

A Toy Model for Inverse Covariance Weighting

James Aguirre, Carina Cheng, and the PAPER Collaboration

In this memo, we focus on the optimal quadratic estimator, defined as:

$$\hat{P}^\alpha = \sum_{\beta} (\mathbf{F}^{-1})^{\alpha\beta} (\hat{q}^\beta - \hat{b}^\beta) \quad (1)$$

where \mathbf{F} is the Fisher matrix (which determines errors on the power spectrum estimate)

$$F^{\alpha\beta} \equiv \frac{1}{2} \text{tr} (\mathbf{C}^{-1} \mathbf{Q}^\alpha \mathbf{C}^{-1} \mathbf{Q}^\beta), \quad (2)$$

\hat{q} is the un-normalized power spectrum estimate

$$\hat{q}^\alpha = \frac{1}{2} \mathbf{x}^\dagger \mathbf{C}^{-1} \mathbf{Q}^\alpha \mathbf{C}^{-1} \mathbf{x}, \quad (3)$$

and \hat{b} is the additive bias

$$\hat{b}^\alpha = \frac{1}{2} \text{tr} (\mathbf{U} \mathbf{C}^{-1} \mathbf{Q}^\alpha \mathbf{C}^{-1}). \quad (4)$$

We will mathematically illustrate its role in estimating the power spectrum of EoR. While there exists detailed literature about quadratic estimators in general (e.g., Liu & Tegmark 2011; Trott et al. 2012; Liu et al. 2014; Dillon et al. 2014), here we focus on two simple cases in order to outline one situation where the estimator successfully suppresses contamination and one where it does not. By describing these two cases, we hope to clarify and motivate the desire to use OQE while also understanding its limitations.

We specifically choose toy models where the data covariance is diagonal, as indeed we expect the EoR signal to be. We assume we have N data points Δ_i which are the sum of a desired signal σ_i and an undesired contaminant v_i

$$\Delta_i = \sigma_i + v_i \quad (5)$$

with

$$\langle \sigma_i \rangle = 0; \langle \sigma_i^2 \rangle = s; \text{ and } \langle \sigma \sigma^T \rangle = s \mathbf{I}_{N \times N} \equiv \mathbf{S}, \quad (6)$$

where we wish to estimate s . The contaminant in this first case has a similar structure (as the EoR) for its covariance, and is assumed uncorrelated with the signal

$$\langle v_i \rangle = 0; \langle v_i^2 \rangle = u; \langle \mathbf{v} \mathbf{v}^T \rangle = u \mathbf{I}_{N \times N} \equiv \mathbf{U}; \text{ and } \langle \sigma_i v_j \rangle = 0. \quad (7)$$

With the covariance matrix given by $\mathbf{C} = \mathbf{S} + \mathbf{U}$, the estimator for s using only the quadratic part of Equation (1) is

$$\hat{s} = \frac{\Delta^T \Delta}{N} \quad (8)$$

and its expectation is

$$\langle \hat{s} \rangle = s + u. \quad (9)$$

Thus, when the covariance structure of the contaminant is identical to the signal ($\frac{\partial \mathbf{S}}{\partial s} = \frac{\partial \mathbf{U}}{\partial u} = \frac{\partial \mathbf{C}}{\partial s}$), the information available to the quadratic portion of the estimator to distinguish between the two is degenerate, and knowledge only of \mathbf{C} and $\frac{\partial \mathbf{C}}{\partial s}$ is inadequate. In order to obtain an unbiased estimate of s , one must also use knowledge of \mathbf{U} . Indeed, computing the linear bias from Equation (4), one finds $b = u$.

Now consider a case, chosen to be very similar to the toy model in ??, in which the data again have an additive contaminant, now given by

$$\Delta_i = \sigma_i + v m_i \quad (10)$$

where the properties of σ_i are as before, but now v is a random variable and m_i is a fixed function of i with

$$\langle v \rangle = 0; \quad \langle v^2 \rangle = u; \quad \langle \mathbf{v} \mathbf{v}^T \rangle = u \mathbf{m} \mathbf{m}^T \equiv \mathbf{U}; \quad \mathbf{m}^T \mathbf{m} = 1; \quad \text{and} \quad \langle \sigma_i v \rangle = 0. \quad (11)$$

Here \mathbf{m} represents a mode which is correlated across many data points (i.e., we are assuming \mathbf{U} need *not* be diagonal), with amplitude given by v . The normalization of \mathbf{m} is a matter of convention, and can be absorbed in the variance u ; the choice above will be convenient for understanding the limiting case $u \gg s$.

We can calculate the quadratic portion of the estimator explicitly by using the Sherman-Morrison identity to invert the covariance matrix. Defining

$$\xi \equiv \frac{u/s}{1 + u/s}, \quad (12)$$

we have

$$\mathbf{C}^{-1} = \frac{1}{s} (\mathbf{I} - \xi \mathbf{m} \mathbf{m}^T) \quad (13)$$

and

$$\hat{s} = \frac{\Delta^T (\mathbf{I} + (\xi^2 - 2\xi) \mathbf{m} \mathbf{m}^T) \Delta}{N + \xi^2 - 2\xi} \quad (14)$$

with expectation

$$\langle \hat{s} \rangle = s + \frac{1 - 2\xi + \xi^2}{N + -2\xi + \xi^2} u. \quad (15)$$

It is worth observing immediately that there is no multiplicative bias on s , and that the additive bias is strictly $< u/N$.

An instructive limit is $u \gg s$, $\xi \rightarrow 1$, in which case the virtue of weighting by \mathbf{C}^{-1} becomes clearer, as it becomes

$$\mathbf{C}^{-1} = \frac{1}{s} (\mathbf{I} - \mathbf{m} \mathbf{m}^T) \quad (16)$$

where $\mathbf{I} - \mathbf{m}\mathbf{m}^T$ is the projection operator, projecting out \mathbf{m} from any vector it acts on, and further, the linear bias tends to 0 as $\xi \rightarrow 1$ (i.e., the projection is “perfect” and not “undone” by the Fisher matrix normalization).

This is the ideal case for the inverse covariance weighting performed in the PAPER analysis, where removal of contamination with a known covariance can be suppressed by a kind of projection of the offending modes. But even in this case, it is worth pointing out that the estimator still has a linear bias for finite u . We have also assumed that the contaminating mode \mathbf{m} is known perfectly; the next appendix takes up the case where the modes are estimated from the data.

REFERENCES

- Dillon, J. S., et al. 2014, Phys. Rev. D, 89, 023002
- Liu, A., Parsons, A. R., & Trott, C. M. 2014, Phys. Rev. D, 90, 023019
- Liu, A., & Tegmark, M. 2011, Phys. Rev. D, 83, 103006
- Trott, C. M., Wayth, R. B., & Tingay, S. J. 2012, ApJ, 757, 101