta}}{P^2(k_{\alpha})} \Theta_{\alpha j}$$

and

$$N_{ij} = \frac{N_{ij}}{P^2(k_{\alpha})} \delta_{ij},$$

where $N_{ij}$ is the total number of modes in a given bin $k_i$. Hence the QML estimator of Eq. (3) reduces to the FKP estimator (Feldman, Kaiser & Peacock 1994) under the assumption that the covariance is constant within the $k$-bin, where several modes (labelled with Greek indices) are combined into bins (denoted by $k_i$ and distinguished with lower case Latin indices) and the absolute values squared of the density field of each bin are summed:

$$\hat{P}(k) = \frac{1}{N_{k_i}} \sum_{k_{\alpha} \in k_i} |F(k_{\alpha})|^2.$$ 

The difference here is that the QML estimator uses a prior of the power spectrum $P(k_{\alpha})$ to weight contributions from each mode optimally, which means that the covariance of the power spectrum is minimal. The FKP estimator is commonly applied even when the assumptions of Eq. (7) to (10) are not valid.
3 REMOVING CONTAMINANTS: MODE DEPROJECTION

We now describe how mode deprojection can be applied to estimate the 3D galaxy power spectrum. The method was first suggested in Rybicki & Press (1992) in the context of noisy, irregularly sampled data. Applications and extensions to angular power spectra can be found for WMAP data in Slosar et al. (2004), for SDSS-III data in Ho et al. (2012), for photometric quasars of the XDQS0z catalogue in Leistedt & Peiris (2014) and Leistedt et al. (2014) and for 2D galaxy clustering in general in Elsner et al. (2016). We use the notation of Elsner et al. (2016) for consistency.

Suppose we estimate the power spectrum using QML and that there is only a single contaminant. Then one can suppress contaminated modes in the covariance matrix updating the covariance matrix as (Elsner et al. 2016)

$$C_{\alpha\beta} \rightarrow \tilde{C}_{\alpha\beta} = C_{\alpha\beta} + \lim_{\sigma \rightarrow \infty} \sigma f(k_\alpha) f^*(k_\beta),$$

(12)
i.e. letting the covariances of contaminated modes tend to infinity. Making use of the Sherman-Morrison matrix inversion lemma (Sherman & Morrison 1950), one can see that if $f(k) \neq 0 \forall k$ the inverse updated covariance matrix converges to

$$\tilde{C}^{-1}_{\alpha\beta} = C^{-1}_{\alpha\beta} \sum_{\mu \nu} C_{\alpha\mu} f(k_\mu) f^*(k_\nu) C_{\nu\beta} \overline{f(k_\nu)} f(k_\beta).$$

(13)

Now supposing that the modes are independent, i.e. Eq. (7) holds, we can insert it into Eq. (13) so that

$$\tilde{C}^{-1}_{\alpha\beta} = \frac{\delta_{\alpha\beta} P(k_\alpha)}{P(k_\alpha)} - \frac{1}{R_P P(k_\alpha) P(k_\beta)} \left[ f(k_\mu) f^*(k_\beta) \sum_{\mu \nu} f(k_\mu) f^*(k_\nu) C^{-1}_{\alpha\beta} \right],$$

(14)

where we have defined

$$R_P \equiv \sum_{\mu} \frac{|f(k_\mu)|^2}{P(k_\mu)},$$

(15)

for simplicity. Taking the derivative of Eq. (14) with respect to $P(k_i)$, we obtain the updated estimator matrix

$$\tilde{E}_{\alpha\beta}(k_i) = \frac{\delta_{\alpha\beta}}{P^2_{\alpha\beta}} \Theta_{\alpha\beta} - \frac{1}{R_P P_{\alpha\beta} P_{\beta\beta}} \left[ \Theta_{\alpha\beta} P_{\alpha\beta} + \Theta_{\beta\alpha} - t_{\alpha\beta} - t_{\beta\alpha} \right],$$

(16)

where

$$t_i \equiv \sum_{k_\alpha \in k_i} \frac{|f(k_\alpha)|^2}{P^2(k_\alpha)}.$$

(17)

After inserting Eq. (16) into Eq. (4), we obtain for the two point function

$$p_{\alpha\beta} = \sum_{k_\alpha \in k_i} \left\{ \frac{|F(k_\alpha)|^2}{P^2(k_\alpha)} - \frac{2}{R_P} \text{Re} \left[ S_P \frac{F^*(k_\alpha) f(k_\alpha)}{P^2(k_\alpha)} \right] \right\} + \frac{|S_P|^2}{R_P} \left| f(k_\alpha) \right|^2$$

(18)

$$= \sum_{k_\alpha \in k_i} \frac{F^*(k_\alpha) f(k_\alpha)}{P^2(k_\alpha)},$$

where we have defined

$$S_P \equiv \sum_{k_\alpha} \frac{F^*(k_\alpha) f(k_\alpha)}{P_\alpha}.$$

(19)

$S_P$ is real, because $F(k)$ and $f(k)$ are Hermitian fields with real Fourier transforms.

Eq. (18) is in a considerably simpler form than Eq. (4) and does not require calculating many matrix elements of the estimator matrix $E$. We show in the next section that we can consider this equation as a best-fit of the contaminants in the data.

We can normalise the updated mode deprojected QML estimator by replacing $C$ by $\tilde{C}$ in Eq. (6). As the term that suppresses contaminated modes from the covariance matrix in Eq. (12) does not depend on the power $P(k)$, we have $\frac{\partial \tilde{E}_{\alpha\beta}}{\partial P(k_\beta)} = \frac{\partial E_{\alpha\beta}}{\partial P(k_\beta)}$ and hence the normalisation is

$$\tilde{N}_{ij} = \sum_{\alpha \mu \nu \rho} \tilde{C}_{ij \alpha \beta} \Theta_{\mu \rho} \tilde{C}^{-1}_{\nu \beta \alpha} \delta_{\mu \alpha} \Theta_{\nu \beta}$$

$$= \sum_{\alpha \mu} \tilde{C}^{-1}_{i \alpha \beta} \Theta_{\mu \rho} \tilde{C}^{-1}_{j \nu \beta} \delta_{\mu \alpha} \Theta_{\nu \beta}$$

$$= \sum_{\alpha \mu} \Theta_{\mu \rho} \tilde{C}^{-1}_{i \alpha \beta} \tilde{C}^{-1}_{j \nu \beta} \delta_{\mu \alpha} \Theta_{\nu \beta}$$

$$+ \left[ \Theta_{\mu \rho} \tilde{C}^{-1}_{i \alpha \beta} \tilde{C}^{-1}_{j \nu \beta} \delta_{\mu \alpha} \Theta_{\nu \beta} \right]$$

(20)

where we have used the Hermitian property of $\tilde{C}^{-1}$ in the third equality. In the first term in the square brackets, $k_\alpha$ has to be in both $k_i$ and $k_j$, hence we can replace one $\Theta$ with $\delta_{ij}$, such that Eq. (20) can be written as a diagonal matrix with diagonal elements $\tilde{n}$ and the outer product of a vector with itself:

$$\tilde{N}_{ij} = \sum_{k_\alpha \in k_i} \frac{\delta_{ij}}{P^2(k_\alpha)} \left( 1 - \frac{2 |f(k_\alpha)|^2}{R_P P(k_\alpha)} \right) + \frac{t_i t_j}{R_P^2},$$

(21)

This means that we can apply the Sherman-Morrison matrix inversion lemma (Sherman & Morrison 1950):

$$\tilde{N}_{ij}^{-1} = \frac{\delta_{ij}}{\tilde{n}_i} - \frac{1}{R_P^2} + \sum_{k_\alpha \in k_i} \frac{t_i t_j}{R_P^2},$$

(22)

As $\tilde{N}^{-1}$ is not diagonal, it does not reduce to a simple FKP style estimator, i.e. if we have $N_{\text{bin}}$ bins, we have to calculate for each bin the $N_{\text{mode}} \times N_{\text{mode}}$ estimator matrix $E$ and we have to invert the $N_{\text{bin}} \times N_{\text{bin}}$ normalisation matrix. This is not feasible for 3D clustering, because of the large number of modes to be considered, especially if we want to choose narrow bins. Including several contaminants makes it even more costly.

One way around this is a new framework introduced by Leistedt & Peiris (2014) which they call extended mode projection and that selectively removes modes based on cross correlations with the data. However, this procedure reintroduces a small bias (Elsner et al. 2016).

Another possibility is using the methodology of the SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS)-collaboration, which is similar to that described in the next section, but applied at the power spectrum level. However, this method is also biased (Elsner et al. 2016). Although Ross et al. (2016) show that, for the Completed SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS DR12), the bias is much smaller than the statistical uncertainty, it was shown in the appendix of Ross et al. (2012) that the bias is significant when one attempts to correct for many systematics. Furthermore, we expect smaller statistical uncertainties with future surveys, so in the next two sections we consider a computationally cheaper way of removing this small bias.
4 REMOVING CONTAMINANTS: MODE SUBTRACTION

Here we will consider mode subtraction and its link to mode deprojection. In order to remove contaminants we start by treating the true, but unknown, amplitude of the contamination $\varepsilon_{\text{true}}$ in Eq. (1) as a free parameter $\varepsilon$, so that an estimate of the true density field $D(k)$ reads

$$\hat{D}(k) = F(k) - \varepsilon f(k). \quad (23)$$

Note that this is different to the template subtraction method introduced by Ho et al. (2012), which is used by the BOSS collaboration and works entirely at the level of power spectra, whereas Eq. (23) works at the map level. We can write a simplified model introduced by Ho et al. (2012), which is used by the BOSS collaboration.

$$-2 \ln L = \ln \left( \prod_k P(k) \right) + \sum_k \frac{|F(k) - \varepsilon f(k)|^2}{P(k)}. \quad (24)$$

We can therefore find $\varepsilon$ by minimising Eq. (24), which is equivalent to simultaneously fitting $\varepsilon$ and the model parameters entering the model power spectrum. The derivative of $\ln L$ with respect to $\varepsilon$ reads

$$\frac{\partial \ln L}{\partial \varepsilon} = \sum_k \frac{\Re[F^*(k)F(k)] - \varepsilon |F(k)|^2}{P(k)}. \quad (25)$$

This expression is equal to zero and the likelihood maximised if

$$\varepsilon^{(\text{BF})} = \frac{S_P}{R_P}. \quad (26)$$

The uncontaminated estimate of the density field is hence given by

$$\hat{D}(k) = F(k) - \frac{S_P}{R_P} f(k), \quad (27)$$

and we can estimate the power as

$$\hat{P}(k_i) = \frac{1}{N_{k_i}} \sum_{k_o} \left| F(k_o) - \frac{S_P}{R_P} f(k_o) \right|^2. \quad (28)$$

This is similar to the mode deprojection result of Eq. (18) with a bias, missing the inverse noise matrix convolution of Eq. (3). The bias of this estimate comes about because $S_P$ is correlated with the true density field $D(k)$. This correlation is similar to that created by the internal linear combination (ILC) method (e.g. Bennett et al. 2003) for the analysis of cosmic microwave background (CMB) data. Based on this knowledge, we build an unbiased FKP-style estimator in the next section.

5 AN UNBIASED FKP-STYLE ESTIMATOR

We present in this section a simple, although sub-optimal, way to remove the bias on the power spectrum estimate resulting from imperfectly removing systematics using either Eq. (18) or (28). A straightforward way to remove the bias consists of calculating the expectation value of the power from each mode analytically, assuming Eq. (1), and divide out the bias. We start with calculating some useful expectations which we need for the final result, summarised in Table 1. With these equations at hand, we can calculate the expectation of Eq. (18) and (28), i.e. the two-point function of Eq. (1):

$$\langle |F(k_o)| - \frac{S_P}{R_P} f(k_o) \rangle^2$$

$$= \langle |F(k_o)|^2 \rangle - \frac{2}{R_P} \langle S_P F(k_o) \rangle \langle f(k_o) \rangle + \langle \varepsilon_{\text{true}}^2 \rangle \langle f(k_o) \rangle^2$$

$$= P(k_o) - \frac{|f(k_o)|^2}{R_P}, \quad (29)$$

hence, we can build an unbiased estimator of the power by dividing each mode in Eq. (18) and (28) by

$$1 - \frac{1}{R_P} \frac{|f(k_o)|^2}{P(k_o)}. \quad (30)$$

If we want to debias the two-point function using this factor, we have to assume a prior power spectrum. Note that the QML approach also requires the prior knowledge of the power spectrum. We will see in Sec. 7 that the impact of adopting a slightly wrong prior is indeed small. Our final estimator of the power spectrum is then

$$\tilde{P}(k_i) = \frac{1}{N_{k_i}} \sum_{k_o} \left| F(k_o) - \frac{S_P}{R_P} f(k_o) \right|^2. \quad (31)$$

Eq. (31) is one of the key results of this article; this is an extension of the FKP estimator that removes potential contaminants from the data in an unbiased way, without the need for large matrices. Moreover, as it is in the same form as the well established FKP estimator, this can easily be folded into estimators for redshift-space clustering such as those by Bianchi et al. (2015) and Scoccimarro (2015).

The same debiasing factor can also be derived from the QML Fisher information matrix $N$, which in the QML approach performs both the debiasing and optimisation effects. Without binning, Eq. (21) simplifies to

$$\tilde{N}_{\alpha\beta} = \frac{\delta_{\alpha\beta}}{P^2(k_o)} \left( 1 - \frac{2 |f(k_o)|^2}{P(k_o)} \right) + \frac{1}{R_P^2} \frac{|f(k_o)|^2}{P^2(k_o)} \frac{|f(k_o)|^2}{P(k_o)} \frac{|f(k_o)|^2}{P(k_o)} \frac{|f(k_o)|^2}{P(k_o)}. \quad (32)$$

The difference between the two approaches is that QML provides an unbiased optimal power estimate, whereas Eq. (31) has been constructed such that it is only unbiased, i.e. the powers in the denominators of Eq. (32) act as optimal weights to each mode. If we allow for some information loss within bins, by assuming the expected power is constant within each bin, we can replace $P^2(k_o)$ by $P(k_o)P(k_o)$, such that

$$\tilde{N}_{\alpha\beta} = \frac{\delta_{\alpha\beta}}{P^2(k_o)} \left( 1 - \frac{2 |f(k_o)|^2}{R_P P(k_o)} \right) + \frac{1}{R_P^2} \frac{|f(k_o)|^2}{P^3(k_o)} \frac{|f(k_o)|^2}{P(k_o)} \frac{|f(k_o)|^2}{P(k_o)} \frac{|f(k_o)|^2}{P(k_o)}. \quad (33)$$

This normalisation is proportional to the Fisher information matrix (Tegmark et al. 1998), from which we marginalise out contributions.
from other modes by summing over all modes \( k_\beta \):

\[
\sum_\beta \tilde{N}_{\alpha\beta} = \frac{1}{P^2(k_\alpha)} \left( 1 - \frac{2 f(k_\alpha)}{R_P P(k_\alpha)} \right)^2 + \frac{1}{R_P^2} \frac{|f(k_\alpha)|^2}{P^2(k_\alpha)}
\]

This is exactly Eq. (30) with a factor of \( \frac{1}{P^2(k_\alpha)} \) that cancels out the difference between Eq. (18) and Eq. (28). We have therefore shown that Eq. (31) is a non-optimal, but unbiased, approximation to using the QML normalisation with mode deprojection. In the limit of narrow bins, when the power spectrum does not change significantly within the bin, Eq. (31) is mathematically identical to the QML result. We shall study the impact of this sub-optimality in examples later in Sec. 7. In fact, we will argue later that this is actually a weaker effect than many common approximations applied when using the FKP estimator, such as ignoring large-scale window effects in the QML approach, when averaging large scale modes.

Note that, in the absence of systematics, we have assumed a diagonal covariance matrix in the derivation of both the mode subtraction and the debiasing step. In practice the covariance matrix has off-diagonal terms due to the effect of the survey window. However, this is usually not included when calculating the data power spectrum but, instead, it is included as a convolution in the model power spectrum. We show in Appendix C1 that Eq. (31) still holds in the general case of having a non-diagonal covariance matrix, as long as \( R_P \) is generalised as in Eq. (8). This generalised \( R_P \) requires the inversion of the full \( N_{\alpha\beta} \) covariance matrix. However, we show in Appendix C2 that the effect of assuming a diagonal covariance matrix is either small, or can be corrected for using the covariance matrix, without inversion.

6 REMOVING MULTIPLE CONTAMINANTS

We have shown the equivalence between mode deprojection and debiased mode subtraction for one contaminant. A realistic survey has several sources of potential contaminants, so we show here this equivalence holds for an arbitrary number of templates. For mode deprojection, we have to update the covariance matrix with a sum over all templates, and thus we have to replace Eq. (12) with

\[
\tilde{C}_{\alpha\beta} = C_{\alpha\beta} + \lim_{\sigma \to \infty} \sum_{A=1}^{N_{\alpha\beta}} \tilde{f}_A(k_\alpha) f_A(k_\beta).
\]

Starting from Eq. (35), we derive in Appendix A the unbinned mode deprojection power spectrum

\[
\tilde{P}(k_\alpha) = \left| F(k_\alpha) - \sum_{A} S_A R_{A\beta} f_B(k_\beta) \right|^2,
\]

where \( R_{A\beta} \equiv \sum_\mu f_A(k_\alpha) f_B(k_\mu) \) and \( S_A \equiv \sum_\mu f_A(k_\mu) f_A(k_\mu) \) are matrix and vector equivalents of \( R_P \) and \( S_P \), respectively, in contaminant space. To apply multiple mode subtraction, we extend the likelihood given in Eq. (24) to

\[
-2 \ln L = \sum_\alpha \left[ \frac{F(k_\alpha)}{P(k_\alpha)} - \sum_A \tilde{f}_A f_A(k_\alpha) \right]^2.
\]

Writing \( \varepsilon \) as a vector, the joint maximum likelihood solution fitting all contaminants is given by (cf. Appendix B)

\[
\varepsilon^{(\text{ML})} = R^{-1} S.
\]

Note that this would require fitting the amplitude of all contaminants simultaneously. The absolute value squared of the best fitting signal is hence equal to Eq. (36). Hence, we also do not need large \( N_{\alpha\beta} \times N_{\alpha\beta} \) matrices when we have to remove several potential contaminants.

We can calculate the debiasing factor

\[
\sum_j \tilde{N}_{ij} P^2(k_i) = 1 - \sum_{AB} \tilde{f}_A(k_i) R_{A\beta}^{-1} f_B(k_i) P(k_i)
\]

analogously to Sec. 4 from the mode deprojection normalisation matrix without binning.

7 TESTING CONTAMINANT REMOVAL

In this section we show how simple contaminants can be removed in power spectrum measurements from simulated density fields, using the hitherto described methodologies.

7.1 Gaussian Spike Contaminant

As a first test, we generate 3-dimensional Gaussian random fields according to an input power spectrum that we calculate using CAMB (Lewis et al. 2000). Each of these fields consists of a \( 16 \times 16 \times 16 \) grid, in a box of length \( 3136 h^{-1} \) Mpc. An example of such a field is shown in the left panel of Fig. 1. We contaminate these Gaussian random fields by adding a real Gaussian spike in \( k \)-space with width \( \sigma = 10^{-3} h \) Mpc, centred around \( k = 0.01 h \) Mpc\(^{-1}\), such that its maximum lies within a bin with sufficiently good statistics. The Fourier transform of this contaminant field is again a Gaussian spike in the centre of the box with some long wavelength fluctuations around it. The amplitude of the real part over-density in \( k \)-space is 100, thus having the same order of magnitude as the “true” density field. An example of this setup can be seen in the central panel of Fig. 1. We calculate four different power spectra:

(i) We do not account for the contaminants and just average the absolute values squared of the density field in each bin (cf. Eq. (11)).

(ii) We perform a naïve mode subtraction, i.e. we subtract off the template, but do not debias the two-point function (cf. Eq. (28)).

(iii) We debias the previous power spectrum by applying Eq. (30).

(iv) We use the full QML estimator with mode deprojection.

In the cases (ii) to (iv), we have to assume a prior power spectrum, which we take as equal to the input power. We shall test the effect of this assumption with the next example. As each bin contains modes of different \( k \)-values, we have to clarify what we mean by the prior power spectrum \( P(k_i) \) for a specific bin. We find that the power spectrum measurements are closest to the input values, when we assume that the input power spectrum \( P(k_i) \) is given by the average of the prior power spectrum values for each mode in the respective bin, i.e.

\[
P(k_i) \equiv \frac{1}{N_{k_i}} \sum_{k_i < k_i} P(k_i).
\]

In Fig. 2, we can clearly see an increase of power in the bins around \( k = 0.01 \) in case (i). Subtracting off the template in the naïve way (method (ii)) is biased in the bins affected by the spike. However, this bias is only a 1 part in a thousand effect. Methods (iii)
and (iv) both reproduce the input power spectrum well, removing the bias. A significant difference between their error bars cannot be observed. It is therefore sufficient in this case to use the FKP-style estimator we introduced in Sec. 5.

7.2 Single Contaminated Mode

As a second example we use Eq. (30) to construct a contaminant that would lead to a strong bias in the recovered $P(k)$ without the debiasing step. Eq. (30) only contains positive quantities and is normalised such that the bias is a value between 0 and 1. 1 corresponds to an unbiased estimate, hence 0 is the maximal bias. This extreme case would be fulfilled if $f$ is large for one mode and 0 otherwise. Therefore, we construct a contaminant that is a large number at the modes corresponding to $k = \pm (0.003, 0.003, 0.003)h$ Mpc$^{-1}$. An example of this setup can be found in Fig. 3. The top panel again shows an uncontaminated Gaussian random field, the central

![Figure 1](image1.png)

**Figure 1.** A slice through a realisation of a Gaussian random field contaminated with a Gaussian spike used in Sec. 7.1. The top panel shows the “clean” Gaussian random field (corresponding to $D(k)$ through Fourier transform) in configuration space. In the central panel, we have plotted the contaminated field (Fourier pair of $F(k)$) with an obvious Gaussian overdensity in the centre. The bottom panel shows the residual, i.e. the difference of the field after mode subtraction (i.e. the Fourier transform of $F(k) - \varepsilon^\text{(BF)} f(k)$, cf. Eq. (28)) and the input field. The best-fitting $\varepsilon^\text{(BF)}$ for this particular realisation amounts to 1.078. Although differences between the top and bottom panels are hard to spot by eye, in Fourier space the differences correspond to the bias in the mode subtraction estimator.

![Figure 2](image2.png)

**Figure 2.** Means and standard deviations of the power spectra of 70,000 realisations of Gaussian random fields contaminated with a real Gaussian spike. The top panel shows the input power spectrum as a solid blue line, as well as the power spectra obtained with methods (i)-(iv) as described in Sec. 7.1. In the lower panel, we plot fractional errors for methods (ii)-(iv).

![Figure 3](image3.png)

**Figure 3.** This plot is similar to Fig. 1, but shows a slice through a field with a single contaminated mode as described in Sec. 7.2. The best-fitting $\varepsilon^\text{(BF)}$ for this particular realisation amounts to 1.005. All panels appear very similar; this is quantified in Fig. 4.
with large numbers of observed modes, but that we can apply QML estimation is not practical for modern surveys focusing on mode deprojection. We have considered methods to remove contaminants when measuring the power spectrum estimation - they are often considered together - arguing that this split makes sense given the mathematical equivalence of mode deprojection and mode subtraction. We argue that the QML estimation is not practical for modern surveys with large numbers of observed modes, but that we can apply mode deprojection to the FKP-estimator, using the mathematical equivalence of mode deprojection and mode subtraction, thus avoiding having to create large estimator and covariance matrices for all modes. The resulting estimate is biased, but can easily be made unbiased with a simple correction, again that can be implemented without the inversion of large matrices. This correction is easily extended to the case of multiple contaminants and is not affected if the modes are correlated even without the effects of contaminants. The final result of our short paper is the suggestion that 3D galaxy power spectrum should be estimated using Eq. (31).

\[
\hat{P}(k_i) = \frac{1}{N_{\text{bins}}} \sum_{k_\alpha} \left| \frac{F(k_\alpha) - \frac{S_P}{P(k)} f(k_\alpha)}{1 - \frac{1}{N_{\text{bins}}}} \right|^2.
\]  

While theoretically it is sub-optimal, in practice the degradation of signal is expected to be less than ignoring window effects in the optimisation of mode averaging when using the standard FKP estimator.

8 CONCLUSIONS

We have considered methods to remove contaminants when measuring the 3D galaxy power spectrum from a given density field, focussing on mode deprojection and mode subtraction. In order to understand how these are related, we have decomposed the problem into separate steps. In particular we have separated mode deprojection from power spectrum estimation - they are often considered together - arguing that this split makes sense given the mathematical equivalence of mode deprojection and mode subtraction. We argue that the QML estimation is not practical for modern surveys with large numbers of observed modes, but that we can apply mode deprojection to the FKP-estimator, using the mathematical equivalence of mode deprojection and mode subtraction, thus avoiding having to create large estimator and covariance matrices for all modes. The resulting estimate is biased, but can easily be made unbiased with a simple correction, again that can be implemented without the inversion of large matrices. This correction is easily extended to the case of multiple contaminants and is not affected if the modes are correlated even without the effects of contaminants. The final result of our short paper is the suggestion that 3D galaxy power spectrum should be estimated using Eq. (31).

\[
\hat{P}(k_i) = \frac{1}{N_{\text{bins}}} \sum_{k_\alpha} \left| \frac{F(k_\alpha) - \frac{S_P}{P(k)} f(k_\alpha)}{1 - \frac{1}{N_{\text{bins}}}} \right|^2.
\]  

While theoretically it is sub-optimal, in practice the degradation of signal is expected to be less than ignoring window effects in the optimisation of mode averaging when using the standard FKP estimator.

ACKNOWLEDGMENTS

The authors would like to thank Franz Elsner, Héctor Gil-Márín, Ashley Ross and the unknown referee for valuable comments.

We used matplotlib (Hunter 2007) to generate plots. The CAMB package (Lewis et al. 2000) has been used to generate model and prior power spectra. We made use of the facilities and staff of the UK Sciafa High Performance Computing cluster supported by the ICG, SEPNet and the University of Portsmouth.

BK thanks the Faculty of Technology of the University of Portsmouth for support during his PhD studies. WJP and DB acknowledge support from UK STFC through the consolidated grant ST/K0090X/1. WJP also acknowledges support from the European Research Council through the Darksurvey grant and the UK Space Agency through grant ST/N00180X/1. LS is grateful for support from SNSF grant SCOPES IZ73Z0-152581, GNSF grant FR/339/6-350/14, and DOE grant DEFG 03-99ER41093.

REFERENCES

Bennett C., et al., 2003, Astrophys. J. Suppl., 148, 97
Bianchi D., Gil-Marn H., Ruggeri R., Percival W. J., 2015, Mon. Not. Roy. Astron. Soc., 453, L11
Elsner F., Leistedt B., Peiris H. V., 2016, Mon. Not. Roy. Astron. Soc., 456, 2095
Feldman H. A., Kaiser N., Peacock J. A., 1994, Astrophys. J., 426, 23
Ho S., et al., 2012, Astrophys. J., 761, 14
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Learejs R., et al., 2011, preprint (arXiv:1110.3193)
Leistedt B., Peiris H. V., 2014, Mon. Not. Roy. Astron. Soc., 444, 2
Leistedt B., Peiris H. V., Roth N., 2014, Phys. Rev. Lett., 113, 221301
Levi M., et al., 2013, preprint (arXiv:1308.0847)
Lewis A., Challinor A., Lasenby A., 2000, Astrophys. J., 538, 473
Ross A. J., et al., 2012, Mon. Not. Roy. Astron. Soc., 424, 564
Ross A. J., et al., 2016, Submitted to: Mon. Not. Roy. Astron. Soc.
Rybicki G. B., Press W. H., 1992, Astrophys. J., 398, 169
Schlegel D., et al., 2011, preprint (arXiv:1106.1706)
Scoccimarro R., 2015, Phys. Rev., D92, 083532
Sherman J., Morrison W. J., 1950, Ann. Math. Stat., 21, 124
**APPENDIX A: DERIVATION OF MODE DEPROJECTION WITH MULTIPLE TEMPLATES**

In this appendix we want to derive Eq. (36) from Eq. (35). We start by rewriting Eq. (35) in matrix notation

\[ \mathbf{C} = \mathbf{C} + \lim_{\sigma \to \infty} \sigma \mathbf{F}_{\text{sys}} \mathbf{f}, \]  

defining an \( N_{\text{mode}} \times N_{\text{sys}} \) matrix \( \mathbf{F}_{\alpha} \equiv f_{\alpha}(k_{\alpha}) \), such that we can invert \( \mathbf{C} \) using the Woodbury matrix identity

\[ \mathbf{C}^{-1} = \mathbf{C}^{-1} - \mathbf{C}^{-1} \lim_{\sigma \to \infty} \sigma \mathbf{F} \left( \mathbf{I}^{N_{\text{sys}}} + \mathbf{F} \mathbf{C}^{-1} \mathbf{F}^T \right)^{-1} \mathbf{F}^T \mathbf{C}^{-1} \]

\[ = \mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{f} \left( \mathbf{f}^T \mathbf{C}^{-1} \mathbf{f} \right)^{-1} \mathbf{f}^T \mathbf{C}^{-1} \]

\[ \equiv \mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{f} \mathbf{f}^T \mathbf{C}^{-1}. \]  

(A2)

If we assume \( \mathbf{C}_{\alpha \beta} = \delta_{\alpha \beta} P(k_{\alpha}) \), \( \mathbf{R} \equiv \mathbf{f} \mathbf{f}^T \) becomes a matrix equivalent to the factor \( R_{\mathbf{f}} \) in previous sections:

\[ \mathbf{R}_{\mathbf{f}} = \sum_{\mu \nu} f_{\mu}^*(k_{\mu}) \frac{\delta_{\mu \nu}}{P(k_{\mu})} f_{\nu}(k_{\nu}) = \sum_{\mu} \frac{f_{\mu}^*(k_{\mu}) f_{\mu}(k_{\mu})}{P(k_{\mu})}. \]

(A3)

The inverse updated covariance matrix then reads

\[ \mathbf{C}_{\alpha \beta}^{-1} = \frac{\delta_{\alpha \beta}}{P(k_{\alpha})} - \sum_{AB} \frac{f_{\alpha}(k_{\alpha}) \mathbf{R}_{AB}^{-1} f_{\beta}(k_{\beta})}{P(k_{\alpha}) P(k_{\beta})}. \]

(A4)

If we do not bin, but apply mode deprojection to each mode separately, the matrix \( \mathbf{E} \) simplifies to

\[ \mathbf{E}_{\alpha \beta} = \sum_{\mu \nu} \mathbf{C}_{\alpha \mu}^{-1} \delta_{\mu \nu} \mathbf{C}_{\nu \beta}^{-1} = \mathbf{C}_{\alpha \beta}^{-1} \]  

where \( \mathbf{C}_{\alpha \beta}^{-1} \) is analogous to \( S_{\alpha \beta}^{-1} \) in the previous appendix.

\[ \mathbf{E}_{\alpha \beta}(k_{\alpha}) = \sum_{\mu \nu} \mathbf{C}_{\alpha \mu}^{-1} \delta_{\mu \nu} \mathbf{C}_{\nu \beta}^{-1} = \mathbf{C}_{\alpha \beta}^{-1}. \]  

(A5)

After inserting Eq. (A4) into Eq. (A5), we obtain

\[ P^{2}(k_{\alpha}) \sum_{\alpha \beta} F^{*}(k_{\alpha}) \mathbf{E}_{\alpha \beta}(k_{\alpha}) F(k_{\beta}) = |F(k_{\alpha})|^2 \]

\[ - \sum_{AB} F^{*}(k_{\alpha}) f_{\alpha}(k_{\alpha}) \mathbf{R}_{AB}^{-1} f_{\beta}(k_{\beta}) F(k_{\beta}) \]

\[ - \sum_{AB} F^{*}(k_{\beta}) f_{\beta}(k_{\beta}) \mathbf{R}_{AB}^{-1} f_{\alpha}(k_{\alpha}) F(k_{\alpha}) \]

\[ + \sum_{ABCD} \frac{F^{*}(k_{\alpha}) f_{\alpha}(k_{\alpha}) \mathbf{R}_{AB}^{-1} f_{\beta}(k_{\beta}) f_{\delta}(k_{\delta}) f_{\beta}(k_{\beta})}{P(k_{\alpha}) P(k_{\beta})} \frac{f_{\delta}(k_{\delta}) F(k_{\delta})}{P(k_{\delta})} \]

\[ = |F(k_{\alpha})|^2 \]

\[ - 2Re \left[ \sum_{AB} S_{A} \mathbf{R}_{AB}^{-1} f_{\beta}(k_{\beta}) F(k_{\beta}) \right] \]

\[ + \sum_{AB} S_{A} \mathbf{R}_{AB}^{-1} f_{\beta}(k_{\beta}) \]

\[ = |F(k_{\alpha})|^2 \sum_{AB} S_{A} \mathbf{R}_{AB}^{-1} f_{\beta}(k_{\beta}) \]

\[ + \sum_{AB} S_{A} \mathbf{R}_{AB}^{-1} f_{\beta}(k_{\beta}) \]

\[ = P^{2}(k_{\alpha}) \sum_{\alpha \beta} F^{*}(k_{\alpha}) \mathbf{E}_{\alpha \beta}(k_{\alpha}) F(k_{\beta}) \]

\[ \equiv \left[ \sum_{\alpha \beta} S_{A} \mathbf{R}_{\alpha \beta}^{-1} f_{\beta}(k_{\beta}) \right]^2, \]  

where we defined \( S_{A} \equiv \sum_{\alpha} f_{\alpha}(k_{\alpha}) \frac{F^{*}(k_{\alpha})}{P(k_{\alpha})} \).  

**APPENDIX B: DERIVATION OF MODE SUBTRACTION WITH MULTIPLE TEMPLATES**

Here we derive the best-fitting \( \epsilon(BF) \) from the likelihood

\[ -2 \ln L = \sum_{\alpha} |F(k_{\alpha}) - \sum_{A} \epsilon_{A} f_{\alpha}(k_{\alpha})|^2 \]

\[ \frac{P(k_{\alpha})}{P(k_{\alpha})} \]

(B1)

to find the same result as in the previous appendix. Taking the derivative with respect to \( \epsilon_{\beta} \) yields

\[ \frac{\partial \chi^2}{\partial \epsilon_{\beta}} = -2 \epsilon_{\beta} \sum_{\alpha} f_{\beta}(k_{\alpha}) F^{*}(k_{\alpha}) - \sum_{A} \epsilon_{A} f_{\beta}(k_{\alpha}) f_{A}(k_{\alpha}) \]

\[ \frac{P(k_{\alpha})}{P(k_{\alpha})} \]

This derivative is zero if

\[ \sum_{\alpha} \frac{f_{\beta}(k_{\alpha}) F^{*}(k_{\alpha})}{P(k_{\alpha})} = \sum_{A} \frac{\epsilon_{A} f_{\beta}(k_{\alpha}) f_{A}(k_{\alpha})}{P(k_{\alpha})}, \]

(B3)

which reads

\[ \mathbf{S} = \mathbf{R} \epsilon. \]

(B4)

in matrix notation. The best fitting \( \epsilon \) value is therefore given by

\[ \epsilon(BF) = \mathbf{R}^{-1} \mathbf{S}. \]

(B5)

The absolute value squared of the best fitting signal is hence equal to Eq. (36):

\[ |F(k_{\alpha}) - \epsilon(BF) f_{\alpha}(k_{\alpha})|^2 = |F(k_{\alpha}) - \sum_{AB} \mathbf{R}_{AB}^{-1} S_{B} f_{A}(k_{\alpha})|^2. \]

(B6)

**APPENDIX C: MODE SUBTRACTION AND THE DEBIAISING STEP FOR A NON-DIAGONAL COVARIANCE MATRIX**

C1 Including the covariance in the calculation of the best-fit mode to subtract

Here we show a derivation similar to the one in Sec. 4 and 5 for the more general case of a non-diagonal covariance matrix. We show that the debiasing works in the same way as in Sec. 5, just with a generalised definition of \( R_{\mathbf{f}} \).

Defining the covariance matrix of the true density

\[ \mathbf{C}_{\alpha \beta} = \langle D_{\alpha} D_{\beta} \rangle \]  

(C1)

As we did in Sec. 4, we introduce a free parameter \( \epsilon \), such that

\[ \mathbf{D}_{\alpha} = f_{\alpha} - \epsilon f_{\alpha}. \]  

(C3)

Assuming that the true density field is Gaussian, its log-likelihood reads

\[ -2 \ln L = \sum_{\alpha \beta} (F_{\alpha} - \epsilon f_{\alpha})^T \mathbf{C}_{\alpha \beta}^{-1} (F_{\beta} - \epsilon f_{\beta}) + \text{const}. \]  

(C4)

To find the best-fitting \( \epsilon(BF) \), we take the derivative of the log-
likelihood with respect to $\epsilon$:

$$-2 \frac{\partial \ln L}{\partial \epsilon} = - \sum_{\alpha \beta} f_\alpha^* C_{\alpha \beta}^{-1} (F_\beta - \epsilon f_\beta) - \sum_{\alpha \beta} (F_\alpha - \epsilon f_\alpha)^* C_{\alpha \beta}^{-1} f_\beta$$

$$= 2 \epsilon \sum_{\alpha \beta} f_\alpha^* C_{\alpha \beta}^{-1} f_\beta - \sum_{\alpha \beta} [f_\alpha^* C_{\alpha \beta}^{-1} f_\beta + f_\alpha^* C_{\alpha \beta}^{-1} f_\beta].$$

(C5)

As $C$ is a covariance matrix of complex random variables, it is Hermitian positive-semidefinite, such that the second sum can be written as

$$\sum_{\alpha \beta} [f_\alpha^* C_{\alpha \beta}^{-1} f_\beta + f_\alpha^* C_{\alpha \beta}^{-1} f_\beta] = 2 \text{Re} \left[ \sum_{\alpha \beta} f_\alpha^* C_{\alpha \beta}^{-1} f_\beta \right].$$

(C6)

For shortness and in analogy to Sec. 4, we call this sum

$$S_P = \sum_{\alpha \beta} \text{Re} \left[ f_\alpha^* C_{\alpha \beta}^{-1} f_\beta \right]$$

(C7)

and the first sum in Eq. (C5) we call

$$R_P = \sum_{\alpha \beta} f_\alpha^* C_{\alpha \beta}^{-1} f_\beta.$$

(C8)

We obtain the best-fitting, i.e. maximum likelihood, value

$$\epsilon^{(BF)} = \frac{S_P}{R_P}$$

(C9)

by equating Eq. (C5) to zero.

Now we want to calculate the expectation value

$$\left\langle \left| F_\alpha - \frac{S_P}{R_P} f_\alpha \right|^2 \right\rangle = \left\langle |F_\alpha|^2 \right\rangle - \frac{2}{R_P} \text{Re} \left[ \langle S_P F_\alpha^* f_\alpha \rangle \right]$$

$$+ \left\langle \frac{S_P^2}{R_P^2} \right\rangle |f_\alpha|^2.$$  

(C10)

We calculate each term separately:

(i) The first term $\langle |F_\alpha|^2 \rangle = C_{\alpha \alpha} + \epsilon_{\text{true}}^2 |f_\alpha|^2$ is a special case of Eq. (C2).

(ii) To calculate the second term, we reexpand $S_P$ and use the fact that $\text{Re} \left[ F_\alpha f_\alpha^* \right] = \text{Re} \left[ F_\alpha^* f_\alpha \right]$:

$$2 \text{Re} \left[ \langle S_P F_\alpha^* f_\alpha \rangle \right] = 2 \text{Re} \left[ \sum_{\gamma \beta} f_\gamma^* C_{\gamma \beta}^{-1} (F_\beta f_\alpha^*) f_\alpha \right].$$

(iii) In the third term, we can again make use of Eq. (C2):

$$\langle S_P^2 \rangle = \sum_{\alpha \beta} \text{Re} \left[ f_\alpha^* C_{\alpha \beta}^{-1} F_\beta F_\alpha^* \right]$$

$$= \sum_{\alpha \beta} \text{Re} \left[ f_\alpha^* C_{\alpha \beta}^{-1} F_\beta C_{\gamma \delta}^{-1} F_\gamma F_\delta \right]$$

$$+ \sum_{\alpha \beta} \epsilon_{\text{true}}^2 |f_\alpha^*|^2 R_P.$$  

(C14)

In the first term, we have again $\sum_{\beta} C_{\alpha \beta}^{-1} C_{\beta \gamma} = \delta_{\alpha \gamma}$, and the second term is equal to $R_P^2$, such that

$$\langle S_P^2 \rangle = \sum_{\alpha \gamma} \text{Re} \left[ f_\alpha^* f_\gamma C_{\gamma \alpha}^{-1} \right] + \epsilon_{\text{true}}^2 R_P^2 = R_P + \epsilon_{\text{true}}^2 R_P^2$$

(C15)

Recollecting 1.-3. and inserting into Eq. (C10) yields

$$\left\langle \left| F_\alpha - \frac{S_P}{R_P} f_\alpha \right|^2 \right\rangle = C_{\alpha \alpha} + \epsilon_{\text{true}}^2 |f_\alpha|^2$$

$$- 2 |f_\alpha|^2 \left( \epsilon_{\text{true}}^2 + \frac{1}{R_P} \right) + \left( \epsilon_{\text{true}}^2 + \frac{1}{R_P} \right) |f_\alpha|^2$$

$$= C_{\alpha \alpha} - \frac{|f_\alpha|^2}{R_P}.$$  

(C16)

As the power spectrum $P(k_\alpha) = C_{\alpha \alpha}$ is defined as the diagonal elements of the covariance matrix, the debiasing step is the same for a non-diagonal covariance matrix as for a diagonal one (cf. Sec. 5), we just have to use the generalised definition of $R_P$ as in Eq. (C8).

C2 The independent mode approximation

We have seen in Appendix C1 that mode subtraction also works when the covariance matrix is non-diagonal. However, to compute the generalised $R_P$, one has to invert the full $N_{\text{mode}} \times N_{\text{mode}}$ covariance matrix, which makes this approach computationally almost as expensive as using the QML estimator. We will argue that, in most cases, Eq. (31) provides a good estimate of the power, even in the presence of covariant modes, and we will provide a further correction term that corrects for using Eq. (31) when off-diagonal covariances are important.

Suppose we apply Eq. (31) assuming a diagonal covariance matrix, even though there are covariances between different modes. Then, we find a best fitting

$$\epsilon_{\text{BF}} = \frac{\sum_{\alpha \beta} f_\alpha^* f_\beta}{\sum_{\alpha \beta} |f_\beta|^2 R_P}$$

(i) instead of the true

$$\epsilon_{\text{BF}} = \frac{\sum_{\alpha \beta} f_\alpha^* C_{\alpha \beta}^{-1} f_\beta}{\sum_{\alpha \beta} f_\alpha^* C_{\alpha \beta}^{-1} f_\beta}.$$  

(C18)

The expectations are the same $\langle \epsilon_{\text{BF}}' \rangle = \langle \epsilon_{\text{BF}} \rangle = \epsilon_{\text{true}}$, but their variances are different. For the approximate estimate we have

$$\langle \epsilon_{\text{BF}}^2 \rangle = \frac{\sum_{\alpha \beta} f_\alpha^* f_\alpha f_\beta f_\beta^*}{R_P^2}$$

$$= \frac{\sum_{\alpha \beta} f_\alpha^* f_\beta^*}{R_P^2} + \frac{\sum_{\alpha \beta} \epsilon_{\text{true}}^2 |f_\alpha|^2 |f_\beta|^2}{R_P^2}$$

$$= \frac{1}{R_P^2} \sum_{\alpha \beta} f_\alpha C_{\alpha \beta} f_\beta^* + \epsilon_{\text{true}}.$$  

(C19)
Unlike in the previous estimates, the covariance matrix does not cancel in the first term. Similarly,

\[ \langle \varepsilon'_B F^*_\alpha F_\alpha \rangle = \frac{1}{R'_P} \left\{ \sum_\beta \frac{f_\beta F^*_\beta F_\alpha}{P_\beta} \right\} = \frac{1}{R'_P} \sum_\beta \frac{f_\alpha C_{\alpha\beta} f^*_\beta}{P_\beta} + \varepsilon^2_{\text{true}} |f_\alpha|^2. \quad (C20) \]

Combining the previous two equations, we obtain

\[ \langle |F_\alpha - \varepsilon'_B f_\alpha|^2 \rangle = C_{\alpha\alpha} - \frac{2}{R'_P} \sum_\beta \frac{f_\alpha C_{\alpha\beta} f^*_\beta}{P_\beta} + \frac{|f_\alpha|^2}{R'_P} \sum_\gamma \sum_\beta \frac{f_\gamma C_{\gamma\beta} f^*_\beta}{P_\gamma P_\beta}. \quad (C21) \]

Splitting the covariance matrix

\[ C_{\alpha\beta} = P_\beta (\delta_{\alpha\beta} + \Delta_{\alpha\beta}) \quad (C22) \]

into a diagonal and off-diagonal elements yields

\[ \langle |F_\alpha - \varepsilon'_B f_\alpha|^2 \rangle = P_\alpha - \frac{|f_\alpha|^2}{R'_P} \left[ 1 + \sum_\gamma \sum_\beta f_\gamma \Delta_{\gamma\beta} f^*_\beta \left( \frac{2\delta_{\alpha\gamma}}{|f_\alpha|^2} - \frac{1}{R'_P P_\gamma} \right) \right]. \quad (C23) \]

Hence, one can perform mode subtraction assuming a diagonal covariance matrix and then apply another correction term which is linear in its off-diagonal elements. The advantage of this procedure is that it does not require any inversion of the \( N^2 \) mode covariance matrix. If the off-diagonal elements are small, then the bias correction reverts back to the form of Eq. (30).

This paper has been typeset from a \TeX/L\LaTeX file prepared by the author.