

hera_pspec Power Spectrum Normalization

by Adrian Liu and the HERA
pspec team

June 4, 2018

Abstract

We derive the power spectrum normalization scalar that relates the dimensionality of the output of `hera_pspec.oqe` module to the dimensionality of the input data. This is derived for the numerical routines as they appear in `hera_pspec` under GIT hash .

1 Conversion from Jy to mK

An interferometric visibility is inherently in units of Jansky (10^{-23} ergs sec $^{-1}$ cm $^{-2}$ Hz $^{-1}$). Visibilities measured by a correlator will almost always be calibrated to a Jansky scale, meaning that after calibration our visibilities will still be in units of Jansky. To convert them to units of milli-Kelvin, we use the definition of the brightness temperature from the Rayleigh-Jeans Law:

$$I_\nu = \frac{2\nu^2 k_b T_b}{c^2} = \frac{2k_b T_b}{\lambda^2} \quad (1)$$

where I is in units of specific intensity (ergs sec $^{-1}$ cm $^{-2}$ Hz $^{-1}$ steradians $^{-1}$) and T is in units of Kelvin. This, however, is a mapping from specific intensity to temperature. What we require is a mapping from Jansky (flux density) to temperature, which we can accomplish by first mapping Jansky to specific intensity. This is done by dividing by the integral of the beam power response, commonly referred to as Ω_p in the literature:

$$\Omega_p(\nu) = \int_{4\pi} d\Omega A(\hat{s}, \nu) \quad (2)$$

where A is unitless, direction-dependent scalar normalized to unity at boresight, and Ω_p carries units of radians 2 or steradians.

To summarize, the full conversion from a visibility on a Jansky scale to a visibility on a mK scale is

$$V_{\text{mK}}(\lambda) = V_{\text{Jy}}(\lambda) 10^{-23} \frac{\lambda^2}{2k_b \Omega_p(\lambda)} 10^3, \quad (3)$$

where the 10^{-23} comes from converting Jansky to flux density in ergs sec $^{-1}$ cm $^{-2}$ Hz $^{-1}$, and the 10^3 comes from the Kelvin to milli-Kelvin conversion.

2 Normalization conventions in the quadratic estimator

Once we have performed the unit conversions outlined in the previous section, our measurement equation is

$$V_b(\nu) = \frac{\gamma(\nu)}{\Omega_p(\nu)} \int d\Omega T(\hat{\mathbf{n}}, \nu) A(\hat{\mathbf{n}}, \nu) e^{-i\frac{2\pi\mathbf{b}\cdot\hat{\mathbf{n}}}{\lambda}}, \quad (4)$$

where $d\Omega$ is a differential solid angle in steradians, T is the brightness temperature on the sky, and A is the primary beam. The (optional) function $\gamma(\nu)$ is some tapering function (e.g., Blackman-Harris) that can be imposed by the data analyst. For the purposes of deriving the normalization conventions, there is no difference between invoking the flat-sky approximation and a full curved-sky treatment. Taking the flat-sky limit, we can say

$$V_b(\nu) = \frac{\gamma(\nu)}{\Omega_p(\nu)} \int d^2\theta T(\boldsymbol{\theta}, \nu) A(\boldsymbol{\theta}, \nu) e^{-i2\pi\mathbf{u}_b\cdot\boldsymbol{\theta}}, \quad (5)$$

where we have also defined $\mathbf{u}_b \equiv \mathbf{b}/\lambda$. Henceforth we will neglect the frequency dependence of \mathbf{u}_b , which is tantamount to ignoring “wedge physics”. This also has no effect on the normalizations. For later convenience, we can also write this relation in Fourier space so that

$$V_b(\nu) = \int d\eta d^2u \frac{\gamma(\nu)}{\Omega_p(\nu)} \tilde{T}(\mathbf{u}, \eta) \tilde{A}(\mathbf{u}_b - \mathbf{u}, \nu) e^{i2\pi\eta\nu}, \quad (6)$$

where η is the Fourier dual to ν , \mathbf{u} is the Fourier dual to $\boldsymbol{\theta}$, and

$$\tilde{T}(\mathbf{u}, \eta) \equiv \int d^2\theta d\nu e^{-i2\pi\mathbf{u}\cdot\boldsymbol{\theta}} e^{-i2\pi\eta\nu} T(\boldsymbol{\theta}, \nu), \quad (7)$$

with \tilde{A} similarly defined as the Fourier transform of A .

We first define a power spectrum \bar{P} in “observer units”. This power spectrum does not depend on cosmological parameters, and is therefore a helpful intermediate quantity. This power spectrum is defined by the relation

$$\langle \tilde{T}(\mathbf{u}, \eta) \tilde{T}^*(\mathbf{u}', \eta') \rangle \equiv \delta^D(\mathbf{u} - \mathbf{u}') \delta^D(\eta - \eta') \bar{P}(\mathbf{u}, \eta), \quad (8)$$

where $\langle \dots \rangle$ signifies an ensemble average and δ^D represents a Dirac delta function. This definition of \bar{P} agrees with how one would define a “naive delay spectrum” for an instrument with a tophat primary beam.

The basic building block of the quadratic estimator is the covariance matrix $\mathbf{C} \equiv \langle \mathbf{x}\mathbf{x}^\dagger \rangle$ of our data \mathbf{x} . If each element of the vector is a different frequency channel of a single baseline’s visibility, then the components of the covariance matrix are given by

$$\begin{aligned} \mathbf{C}_{ij} &\equiv \langle V_b(\nu_i) V_b^*(\nu_j) \rangle \\ &= \int d\eta d^2u \bar{P}(\mathbf{u}, \eta) e^{i2\pi\eta(\nu_i - \nu_j)} \tilde{A}(\mathbf{u}_b - \mathbf{u}, \nu_i) \tilde{A}^*(\mathbf{u}_b - \mathbf{u}, \nu_j) \frac{\gamma(\nu_i)}{\Omega_p(\nu_i)} \frac{\gamma(\nu_j)}{\Omega_p(\nu_j)}. \end{aligned} \quad (9)$$

We now make two approximations. The first is that the primary beam is fairly broad, so that the \tilde{A} is compact as a function of \mathbf{u} . This means that the integral receives most of its contribution from $\mathbf{u} \approx \mathbf{u}_b$, and we can evaluate $\bar{P}(\mathbf{u}, \eta)$ there, factoring it out of the integral:

$$\mathbf{C}_{ij} \approx \int d\eta \bar{P}(\mathbf{u}_b, \eta) e^{i2\pi\eta(\nu_i - \nu_j)} \int d^2u \tilde{A}(\mathbf{u}_b - \mathbf{u}, \nu_i) \tilde{A}^*(\mathbf{u}_b - \mathbf{u}, \nu_j) \frac{\gamma(\nu_i)}{\Omega_p(\nu_i)} \frac{\gamma(\nu_j)}{\Omega_p(\nu_j)}. \quad (10)$$

Next we assume that \bar{P} is piecewise constant over some discrete bins in η , which gives

$$\begin{aligned} \mathbf{C}_{ij} &\approx \sum_{\alpha} \bar{P}_{\alpha} \int_{\eta_{\alpha}} d\eta e^{i2\pi\eta(\nu_i - \nu_j)} \int d^2u \tilde{A}(\mathbf{u}_b - \mathbf{u}, \nu_i) \tilde{A}^*(\mathbf{u}_b - \mathbf{u}, \nu_j) \frac{\gamma(\nu_i)}{\Omega_p(\nu_i)} \frac{\gamma(\nu_j)}{\Omega_p(\nu_j)} \\ &\approx \sum_{\alpha} \bar{P}_{\alpha} \Delta\eta e^{i2\pi\eta_{\alpha}(\nu_i - \nu_j)} \int d^2u \tilde{A}(\mathbf{u}_b - \mathbf{u}, \nu_i) \tilde{A}^*(\mathbf{u}_b - \mathbf{u}, \nu_j) \frac{\gamma(\nu_i)}{\Omega_p(\nu_i)} \frac{\gamma(\nu_j)}{\Omega_p(\nu_j)} \\ &\approx \sum_{\alpha} \bar{P}_{\alpha} \Delta\eta e^{i2\pi\eta_{\alpha}(\nu_i - \nu_j)} \int d^2\theta A(\theta, \nu_i) A(\theta, \nu_j) \frac{\gamma(\nu_i)}{\Omega_p(\nu_i)} \frac{\gamma(\nu_j)}{\Omega_p(\nu_j)} \end{aligned} \quad (11a)$$

$$\equiv \sum_{\alpha} \bar{P}_{\alpha} \mathbf{Q}_{ij}^{\alpha}. \quad (11b)$$

The matrix \mathbf{Q}^{α} is defined as $\partial\mathbf{C}/\partial\bar{P}_{\alpha}$, and encodes the response of the data covariance to the α th bandpower \bar{P}_{α} . It is the bridge between the input vector space that data vectors inhabit and the output vector space that bandpowers inhabit.

The quadratic estimator formalism instructs us to form an estimate of the power spectrum bandpowers by first computing

$$\hat{q}_{\alpha} \equiv \frac{1}{2} \mathbf{x}^{\dagger} \mathbf{Q}^{\alpha} \mathbf{x}. \quad (12)$$

Taking the expectation value of this gives

$$\langle \hat{q}_{\alpha} \rangle = \frac{1}{2} \text{tr}(\mathbf{Q}^{\alpha} \mathbf{C}) = \frac{1}{2} \sum_{\beta} \text{tr}(\mathbf{Q}^{\alpha} \mathbf{Q}^{\beta}) \bar{P}_{\beta}. \quad (13)$$

This shows that on average, \hat{q}_{α} measures a weighted sum of a true bandpowers. To arrive at a properly normalized power spectrum, the weights of the sum must add to unity. Alternatively, we can simply divide our estimator with the sum of our weights, obtaining

$$\widehat{\bar{P}}_{\alpha} = \frac{\mathbf{x}^{\dagger} \mathbf{Q}^{\alpha} \mathbf{x}}{\sum_{\gamma} \text{tr}(\mathbf{Q}^{\alpha} \mathbf{Q}^{\gamma})}. \quad (14)$$

If we computed our power spectra in this way, our answers would be automatically normalized by construction. However, this is *not* what we typically do in our power spectrum pipelines. Instead of comparing Equations (11a) and (11b)

to obtain the \mathbf{Q}^α that is inserted into Equation (12), we observe that the main function of \mathbf{Q}^α is to Fourier transform the two data vectors that act on it and define

$$\mathbf{Q}_{ij}^{\text{alt},\alpha} \equiv D e^{i2\pi\eta_\alpha(\nu_i - \nu_j)}, \quad (15)$$

where D is some constant that we can pick later. We then insert this into our quadratic form in place of \mathbf{Q}^α so that

$$\hat{q}_\alpha^{\text{alt}} \equiv \frac{1}{2} \mathbf{x}^\dagger \mathbf{Q}^{\text{alt},\alpha} \mathbf{x}, \quad (16)$$

which has an expectation value of

$$\langle \hat{q}_\alpha \rangle = \frac{1}{2} \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{C}) = \frac{1}{2} \sum_\beta \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^\beta) \bar{P}_\beta. \quad (17)$$

With this expression, one would normally write an estimator of the form

$$\hat{P}_\alpha^{\text{alt}} = \frac{\mathbf{x}^\dagger \mathbf{Q}^{\text{alt},\alpha} \mathbf{x}}{\sum_\gamma \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^\gamma)}. \quad (18)$$

It turns out that this *still* isn't what's done in the code! Instead, the code simply assumes that it is correct to replace \mathbf{Q}^α with every instance of $\mathbf{Q}^{\text{alt},\alpha}$ (ignoring the fact that Equation 17, for instance, has \mathbf{Q}^α and $\mathbf{Q}^{\text{alt},\alpha}$ serving distinct roles). To compensate for the fact that this is incorrect, a final normalization factor κ is inserted, so that our expression becomes

$$\hat{P}_\alpha^{\text{code}} = \frac{\mathbf{x}^\dagger \mathbf{Q}^{\text{alt},\alpha} \mathbf{x}}{\kappa \sum_\gamma \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^{\text{alt},\gamma})}. \quad (19)$$

Comparing the expression for $\hat{P}_\alpha^{\text{alt}}$ with the expression for $\hat{P}_\alpha^{\text{code}}$, we see that $\hat{P}_\alpha^{\text{code}}$ will only give correctly normalized power spectra if

$$\kappa \sum_\gamma \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^{\text{alt},\gamma}) = \sum_\gamma \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^\gamma). \quad (20)$$

We may use this to solve for the correct κ and/or D . (There is some freedom as to where one inserts the correction factors). Evaluating the LHS, we have

$$\begin{aligned} \kappa \sum_\gamma \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^{\text{alt},\gamma}) &= \kappa \sum_{\gamma ij} \mathbf{Q}_{ij}^{\text{alt},\alpha} \mathbf{Q}_{ji}^{\text{alt},\gamma} = \kappa D^2 \sum_{\gamma ij} e^{i2\pi\eta_\alpha(\nu_i - \nu_j)} e^{i2\pi\eta_\gamma(\nu_j - \nu_i)} \\ &= \kappa D^2 N_{\text{freq}} \sum_{ij} e^{i2\pi\eta_\alpha(\nu_i - \nu_j)} \delta_{ij} = \kappa D^2 N_{\text{freq}}^2, \end{aligned} \quad (21)$$

where in evaluating the sum over γ , we used the symmetries of the discrete Fourier transform to say that $\sum_\gamma e^{i2\pi\eta_\gamma(\nu_j - \nu_i)} = N_{\text{freq}} \delta_{ij}$. The RHS evaluates

to

$$\begin{aligned}
\sum_{\gamma} \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^{\gamma}) &= \sum_{\gamma ij} \mathbf{Q}_{ij}^{\text{alt},\alpha} \mathbf{Q}_{ji}^{\gamma} \\
&= D\Delta\eta \sum_{\gamma ij} e^{i2\pi\eta_{\alpha}(\nu_i - \nu_j)} e^{i2\pi\eta_{\gamma}(\nu_j - \nu_i)} \int d^2\theta A(\boldsymbol{\theta}, \nu_j) A(\boldsymbol{\theta}, \nu_i) \\
&= D\Delta\eta N_{\text{freq}} \sum_{ij} e^{i2\pi\eta_{\alpha}(\nu_i - \nu_j)} \delta_{ij} \int d^2\theta A(\boldsymbol{\theta}, \nu_i) A(\boldsymbol{\theta}, \nu_j) \frac{\gamma(\nu_i)}{\Omega_p(\nu_i)} \frac{\gamma(\nu_j)}{\Omega_p(\nu_j)} \\
&= D\Delta\eta N_{\text{freq}} \sum_i \frac{\gamma^2(\nu_i)}{\Omega_p^2(\nu_i)} \int d^2\theta A^2(\boldsymbol{\theta}, \nu_i). \tag{22}
\end{aligned}$$

Equating the two expressions then gives

$$\kappa D = \Delta\eta \frac{1}{N_{\text{freq}}} \sum_i \frac{\gamma^2(\nu_i)}{\Omega_p^2(\nu_i)} \int d^2\theta A(\boldsymbol{\theta}, \nu_i)^2. \tag{23}$$

Now, for a numerical FFT, $\Delta\eta = 1/B$ and $N_{\text{freq}} = B/\Delta\nu$. Thus, one obtains

$$\kappa D = \frac{1}{B^2} \sum_i \Delta\nu \frac{\gamma^2(\nu_i)}{\Omega_p^2(\nu_i)} \int d^2\theta A(\boldsymbol{\theta}, \nu_i)^2 \approx \frac{1}{B^2} \int d\nu \frac{\gamma^2(\nu)}{\Omega_p^2(\nu)} \int d^2\theta A(\boldsymbol{\theta}, \nu)^2. \tag{24}$$

This is precisely the form of the “usual” power spectrum scalar (minus the cosmological scalings), as defined in previous memos. Since κ and D are complete degenerate here, one can choose for simplicity to set $D = 1$. This is what is done in the code.

Note that if we write out the form of our estimator, we have

$$\widehat{P}_{\alpha}^{\text{code}} = \frac{\sum_{ij} e^{i2\pi\eta_{\alpha}(\nu_i - \nu_j)} V_b(\nu_j) V_b^*(\nu_i)}{\kappa N_{\text{freq}}^2} = \frac{1}{\kappa} \left| \frac{1}{N_{\text{freq}}} \sum_i e^{i2\pi\eta_{\alpha}\nu_i} V_b^*(\nu_i) \right|^2. \tag{25}$$

This proves a piece of lore from before—that a correctly normalized power spectrum estimate can be obtained by using a *inverse* Fourier transform (which divides by N_{freq} under `numpy`’s Fourier convention) on complex conjugated data, squaring, and then dividing by κ .

The thoroughly confusing aspect about this is that in the `hera_pspec.oqe` module, one does not see any inverse Fourier transforms. This is because Equation (15) is the forward Fourier transform (i.e., without the factor of N_{freq}) if $D = 1$, so the `use_fft` option of the `get_Q` function implements it as such. What ends up happening is that the normalization factor ends up being proportional to $1/N_{\text{freq}}^2$, which effectively converts the forward Fourier transforms into inverse Fourier transforms (up to a complex conjugation).

Note that if we had wanted, we *could’ve* set $D = 1/N_{\text{freq}}^2$ so that $\mathbf{x}^{\dagger} \mathbf{Q}^{\alpha} \mathbf{x}$ corresponds to an inverse FFT. Following the mathematics through, the denominator of Equation (19) then just reduces to only the power spectrum scalar, so everything is consistent.

3 Window functions

Taking the ensemble average of Equation (19), one can show that

$$\langle \widehat{P}_\alpha^{\text{code}} \rangle = \frac{\sum_\beta \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^\beta) \overline{P}_\beta}{\kappa \sum_\gamma \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^{\text{alt},\gamma})}. \quad (26)$$

This means that the window functions are given by

$$W_{\alpha\beta} = \frac{\text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^\beta)}{\kappa \sum_\gamma \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^{\text{alt},\gamma})}. \quad (27)$$

4 What if there are weighting matrices?

In general, one can imagine a more general estimator that weights the data, such that we have

$$\hat{q}_\alpha^{\text{alt}} \equiv \frac{1}{2} \mathbf{x}^\dagger \mathbf{R}_1 \mathbf{Q}^{\text{alt},\alpha} \mathbf{R}_2 \mathbf{x} \quad (28)$$

rather than Equation (16). When one uses non-trivial (i.e., non-identity) weighting matrices like this, the code ends up forming

$$\widehat{P}_\alpha^{\text{code}} = \frac{\mathbf{x}^\dagger \mathbf{R}_1 \mathbf{Q}^{\text{alt},\alpha} \mathbf{R}_2 \mathbf{x}}{\kappa \sum_\gamma \text{tr}(\mathbf{R}_1 \mathbf{Q}^{\text{alt},\alpha} \mathbf{R}_2 \mathbf{Q}^{\text{alt},\gamma})}. \quad (29)$$

Going through the same exercise as before, we need to make sure that this is equal to

$$\widehat{P}_\alpha^{\text{alt}} = \frac{\mathbf{x}^\dagger \mathbf{R}_1 \mathbf{Q}^{\text{alt},\alpha} \mathbf{R}_2 \mathbf{x}}{\sum_\gamma \text{tr}(\mathbf{R}_1 \mathbf{Q}^{\text{alt},\alpha} \mathbf{R}_2 \mathbf{Q}^\gamma)}. \quad (30)$$

Equating these, we end up with

$$\kappa = \frac{\sum_\gamma \text{tr}(\mathbf{R}_1 \mathbf{Q}^{\text{alt},\alpha} \mathbf{R}_2 \mathbf{Q}^\gamma)}{\sum_\gamma \text{tr}(\mathbf{R}_1 \mathbf{Q}^{\text{alt},\alpha} \mathbf{R}_2 \mathbf{Q}^{\text{alt},\gamma})}. \quad (31)$$

For general \mathbf{R}_1 , \mathbf{R}_2 , tapers, and primary beams, this will *not* reduce to a simple scalar independent of α , \mathbf{R}_1 and \mathbf{R}_2 . Being able to normalize out the primary beam effects, tapers, etc. with a scalar only works if

- $\mathbf{R}_1 = \mathbf{R}_2 = \mathbf{I}$. This is the limiting case from the previous section.

OR

- No tapering function is used (i.e., $\gamma(\nu) = 1$) and the primary beams can be approximated as being frequency-independent.

There is one further case where the scalar is independent of α (though still dependent on \mathbf{R}_1 and \mathbf{R}_2). This is when \mathbf{R}_1 and \mathbf{R}_2 are both diagonal. Suppose $R_{1,ij} \equiv r_i^{(1)} \delta_{ij}$ and similarly for \mathbf{R}_2 . Then κ reduces to

$$\kappa = \frac{1}{B} \frac{\sum_i r_i^{(1)} r_i^{(2)} \gamma^2(\nu_i) \int d^2\theta A(\boldsymbol{\theta}, \nu)^2 / \Omega_p^2(\nu_i)}{\sum_j r_j^{(1)} r_j^{(2)}}. \quad (32)$$

This is very similar to the scalar normalization from before, except that the integrals over frequency from before are weighted by the weighting functions.

5 Some notation

In the code, we employ the notation

$$G_{\alpha\beta} \equiv \frac{1}{2} \text{tr} (\mathbf{R}_1 \mathbf{Q}^{\text{alt},\alpha} \mathbf{R}_2 \mathbf{Q}^{\text{alt},\beta}) \quad (33)$$

and

$$H_{\alpha\beta} \equiv \frac{1}{2} \text{tr} (\mathbf{R}_1 \mathbf{Q}^{\text{alt},\alpha} \mathbf{R}_2 \mathbf{Q}^{\text{tapered},\beta}) \quad (34)$$

where

$$\mathbf{Q}^{\text{tapered},\beta} \equiv e^{i2\pi\eta_\alpha(\nu_i - \nu_j)} \gamma(\nu_i) \gamma(\nu_j) \quad (35)$$

6 Accounting for a different number of delay bins

Implementing the Fourier transforms as FFTs requires that the number of delay bins N_{dlys} is equal to N_{freq} . We may want to use fewer delay bins, so that each bin is thicker. In general, the quadratic estimator formalism allows arbitrary bin sizes. These bins can even be non-uniform. However, it turns out that for our single scalar normalization factor κ to work, the bins must be uniform in size. (Which makes sense—intuitively, if we had one bin that was twice as wide as all other bins, one would need to account for that in the normalization). When using evenly spaced bins, we pick

$$\eta_\alpha \equiv \frac{\alpha}{N_{\text{dlys}} \Delta\nu}. \quad (36)$$

This has implications for computing the scalar normalization. Our previous derivation used the fact that

$$\sum_{\gamma} e^{i2\pi\eta_\gamma(\nu_j - \nu_i)} = N_{\text{freq}} \delta_{ij}, \quad (37)$$

but this is *not* true in general any more (even if we replace N_{freq} with N_{dlys}). To see this, let us explicitly evaluate the sum

$$S_{ij} \equiv \sum_{\gamma} e^{i2\pi\eta_\gamma(\nu_j - \nu_i)} = \sum_{\gamma=0}^{N_{\text{dlys}}-1} e^{i2\pi\gamma(i-j)/N_{\text{dlys}}}, \quad (38)$$

where in the last equality we explicitly put in our new η spacings. If $i = j$, then one does get N_{dlys} . However, this sum can also evaluate to non-zero values for certain $i \neq j$ combinations. Recall that i and j are indices for frequency, so

they go from 0 to $N_{\text{freq}} - 1$. Since $N_{\text{dlys}} < N_{\text{freq}}$, it is possible for $(i - j)/N_{\text{dlys}}$ to be equal to an integer. If such a condition is fulfilled, one sees that the sum will also evaluate to N_{dlys} . If the condition is not fulfilled, we have a geometric series that can be summed explicitly to give. If $i \neq j$, the sum is a geometric series that can be summed explicitly to give

$$S_{ij} = \frac{1 - e^{i2\pi(i-j)}}{1 - e^{i2\pi\gamma(i-j)/N_{\text{dlys}}}} = 0. \quad (39)$$

This means that

$$S_{ij} = \begin{cases} N_{\text{dlys}} & \text{if } (i - j)/N_{\text{dlys}} \text{ is an integer} \\ 0 & \text{otherwise.} \end{cases} \quad (40)$$

This is less convenient than what we had before, but that's ok. Pushing forward, we can compute one of the pieces that we need for the scalar normalization (see Equation 20):

$$\text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^{\text{alt},\gamma}) = \sum_{ij} e^{i2\pi\eta\alpha(\nu_i - \nu_j)} S_{ij} = \sum_{ij} e^{i2\pi\alpha(i-j)/N_{\text{dlys}}} S_{ij}. \quad (41)$$

Notice that regardless of α , the complex exponential factor is equal to 1 whenever $S_{ij} = N_{\text{dlys}}$, and 0 whenever $S_{ij} = 0$. This implies that $\text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^{\text{alt},\gamma})$ is independent of the value of α .

The other piece that we need is $\sum_{\gamma} \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^{\gamma})$. With very thick bins, it is prudent to revisit one of the approximations that we made in forming \mathbf{Q}^{γ} . In particular, we made the approximation that

$$\int_{\eta_{\alpha}} d\eta e^{i2\pi\eta(\nu_i - \nu_j)} \approx \Delta\eta e^{i2\pi\eta_{\alpha}(\nu_i - \nu_j)}, \quad (42)$$

which is appropriate when the bins are thin. To be more exact when the bins are thick, one can just evaluate the integral explicitly to give

$$\int_{\eta_{\alpha}} d\eta e^{i2\pi\eta(\nu_i - \nu_j)} = \Delta\eta e^{i2\pi\eta_{\alpha}(\nu_i - \nu_j)} \text{sinc}[\pi\Delta\eta(\nu_i - \nu_j)]. \quad (43)$$

Redoing our derivation for $\sum_{\gamma} \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^{\gamma})$ with this expression gives

$$\begin{aligned} \sum_{\gamma} \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^{\gamma}) &= \sum_{ij} e^{i2\pi\alpha(i-j)/N_{\text{dlys}}} S_{ij} \Delta\eta \text{sinc}[\pi\Delta\eta(\nu_i - \nu_j)] \\ &\quad \times \int d^2\theta A(\boldsymbol{\theta}, \nu_i) A(\boldsymbol{\theta}, \nu_j) \frac{\gamma(\nu_i)}{\Omega_p(\nu_i)} \frac{\gamma(\nu_j)}{\Omega_p(\nu_j)}, \end{aligned} \quad (44)$$

which by the same logic as above, must be independent of α .

The fact that $\sum_{\gamma} \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^{\gamma})$ and $\sum_{\gamma} \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^{\text{alt},\gamma})$ are independent of α is crucial, because it implies that our normalizing constant

$$\kappa \equiv \frac{\sum_{\gamma} \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^{\gamma})}{\sum_{\gamma} \text{tr}(\mathbf{Q}^{\text{alt},\alpha} \mathbf{Q}^{\text{alt},\gamma})} \quad (45)$$

is also just a constant, independent of α . This was built into the infrastructure of our code.

One handy feature of our new expressions is that even though κ is now in general more complicated, if the primary beam is frequency-independent over our band, our new normalization constant κ_{new} can be related to our old normalization constant κ_{old} by some multiplicative factor that does not involve the beam.

With a frequency-independent beam, the old normalization constant takes the form

$$\kappa_{\text{old}} = \frac{1}{BN_{\text{freq}}} \left(\int d^2\theta \frac{A(\boldsymbol{\theta})^2}{\Omega_p^2} \right) \sum_i \gamma^2(\nu_i), \quad (46)$$

while the new normalization is of the form

$$\kappa_{\text{new}} = \left(\int d^2\theta \frac{A(\boldsymbol{\theta})^2}{\Omega_p^2} \right) \frac{\sum_{ij} e^{i2\pi\alpha(i-j)/N_{\text{dlys}}} S_{ij} \Delta\eta \text{sinc}[\pi\Delta\eta(\nu_i - \nu_j)] \gamma(\nu_i)\gamma(\nu_j)}{\sum_{ij} e^{i2\pi\alpha(i-j)/N_{\text{dlys}}} S_{ij}}. \quad (47)$$

Using the fact that $\Delta\eta$ is now $1/(N_{\text{dlys}}\Delta\nu)$ rather than $1/(N_{\text{freqs}}\Delta\nu)$, we have

$$\frac{\kappa_{\text{new}}}{\kappa_{\text{old}}} = \frac{N_{\text{freq}}^2}{N_{\text{dlys}}} \frac{\sum_{ij} e^{i2\pi\alpha(i-j)/N_{\text{dlys}}} S_{ij} \text{sinc}[\pi\Delta\eta(\nu_i - \nu_j)] \gamma(\nu_i)\gamma(\nu_j)}{[\sum_i \gamma^2(\nu_i)] \left(\sum_{ij} e^{i2\pi\alpha(i-j)/N_{\text{dlys}}} S_{ij} \right)}, \quad (48)$$

so the primary beam cancels out in this adjustment factor.

7 Converting to cosmological coordinates

The power spectrum \bar{P} defined in Equation (8) can be thought of as the power spectrum in “telescope coordinates”, with units of $\text{mK}^2 \text{Sr MHz}$. Ultimately, we are interested in a power spectrum P in cosmological coordinates, with units of $\text{mK}^2 h^{-3} \text{Mpc}^3$. Such a power spectrum is defined by the relation

$$\langle \check{T}(\mathbf{k}_\perp, k_\parallel) \check{T}^*(\mathbf{k}'_\perp, k'_\parallel) \rangle \equiv (2\pi)^3 \delta^D(\mathbf{k} - \mathbf{k}') P(\mathbf{k}_\perp, k_\parallel), \quad (49)$$

where

$$\check{T}(\mathbf{k}_\perp, k_\parallel) \equiv \int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}} T(\mathbf{r}) \quad \text{and} \quad T(\mathbf{r}) = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{r}} \check{T}(\mathbf{k}_\perp, k_\parallel). \quad (50)$$

To translate between \bar{P} and P , two things need to be done:

1. The \mathbf{u} and η dependencies need to be mapped to \mathbf{k}_\perp and k_\parallel respectively, using the relations

$$\mathbf{k}_\perp = \frac{2\pi\mathbf{u}}{D_c} \equiv \frac{2\pi\mathbf{u}}{X} \quad \text{and} \quad k_\parallel = \frac{2\pi\nu_{21}H_0E(z)}{c(1+z)^2} \eta \equiv \frac{2\pi\eta}{Y}. \quad (51)$$

2. The estimated \bar{P} needs to be multiplied by X^2Y , i.e.,

$$P(\mathbf{k}_\perp, k_\parallel) = \frac{c(1+z)^2 D_c^2}{\nu_{21} H_0 E(z)} \bar{P}(\mathbf{u}, \eta) = X^2 Y \bar{P}(\mathbf{u}, \eta). \quad (52)$$

Derivations of these results are provided in Appendix A of Liu et al., PRD **90**, 023018 (2014), so we won't repeat them here. Note that in principle, X^2Y is frequency-dependent. Liu et al., ApJ **833**, 242 (2016) shows that this can be dealt with by absorbing the X^2Y term into the scalar correction factor from the previous section and moving it inside the ν integral.