Forecasting performance of CMB experiments in the presence of complex foreground contaminations

Forecasting performance of CMB experiments in the presence of complex foreground contaminations.

Radek Stompor*,[1] Josquin Errard†,[2,3] and Davide Poletti‡[1]

[1] AstroParticule et Cosmologie, Univ Paris Diderot, CNRS/IN2P3,CEA/Irfu, Obs de Paris, Sorbonne Paris Cité, France
[2] Sorbonne Universités Institut Lagrange de Paris (ILP), 98 bis Boulevard Arago 75014 Paris, France
[3] LPNHE, CNRS-IN2P3 and Universités Paris 6 & 7, 4 place Jussieu F-75252 Paris, Cedex 05, France

(Dated: November 2, 2016)

We present a new, semi-analytic framework for estimating the level of residuals present in CMB maps derived from multi-frequency Cosmic Microwave Background (CMB) data and forecasting their impact on cosmological parameters. The data are assumed to contain non-negligible signals of astrophysical and/or Galactic origin, which we clean using parametric component separation technique. We account for discrepancies between the foreground model assumed during the separation procedure and the true one, allowing for differences in scaling laws and/or their spatial variations. Our estimates and their uncertainties include both systematic and statistical effects and are averaged over the instrumental noise and CMB signal realizations. The framework can be further extended to account self-consistently for existing uncertainties in the foreground models. We demonstrate and validate the framework on simple study cases which aim at estimating the tensor-to-scalar ratio, r. The proposed approach is computationally efficient permitting an investigation of hundreds of set-ups and foreground models on a single CPU.

I. INTRODUCTION.

Forecasting performance of current and future CMB experiments is a necessary step in conception, design and optimization of their hardware as well as operations. Ideally, a forecasting procedure should be both reliable and efficient permitting scrutiny of broad swaths of parameter space in order to quickly zoom on a limited subset of the most promising configurations. This subset should be small enough to facilitate their further, more detailed investigation, typically employing numerical simulations, which while permitting a higher level of realism and detail are significantly more time and resource consuming.

Reliable forecasting for high precision CMB experiments is difficult due to the presence of the non-CMB signals, which unavoidably contribute to the measurements registered by the CMB instruments. Indeed, the multi-frequency observations from the Planck and WMAP satellites indicate that foreground emissions originating from our Galaxy or extra-galactic sources represent a major contaminant, e.g., [1–3], which current and future CMB polarization experiments will have to deal explicitly with. Methods employed for this purpose will thus have to ensure precision matching sensitivity envisaged for these forthcoming efforts and set by very ambitious science goals, which the CMB community world-wide is preparing to address. These goals include a detection and a characterization of the B-mode signal over a broad range of angular scales with a special emphasis on its large angular scale part, which is thought to be generated by primordial gravity waves present in the early Universe. The key parameter in this latter case is the so-called tensor-to-scalar ratio, r, and for concreteness in the following we will couch our presentation as targeting constraints on this parameter. The approach we introduce is however fully general and generalization to other parameters is straightforward.

The standard CMB forecasting tools are ill-adapted to tackle cases with non-negligible foreground contributions. Their impact is therefore often either modelled or assessed by some simplified means either in respect to estimating the residuals or their impact on the detection of r, e.g., [3–10]. Alternately, the issue is investigated with help of numerical simulations, which are typically computationally heavy and thus only allow for a limited number of studied cases [11–13].

Against this background, Errard et al (2011) [14] has proposed a semi-analytic framework, which attempts to propagate strictly statistical uncertainties incurred as a result of a component separation procedure to the final estimate of r. Their component separation of choice is a maximum likelihood parametric component separation approach [15–17], which assumes a parametrization of the frequency scalings for each considered sky component. Though self-consistent this approach is only capable of dealing with the statistical uncertainties and therefore its conclusions are limited in their validity and the results should be interpreted with caution. Specifically, this approach requires that the parametrization assumed for the frequency scaling of the sky components is sufficiently flexible and general that the actual frequency scaling laws of the true sky signals is included as its special case. Nonetheless, the framework has proven to be helpful in enabling studies of numerous experimental set-ups in a uniform fashion [18, 19], providing useful insights and intuitions and informing multiple instrument designs.

In this work we develop a framework capable of accounting for differences between these two sky signal

*E-mail: radek@apc.univ-paris-diderot.fr
†E-mail: josquin.errard@lpnhe.in2p3.fr
‡E-mail: Davide.Poletti@apc.univ-paris7.fr

models. As in [14], we assume that the components are separated with help of the parametric component separation technique and we estimate by semi-analytic means both the bias and statistical uncertainty, which are both present whenever the two sky models do not match. This can be either due to differences in the frequency scaling laws for some of the components or their spatial variability. The framework also permits incorporating the uncertainty related to our ignorance of the foreground signals and/or shortcomings of our models.

The bias and statistical uncertainties are then propagated to the second step of the procedure, where their impact on r is calculated. The new approach is equivalent to that of Errard et al (2011) [14], if the sky model and the true sky are consistent, and in this sense it extends and completes this earlier work.

We present the formalism in Sect. II and demonstrate and validate it in Sect. III. We leave a thorough investigation of different experimental set-ups and foreground models for future work. For convenience, we define symbols most commonly used in this paper in Table I.

II. FRAMEWORK

A. Data model.

The outline of our approach is as follows. Our input data are assumed to consist of a set of multi-frequency maps. These are collated together in a single data vector, $\mathbf{d}$, and are assumed to be a linear combination of sky component amplitudes in corresponding sky pixels. Collecting these together in a single sky component vector, $\hat{\mathbf{s}}$, we can therefore write,

$$\mathbf{d} = \hat{\mathbf{d}} + \mathbf{n} \equiv \hat{\mathbf{A}}\,\hat{\mathbf{s}} \,+\, \mathbf{n}, \tag{1}$$

where $\hat{\mathbf{d}}$ denotes true sky (noiseless) signal, matrix $\hat{\mathbf{A}}$ stands for the true mixing matrix of our data, and $\mathbf{n}$ – noise. We therefore have for a specific pixel, p,

$$\mathbf{d}_p \;=\; \hat{\mathbf{d}}_p + \mathbf{n}_p \equiv \hat{\mathbf{A}}_p\,\hat{\mathbf{s}}_p \,+\, \mathbf{n}_p, \tag{2}$$

where $\mathbf{d}_p$ stands for a vector of sky signal amplitudes measured at all observed frequencies in pixel p. Similarly, for a channel, k,

$$\mathbf{d}^{(k)} \;=\; \hat{\mathbf{d}}^{(k)} + \mathbf{n}^{(k)} \equiv \hat{\mathbf{A}}^{(k)}\,\hat{\mathbf{s}}^{(k)} \,+\, \mathbf{n}^{(k)}, \tag{3}$$

where $\mathbf{d}^{(k)}$ is a single frequency map of the observed sky in a frequency band defined by k.

We assume hereafter that the actual mixing matrix, $\hat{\mathbf{A}}$, is not available to us and instead we have to rely on some model of it to represent the available data. We denote this assumed mixing matrix as $\mathbf{A}$ and the corresponding component vector as $\mathbf{s}$. Our assumed data model therefore states that,

$$\mathbf{d}_p \;=\; \mathbf{A}_p(\beta)\,\mathbf{s}_p \,+\, \mathbf{n}_p, \tag{4}$$

where β denotes parameters used to parametrize the mixing matrix in order to reduce the number of unknowns.

Hereafter, we allow for a pixel-dependence of the model mixing matrix, $\mathbf{A}$. This could be either due to allowing for different values of the same physical parameters in different subsets of pixels or due to adopting different physical laws and parametrizations in different pixels. In any case, hearafter, a parameter, which is allowed to have a different value in two different pixels is treated as two different parameters rather than as a single parameter, which is pixel-dependent. This perspective will be helpful in the following.

Note that, whenever sky-variability of the scaling parameters is considered later on, we will assign different parameters to subsets of all pixels and make an implicit assumption hereafter that these are composed of a rather large number of pixels, covering well-behaved, singly-connected, compact sky patches. Though, the formalism is applicable more generally, its implementation and the interpretation of its results are both aided by this assumption.

We emphasize that in the presented formalism no assumption is made about the true sky signals and mixing matrices. Though, these are obviously needed for any specific application of the formalism in order to define the statistical properties of the data and play a crucial role in determining the resulting forecasts.

B. Data likelihood.

Following our data model we can write the standard likelihood function for the data given our model, Eq. (4). This reads up to an irrelevant constant as,

$$\mathcal{S}_{map} = \sum_p (\mathbf{d}_p - \mathbf{A}_p(\beta)\,\mathbf{s}_p)^t\,\mathbf{N}_p^{-1}(\mathbf{d}_p - \mathbf{A}_p(\beta)\,\mathbf{s}_p), \tag{5}$$

where here and in the following we define $\mathcal{S}$ as equal to $-2\ln\mathcal{L}$ up to some constant. We note that in the above expression the noise covariance is explicitly assumed to be uncorrelated between pixels. This is clearly not always the case and may need to be taken into account (see, e.g., a relevant discussion in [13]). Mathematically, the formalism presented hereafter can be easily generalized to permit correlations between pixels, however, in actual implementation treating full pixel-pixel covariance matrices becomes quickly prohibitively expensive. As our goal here is to provide a quick, performance forecasting tool, we will thus neglect the potential presence of such correlations.

If the multifrequency maps do not conform with our assumed model, i.e., $\mathbf{A} \neq \hat{\mathbf{A}}$, this likelihood is obviously incorrect as there is no value of β for which $\mathbf{d}_p - \mathbf{A}_p(\beta)\mathbf{s}_p$ could be merely a noise. However, this is the likelihood we would have adopted for component separation in the absence of any other information about the true sky. If discrepancies between the assumed and true sky signal

TABLE I: Notations

symbols	$\mathbf{d}$	$\hat{\mathbf{s}}$	$\mathbf{n}$, $\mathbf{N}$	$\hat{\mathbf{d}}$	$\hat{\mathbf{A}}$	p	k	β	$\mathbf{A}$	$\mathbf{s}$	$\bar{\mathbf{m}}_p$	$\mathbf{r}_p$, $\mathbf{r}^{\rm cmb}(\beta)$, $C_\ell^{\rm res}$	$\hat{\mathbf{f}}_p$, $\mathbf{F}_{pk}$, $\mathcal{F}_\ell^{\rm fore}$
definition	set of observed multi-frequency maps	true sky noiseless signal	noise, covariance	true sky component amplitudes	true mixing matrix	sky pixel	frequency channel	spectral parameters	model mixing matrix	model sky component amplitudes	noiseless estimates of the components	noiseless residuals	foregrounds signal

models are present, we expect that relying on this likelihood will lead in general to both systematic and statistical uncertainties in the derived results and in particular to systematic and statistical foreground residuals in the separated CMB map. In the approach proposed here we aim at estimating both these residuals and evaluating their impact on a value of the tensor-to-scalar ratio parameter, r, derived from the separated CMB map.

The proposed procedure involves two main steps: component separation and parameter estimation, which we describe in detail in the following.

C. Parametric component separation.

1. *Spectral parameters*

Following [14, 17] we invoke the spectral likelihood, which we will use to determine the spectral parameters of the scaling relations,

$$\mathcal{S}_{spec}=-\sum_p \left(\mathbf{A}_p\mathbf{N}_p^{-1}\mathbf{d}_p\right)^t \left(\mathbf{A}_p^t\mathbf{N}_p^{-1}\mathbf{A}_p\right)^{-1}\mathbf{A}_p\mathbf{N}_p^{-1}\mathbf{d}_p. \quad (6)$$

This is a profile likelihood obtained by maximizing the map likelihood in Eq. (5) with respect to the sky signal, $\mathbf{s}$. As such it peaks at exactly the same values as the full likelihood. $\mathcal{S}_{spec}$ can be maximized case-by-case for any given data, what indeed is implicitly or explicitly done in the parametric component separation codes, e.g. [16], or, instead, first averaged over the statistical ensemble of plausible input data and then maximized over the spectral parameters to yield both their average estimate and error on them. In the simplest case the statistical ensemble can merely include realizations of the noise. However, more generally, instead of a single true model of the foreground signals we may prefer to consider a family of models defined by their stochastic properties. This can be either due to our imperfect understanding of the foreground physics or due to the actual complexity of the foreground, which may be easier to sum up by statistical means [20, 21]. The formalism presented here lends itself straightforwardly to this kind of extensions. Nonetheless, we leave their exploitation to future work and in this paper we focus on the spectral likelihood averaged over a statistical ensemble of the noise realizations, which is then given by,

$$\langle\mathcal{S}_{spec}\rangle = -{\rm tr}\sum_p\left\{(\mathbf{N}_p^{-1}-\mathbf{P}_p)\Big(\hat{\mathbf{d}}_p\hat{\mathbf{d}}_p^t+\mathbf{N}_p\Big)\right\}. \quad (7)$$

Here, the dependence on the spectral parameters is confined to the projection operator, $\mathbf{P}_p$,

$$\mathbf{P}_p \equiv \mathbf{N}_p^{-1}-\mathbf{N}_p^{-1}\mathbf{A}_p\left(\mathbf{A}_p^t\mathbf{N}_p^{-1}\mathbf{A}_p\right)^{-1}\mathbf{A}_p^t\mathbf{N}_p^{-1}. \quad (8)$$

This likelihood can be maximized very efficiently numerically, given that in most applications the number of unknown spectral parameters is rather limited and capitalizing on the analytical derivatives of the likelihood as derived in Appendix A. In the approach proposed here, these maximum likelihood values define the average values of the spectral parameters as could be derived from the actual data while the curvature of the likelihood computed at its peak quantifies the uncertainty expected due to the instrumental noise. Denoting the latter as $\mathbf{\Sigma}$, we have,

$$\left(\mathbf{\Sigma}^{-1}\right)_{\beta\beta'} \equiv \left\langle\frac{\partial^2\mathcal{S}}{\partial\beta\partial\beta'}\right\rangle. \quad (9)$$

We note that the proposed procedure can be always applied, whether the assumed and true sky models match or not. When the true mixing matrix, $\hat{\mathbf{A}}$, agrees with the assumed one, $\mathbf{A}(\beta)$, for some values of the parameters, β, then the estimated values agree with these and the estimator is unbiased. In this case, only statistical residual, related to the statistical scatter of the determined values of β due to the instrumental noise, is present in the cleaned CMB map, which, if properly accounted for, will merely increase the statistical uncertainty of subsequently estimated cosmological parameters, without biasing their values [14].

When the assumed and true mixing matrices do not perfectly coincide for any values of the parameters for some or all pixels, be it due the inconsistency of the scaling laws or their spatial variability or both, there is a systematic residual unavoidably present in the estimated CMB map, which in turn may lead to biases in the estimated values of the cosmological parameters. We note that in such cases there are in general no 'true' values of the spectral parameters and their estimated, 'effective' values depend not only on the assumed and true mixing matrices but also instrument characteristics such as observational frequency bands, etc.

The statistical error matrix, $\mathbf{\Sigma}$, can be computed semi-analytically using Eq. (A5), or numerically. Indeed, many maximization routines construct a numerical approximation to the curvature matrix as part of the procedure. We have found that in our test cases both approaches produced results in very good agreement. We note that in the absence of calibration errors and assuming that the spectral dependence of the CMB component is known completely, neither the estimated spectral

parameter values nor their uncertainties depend on the CMB signal present in the data. This observation is analogous to the one pointed out in [14] and is elaborated on in Appendix A.

2. Residuals

Given $\mathbf{A}_p$ computed for some value of β, we can express noiseless estimates of the components as [22],

$$\bar{\mathbf{m}}_p = (\mathbf{A}_p^t \mathbf{N}_p^{-1} \mathbf{A}_p)^{-1} \mathbf{A}_p^t \mathbf{N}_p^{-1} \hat{\mathbf{d}}_p \equiv \mathbf{W}_p(\beta) \hat{\mathbf{d}}_p. \quad (10)$$

The noiseless residuals in the estimates then read as,

$$\mathbf{r}_p = \bar{\mathbf{m}}_p - \hat{\mathbf{s}}_p = \mathbf{W}_p(\beta) \hat{\mathbf{d}}_p - \hat{\mathbf{s}}_p. \quad (11)$$

Hereafter, we assume that the CMB corresponds to the 0th element of any multicomponent vector and split the multi-component vectors and the mixing matrices into CMB and foreground parts as follows,

$$\hat{\mathbf{A}} \equiv [\hat{\mathbf{A}}^{\rm cmb}, \hat{\mathbf{A}}^{\rm fore}],\ \mathbf{A} \equiv [\mathbf{A}^{\rm cmb}, \mathbf{A}^{\rm fore}],\ \hat{\mathbf{s}}_p \equiv \begin{bmatrix} \hat{\mathbf{s}}^{\rm cmb} \\ \hat{\mathbf{s}}^{\rm fore} \end{bmatrix}. \quad (12)$$

We can now represent $\hat{\mathbf{d}}_p$ as,

$$\hat{\mathbf{d}}_p = \hat{\mathbf{A}}_p^{\rm cmb} \hat{\mathbf{s}}_p^{\rm cmb} + \hat{\mathbf{A}}_p^{\rm fore} \hat{\mathbf{s}}_p^{\rm fore} \equiv \hat{\mathbf{A}}_p^{\rm cmb} \hat{\mathbf{s}}_p^{\rm cmb} + \hat{\mathbf{f}}_p, \quad (13)$$

where $\hat{\mathbf{f}}$ denotes a true noise-free contribution of the foregrounds to all single-frequency maps. As we assume throughout that the CMB scaling is the same in the model and true sky, we have,

$$\left[\mathbf{W}_p \hat{\mathbf{A}}^{\rm cmb}\right]_{00} = 1. \quad (14)$$

This emphasizes the fact that all of the CMB signal will remain in the estimated CMB component, which however will be contaminated by contributions from the other, non-CMB signals.

Indeed, on rewriting Eq. (11) and specializing it for the CMB component residual only we have,

$$\mathbf{r}_p^{\rm cmb} = \sum_k \mathbf{W}_p^{0k}(\beta) \hat{\mathbf{f}}_p^{(k)} \equiv \sum_k \mathbf{W}_p^{0k}(\beta) \mathbf{F}_{pk} \quad (15)$$

which, as expected, is explicitly free of the CMB signal. Here, we have introduced a foreground matrix, $\mathbf{F}$, kth column of which defines the total foreground contribution to the kth frequency channel.

We can now perform a Taylor expansion of the residuals with respect to the scaling parameters but around their estimated, maximum likelihood values, $\bar{\beta}$, obtaining,

$$\begin{aligned} \mathbf{r}_p^{\rm cmb}(\beta) \simeq & \sum_k \mathbf{W}_p^{0k}(\bar{\beta}) \mathbf{F}_{pk} + \sum_{k,\beta} \delta\beta \left.\frac{\partial \mathbf{W}_p^{0k}}{\partial \beta}\right|_{\bar{\beta}} \mathbf{F}_{pk} \\ & + \sum_{k,\beta,\beta'} \delta\beta\delta\beta' \left.\frac{\partial^2 \mathbf{W}_p^{0k}}{\partial\beta\partial\beta'}\right|_{\bar{\beta}} \mathbf{F}_{pk}, \end{aligned} \quad (16)$$

where we need to go up to the second order to have a consistent, up to the second order, approximation of the data covariance matrix, $\mathbf{E}$, in Eq. (38). On introducing pixel-domain objects: a vector, $\mathbf{y}$, two dimensional, $\mathbf{Y}^{(1)}$, and three dimensional, $\mathbf{Y}^{(2)}$, arrays, defined as,

$$\begin{aligned} \mathbf{y}_p &\equiv \sum_k \mathbf{W}_p^{0k}(\bar{\beta})\, \mathbf{F}_{pk}, \\ \mathbf{Y}^{(1)}_{p\beta} &\equiv \sum_k \left.\frac{\partial \mathbf{W}_p^{0k}}{\partial \beta}\right|_{\bar{\beta}} \mathbf{F}_{pk}, \\ \mathbf{Y}^{(2)}_{p\beta\beta'} &\equiv \sum_k \left.\frac{\partial^2 \mathbf{W}_p^{0k}}{\partial\beta\partial\beta'}\right|_{\bar{\beta}} \mathbf{F}_{pk}, \end{aligned} \quad (17)$$

we can rewrite this last expression as,

$$\mathbf{r}^{\rm cmb}(\beta) \equiv \mathbf{y} + \sum_\beta \delta\beta\, \mathbf{Y}^{(1)}_\beta + \sum_{\beta,\beta'} \delta\beta\delta\beta'\, \mathbf{Y}^{(2)}_{\beta\beta'}, \quad (18)$$

where for shortness we use $\mathbf{Y}^{(1)}_\beta$ and $\mathbf{Y}^{(2)}_{\beta\beta'}$ to denote pixel-domain vectors given by the elements of the arrays, $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ for which the spectral parameter indices – β and β, β' – are fixed.

We point out that $\delta\beta$ is explicitly pixel-independent. This is so thanks to the way we define the total spectral parameter set as discussed following Eq. (4), where every parameter appears as many times in the parameter set as many independent values it is allowed to take. Clearly, not all the parameters defined in this way will in general be relevant for all pixels. This is encoded in the multi-dimensional arrays, $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$, which will have all entries corresponding to such pixels set to zero.

Eq. (18) is a generalization of Eq. (10) of Errard et al. (2011) and Eq. (24) of Stivoli et al. (2010) [22]. The generalization concerns two aspects:

1. first, it includes the bias in the estimated component maps due to the residual foregrounds, which does not disappear when averaged over the statistical ensemble of noise and foreground realizations. This is given by $\mathbf{y}$.

2. second, this equation has been derived in a way, which did not invoke any assumptions about the pixel-dependence of the true sky mixing matrix, $\hat{\mathbf{A}}_p$, which can be therefore arbitrary.

Though this expression is derived in the pixel domain we can rewrite it in the harmonic domain owing to the fact that $\delta\beta$ are pixel independent,

$$\tilde{\mathbf{r}}^{\rm cmb}(\beta) \equiv \tilde{\mathbf{y}} + \sum_\beta \delta\beta\, \tilde{\mathbf{Y}}^{(1)}_\beta + \sum_{\beta,\beta'} \delta\beta\delta\beta'\, \tilde{\mathbf{Y}}^{(2)}_{\beta\beta'}, \quad (19)$$

where we use a tilde to denote vectors of harmonic multipoles. For definiteness they are arranged in such a way that the multipoles with the same ℓ are ordered consecutively with m increasing from $-\ell$ to ℓ and are followed

by the modes with $\ell' = \ell + 1$. So the relations between the multipole numbers, (ℓ, m), of a multipole and its position, j, in the harmonic vector are as follows,

$$\begin{aligned} j &= \ell^2 + \ell + m, \\ \ell &= \text{round}[(-1 + \sqrt{1+4j})/2], \\ m &= j - \ell\,(\ell+2)\,, \end{aligned} \tag{20}$$

where j goes from 0 to $(\ell_{max}+1)^2 - 1$ and the function *round* rounds a real number to the *closest* integer.

We note that going from the pixel domain vectors, $\mathbf{y}, \mathbf{Y}^{(1)}_{\beta}, \mathbf{Y}^{(2)}_{\beta\beta'}$, to their harmonic domain counterparts, $\tilde{\mathbf{y}}, \tilde{\mathbf{Y}}^{(1)}_{\beta}, \tilde{\mathbf{Y}}^{(2)}_{\beta\beta'}$, is only straightforward, if the pixel-domain vectors corresponds to the full-sky maps. This obviously is rarely the case in practice, as even for the full sky observations, the parts most affected by the foregrounds, i.e., Galactic plane and point sources, need to be typically masked out and are not used for the component separation. In the spirit of the Fisher approaches we however ignore this difficulty hereafter, assuming that this is possible and merely comes at the cost of the increased statistical uncertainty due to a fewer number of available modes and some rough cut-off scale at low-ℓ. In practice, as mentioned earlier, see the discussion after Eq. (4), this implies that the adopted mixing matrix instead of being permitted to change freely from pixel-to-pixel is taken to be the same for sufficiently large and regular sky patches.

For future convenience, we can write Eq. (19) for each harmonic mode, j $(= \ell^2 + \ell + m)$, Eqs. (20), as,

$$\tilde{\mathbf{r}}^{\rm cmb}_j(\beta) \equiv \tilde{\mathbf{y}}_j + \tilde{\mathbf{Y}}^{(1)}_j \boldsymbol{\delta} + \boldsymbol{\delta}^t \tilde{\mathbf{Y}}^{(2)}_j \boldsymbol{\delta}, \tag{21}$$

where $\tilde{\mathbf{Y}}^{(1)}_j$ and $\tilde{\mathbf{Y}}^{(2)}_j$ stand for a vector and a matrix respectively made of elements of $\tilde{\mathbf{Y}}^{(1)}_{j\beta}$ and $\tilde{\mathbf{Y}}^{(2)}_{j\beta\beta'}$ for the given j and $\boldsymbol{\delta}$ is a vector of uncertainties on spectral parameters, β, around the estimated values of the parameters. We therefore have

$$\langle \boldsymbol{\delta}\boldsymbol{\delta}^t \rangle = \boldsymbol{\Sigma} \quad \text{and} \quad \langle \boldsymbol{\delta} \rangle = 0. \tag{22}$$

Using these two last equations we can now rewrite an expression for the typical level of the residuals in the power spectrum domain. Indeed, we have, (see Appendix B for details),

$$\begin{aligned} C^{\rm res}_\ell \;\simeq\; & \otimes_\ell(\tilde{\mathbf{y}},\tilde{\mathbf{y}}) + \otimes_\ell(\tilde{\mathbf{y}},\tilde{\mathbf{z}}) + \otimes_\ell(\tilde{\mathbf{z}},\tilde{\mathbf{y}}) \\ & + \text{tr}\Big[\boldsymbol{\Sigma}\, \otimes_\ell(\tilde{\mathbf{Y}}^{(1)}, \tilde{\mathbf{Y}}^{(1)})\Big], \end{aligned} \tag{23}$$

where $\tilde{\mathbf{z}}$ is defined as,

$$\tilde{\mathbf{z}}_j \equiv \text{tr}\left[\tilde{\mathbf{Y}}^{(2)}_j \boldsymbol{\Sigma}\right], \tag{24}$$

and can be computed as the harmonic representation of the pixel domain object, $\mathbf{z}$, defined as,

$$\begin{aligned} \mathbf{z}_p &= \text{tr}\left[\mathbf{Y}^{(2)}_p \boldsymbol{\Sigma}\right] = \sum_{\beta,\beta'} \sum_k \left.\frac{\partial^2 \mathbf{W}^{0k}_p}{\partial\beta\partial\beta'}\right|_{\bar\beta} \mathbf{F}_{pk}\, \boldsymbol{\Sigma}_{\beta'\beta} \\ &= \sum_k \mathbf{F}_{pk} \left[\sum_{\beta,\beta'} \left(\left.\frac{\partial^2 \mathbf{W}^{0k}_p}{\partial\beta\partial\beta'}\right|_{\bar\beta} \boldsymbol{\Sigma}_{\beta'\beta} \right)\right]. \end{aligned} \tag{25}$$

In addition, we have also introduced symbol $\otimes$ to denote a power spectrum of two sets of harmonic coefficients provided as input parameters, i.e.,

$$\otimes_\ell(\tilde{\mathbf{X}}, \tilde{\mathbf{Z}}) \equiv \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \tilde{\mathbf{X}}^\dagger_j \tilde{\mathbf{Z}}_j, \quad j = \ell^2 + \ell + m. \tag{26}$$

We note that whenever, $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Z}}$ are multi-dimensional arrays of spectral coefficients indexed by j and the outcome of the operations is a matrix containing ℓth multipole of all the (cross)spectra of all these coefficients.

Eq. (23) permits computation of the power spectra of the typical residual in the presence of: (1) spatial variability of the spectral indices, in both the model and the actual sky, (2) discrepancies in the spatial and frequency behavior between the two, and (3) inhomogeneity of the measurement noise, which is however assumed to be uncorrelated between pixels. The inputs required for this are the relevant auto- and cross- spectra of $\tilde{\mathbf{y}}, \tilde{\mathbf{Y}}^{(1)}$, and $\tilde{\mathbf{z}}$. These spectra, in general, do not have any simple physical interpretation as they conflate all the different pixel-dependent effects together. However, if the assumed sky model is pixel-independent and the noise is white, all the spectra can be directly related to the spectra of the combined foreground signals as measured in different frequency bands, Appendix D, i.e.,

$$\mathcal{F}^{\rm fore}_\ell \equiv \otimes_\ell(\tilde{\mathbf{F}}, \tilde{\mathbf{F}}), \tag{27}$$

where $\tilde{\mathbf{F}}$ is a harmonic representation of $\mathbf{F}$ defined in Eq. (13) and therefore it is a matrix of as many columns as the assumed frequency bands with each column representing a combined foreground signal in each frequency band. Consequently, $\mathcal{F}^{\rm fore}_\ell$ is then a matrix containing multipole ℓ of all auto- and cross-spectra between foreground signals in all the frequency bands. Furthermore, if the true sky frequency scaling is also pixel independent, then the latter spectra can be related straightforwardly to those of the sky components at some fiducial frequency, as it was the case in [18].

We can also compute a dispersion around the average spectrum of the residual, Appendix B, which can be approximated as follows,

$$\begin{aligned} \text{Var}(C^{\rm res}_\ell) \;\simeq\; & 2 \left[\text{tr}\left(\boldsymbol{\Sigma}\, \otimes_\ell(\tilde{\mathbf{Y}}^{(1)}, \tilde{\mathbf{Y}}^{(1)})\right)\right]^2 \\ & +\; 2\, \otimes_\ell(\tilde{\mathbf{y}}, \tilde{\mathbf{Y}}^{(1)})\, \boldsymbol{\Sigma}\, \otimes_\ell(\tilde{\mathbf{Y}}^{(1)}, \tilde{\mathbf{y}}). \end{aligned} \tag{28}$$

3. *Noise*

The instrumental noise present in the cleaned CMB map, $\mathbf{n}^{\rm cmb}$, is given by,

$$\mathbf{n}^{\rm cmb}_p \;=\; \sum_k \mathbf{W}^{0k}_p(\bar\beta) \mathbf{n}_{pk}, \tag{29}$$

where $\mathbf{n}_{pk}$ denotes the noise of the kth frequency map. Given that we assume that the noise of each single frequency map is Gaussian, uncorrelated between pixels and

its pixel covariance is given by $\mathbf{N}_p$, the variance of the CMB map noise can be expressed as,

$$\sigma^2_{\mathrm{cmb,\,p}} \equiv \left[\left(\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}\right)^{-1}\right]_{\mathrm{CMB\ x\ CMB}} = \sum_{k,k'} \mathbf{W}_p^{0k}(\bar{\beta})\mathbf{W}_p^{0k'}(\bar{\beta})\mathbf{N}_{p,\,kk'}. \quad (30)$$

The CMB map noise will be inhomogeneous and its variance – pixel-dependent, whenever the noise of the single-frequency maps is inhomogeneous and/or the assumed scaling laws are pixel-dependent. In this latter case the coefficient of the matrix $\mathbf{W}$ will depend on the pixel, giving rise to the CMB map noise inhomogeneity even if the single-frequency maps noise is homogeneous. As mentioned earlier the formalism as developed until now is capable of handling the cases of both these kinds. Nevertheless, the noise inhomogeneity potentially leads to two problems. On the technical level, the noise which is inhomogeneous in the pixel domain results in correlations of the noise in the harmonic domain, which therefore usually can not be described in a compact, computationally-manageable manner. In particular, the noise power spectrum does not anymore provide a sufficient description of the noise properties in the harmonic domain. This is however in practice a necessary assumption for the forecasting approach as presented here. We discuss this issue in the next Section.

There is also another, more fundamental problem related to the noise inhomogeneities, which is pertinent to the robustness of the performance forecasts in the presence of the foregrounds and is not specific to this particular approach but is applicable to any forecasts based on the pixel-domain parametric approach and its applications. The constraints on the spectral parameters are tight whenever the foreground signal is high, or the noise level is low or both. The wildly inhomogeneous noise patterns are thus likely to lead to very different constraints depending on the fortuitous overlap of the densely observed sky areas with high foreground areas or lack thereof. Consequently, the predicted levels of both systematic and statistical residuals will not only very wildly but also will depend on the details of modelling, which can be difficult to control. As such they may not be particularly illuminating and useful in the performance forecasting, whatever is the specific way of obtaining those. Indeed, though such coincidental alignments may happen in the analysis of an actual data sets, it is likely that if the constraints on spectral parameters are found to be dominated by a handful of pixels, in which the foreground levels happen to be high and the instrumental noise – low, the best way forward could be to isolate these pixels in the analysis by for instance assigning to them a new set of parameters. Obviously, this effect is also present when the noise is homogeneous, however the noise fluctuations tend to amplify its role. In the former case, the effect is thought to be minimized by usually an implicit assumption that the foregrounds in the observed sky area are typical and that the area is large enough to include a range of typical foreground features.

The observational strategies of the CMB experiments, including those of the satellite missions [23, 24], however, commonly lead to inhomogeneous distributions of the observations over the covered sky area. Therefore, estimating the effects of the inhomogeneities on the forecasts is of clear importance. One way of proceeding here could be to derive the constraints on spectral parameters not tied up to specific foreground templates but averaging over plausible foreground morphologies. Though, as mentioned earlier this is something what can be readily incorporated in the proposed formalism, we leave an exploration of this aspect of the approach for future work.

Instead, for time being we make an implicit assumption that both the noise inhomogeneities and frequency scaling laws change only slowly with the position on the sky and across the observed sky, the noise power spectrum is expected to provide a description of the noise of the CMB map, which is sufficient for the estimation of the cosmological parameters based on this map. We calculate this spectrum as,

$$C_\ell^{\mathrm{noise}} = \frac{1}{n_{pix}} \sum_p \sigma^2_{p,\,\mathrm{cmb}}, \quad (31)$$

with the pixel-variance as given by Eq. (30). So the only effect due to the noise inhomogeneity or spatial dependence of the scaling laws, which is taken into account here, is an increased noise level.

In the case of the homogeneous noise and a single scaling law assumed for the entire observed patch, as used in the specific examples studied later on, this can be rewritten as,

$$C_\ell^{\mathrm{noise}} = \left[\left(\mathbf{A}^T\mathbf{N}_\ell^{-1}\mathbf{A}\right)^{-1}\right]_{\mathrm{CMB\ x\ CMB}} \quad (32)$$

where matrices N_ℓ are the harmonic space counterparts of N_p and describe the noise spectra of single frequency maps accounting for their resolution,

$$\mathbf{N}_\ell^{ij} \equiv (w_i)^{-1} \exp\left(\ell(\ell+1)\frac{\mathrm{FWHM_i}^2}{8\log 2}\right)\delta_i{}^j \quad (33)$$

with $(w_i)^{-1}$ the sensitivity of the frequency channel i in $(\mu\mathrm{K_{RJ}} - \mathrm{rad})^2$. We note that the beam effects appearing explicitly in this equation are included in there by hand as they are not fully consistent with the formalism presented earlier, which assumes no pixel-domain correlations. Indeed, the latter requirement implies that no beam-deconvolution procedure of any sort can be applied to the single-frequency maps prior to the component separation and instead the latter would have to be performed instead on the maps smoothed to the largest experimental beam. However, once the spectral parameter estimation is done, and thus the constraints on them are set using only the underpixelized data, both the noise and the residuals can be extrapolated to higher multipoles as

long as the number of frequency channels with a resolution high enough is sufficient to ensure that the matrix on the rhs of Eq. (32) is invertible.

D. Parameter estimation

As before, to estimate the parameters, we proceed as we would have done it if we did not know that our maps are potentially systematically biased. We therefore start off from the standard Gaussian likelihood, which accounts only for the presence of the noise, the CMB signal and the statistical foreground residual in the recovered CMB map, assuming that they all are Gaussian with the total covariance given by $\mathbf{C}$, i.e.,

$$\mathcal{S}^{par} \equiv -2\,\ln\mathcal{L}_{par} = \mathbf{a}^t\mathbf{C}^{-1}\mathbf{a} + \ln\det\mathbf{C}. \quad (34)$$

Here, $\mathbf{a}$ is a harmonic representation of the available CMB map. In our case this is the map obtained from the component separation procedure, and which therefore may include in addition to the CMB signal, the measurement noise and the statistical residual also systematic bias. This latter is however ignored in the *assumed* data covariance matrix, $\mathbf{C}$, which as in [14] includes only first three of these contributions: the CMB signal, the noise and the statistical foreground residual. The explicit form of $\mathbf{C}$ is given later on. The cosmological parameters we are after here enter only in the expression for the CMB covariance. The parameter likelihood averaged over the instrumental noise and CMB signal realizations is given by,

$$\langle\mathcal{S}^{par}\rangle = \mathrm{tr}\,\mathbf{C}^{-1}\mathbf{E} + \ln\det\mathbf{C}, \quad (35)$$

where $\mathbf{E} \equiv \langle\mathbf{a}\mathbf{a}^t\rangle$ is the correlation matrix of the data. The values of the cosmological parameters, which maximize Eq. (35), are those, for which the likelihood gradient vanishes. This reads, e.g., [25],

$$\langle\mathcal{S}^{par}_{,i}\rangle = \mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,i} - \mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{E}\right]. \quad (36)$$

Given that the harmonic coefficients of the CMB map, $\mathbf{a}$, can be represented as,

$$\begin{aligned}\mathbf{a}_j &= \mathbf{a}_j^{\rm cmb} + \mathbf{a}_j^{\rm noise} + \tilde{\mathbf{r}}_j^{\rm cmb} \\ &= \mathbf{a}_j^{\rm cmb} + \mathbf{a}_j^{\rm noise} + \tilde{\mathbf{y}}_j + \tilde{\mathbf{Y}}_j^{(1)}\boldsymbol{\delta} + \boldsymbol{\delta}^t\tilde{\mathbf{Y}}_j^{(2)}\boldsymbol{\delta}, \quad (37)\end{aligned}$$

we can write up to the second order in $\boldsymbol{\delta}$,

$$\begin{aligned}\mathbf{E}_{jj'} \equiv\, & \mathbf{D}_{jj'} + \tilde{\mathbf{y}}_j\tilde{\mathbf{y}}^*_{j'} + \tilde{\mathbf{y}}_j\,\mathrm{tr}\left[\tilde{\mathbf{Y}}^{(2)*}_{j'}\boldsymbol{\Sigma}\right] \\ &+ \mathrm{tr}\left[\tilde{\mathbf{Y}}^{(2)}_j\boldsymbol{\Sigma}\right]\tilde{\mathbf{y}}^*_{j'} + \tilde{\mathbf{Y}}^{(1)t}_j\boldsymbol{\Sigma}\tilde{\mathbf{Y}}^{(1)*}_{j'} \\ =\, & \mathbf{D}_{jj'} + \tilde{\mathbf{y}}_j\tilde{\mathbf{y}}^*_{j'} + \tilde{\mathbf{y}}_j\,\tilde{\mathbf{z}}^*_{j'} + \tilde{\mathbf{z}}_j\tilde{\mathbf{y}}^*_{j'} + \tilde{\mathbf{Y}}^{(1)}_j\,\boldsymbol{\Sigma}\,\tilde{\mathbf{Y}}^{(1),\,\dagger}_{j'}, \quad (38)\end{aligned}$$

where the cross-terms vanish given that $\langle a_{lm}\rangle = 0$ for both the noise and CMB, and $\langle\delta\beta\rangle = \langle a_{lm}\delta\beta\rangle = 0$ by definition. Here,

$$\begin{aligned}\mathbf{D}_{jj'} &\equiv \langle\mathbf{a}_j^{\rm cmb}\mathbf{a}_{j'}^{\rm cmb,\,\dagger}\rangle + \langle\mathbf{a}_j^{\rm noise}\mathbf{a}_{j'}^{\rm noise,\,\dagger}\rangle \\ &= C_\ell^{\rm cmb}\delta_{jj'} + C_\ell^{\rm noise}\delta_{jj'} \equiv C_\ell\delta_{jj'}, \quad (39)\end{aligned}$$

is the CMB plus noise only covariance, which is assumed hereafter to be diagonal in the harmonic space, corresponding therefore to the assumption of the stationary pixel-domain noise.

We can rewrite the expression for the true data covariance, Eq. (38), in the matrix form as,

$$\mathbf{E} = \mathbf{D} + \tilde{\mathbf{y}}\tilde{\mathbf{y}}^\dagger + \tilde{\mathbf{z}}\tilde{\mathbf{y}}^\dagger + \tilde{\mathbf{y}}\tilde{\mathbf{z}}^\dagger + \tilde{\mathbf{Y}}^{(1)}\boldsymbol{\Sigma}\,\tilde{\mathbf{Y}}^{(1)\dagger}. \quad (40)$$

In contrast, the assumed covariance, $\mathbf{C}$, will be like above but with the terms due to the bias omitted, i.e.,

$$\mathbf{C} = \mathbf{D} + \tilde{\mathbf{Y}}^{(1)}\boldsymbol{\Sigma}\,\tilde{\mathbf{Y}}^{(1)\dagger}. \quad (41)$$

This last expression resembles the one derived in [14], Eq. (B5), however it extends it by accounting correctly for the possible presence of multiple foreground components. We can now use Eqs. (41) and (38) to calculate the ensemble averages of derivatives of the likelihood given by Eq. (34). The full expressions are quite lengthy and are collected in Appendix C.

We note that though a computation of the explicit form of the covariance matrices, $\mathbf{C}$ and $\mathbf{E}$, Eqs. (41) and (40), respectively, requires knowledge of all the harmonic modes of the foreground components, for the calculation of the ensemble averaged likelihood and its derivatives, we need only various cross-spectra of pixel-domain objects defined by $\tilde{\mathbf{y}}$, $\tilde{\mathbf{z}}$ and columns of $\tilde{\mathbf{Y}}^{(1)}$. This is a general observation, which stems merely from the assumptions about the diagonality of the CMB signal and noise covariance matrices in the harmonic domain, and thus their stationarity in the pixel domain, and does not involve any specific assumptions about the foregrounds themselves. This can be intuitively understood, as whenever the CMB signal and noise are both stationary in the pixel domain the constraints on the cosmological parameters parametrizing the CMB spectrum can only depend on foreground properties averaged over the observed patch such as their power spectra, and not on their morphology or phase-dependent information. This is the case, whatever is the actual statistics of the foreground templates. If the noise is inhomogeneous and anisotropic in the pixel domain, it will give preference to some selected modes over others and the results of the parameter estimation will depend on both the power and morphology of the foregrounds, making the forecasting dependent on subtle details of the modelling, many of which are still poorly known at this time. This is similar to the case discussed earlier in the context of the component separation. Unlike in that latter case now the assumption of the pixel-domain noise stationarity not only makes our forecasts less detail dependent but it is in fact necessary in order to facilitate the analytic calculations, which in turn are essential for the numerical efficiency of the proposed approach.

Hereafter, we will thus employ the noise spectra as given in Eq. (31) accepting that some of the information is lost in this process. Our forecast will therefore be pessimistic in some sense but more reliable. The information is not lost only when the noise of the recovered CMB

map is homogeneous, which is equivalent to the case of the homogeneous noise in the frequency maps and global scalings laws, and where our assumptions are automatically fulfilled.

We note that more general noise power spectra than the white noise cases can be studied using this formalism, for instance, the spectra with excess power at low-ℓ end of the spectrum, devised to mimic the potential effects of time-domain noise correlations and/or time-domain filtering. As no noise correlations are included in our component separation step this will not be a fully consistent approach, however, as the pixel-stationary noise correlations are expected to have a bigger impact on the cosmological parameter estimation than the component separation step, such ad hoc adjustments can be expected to provide useful and meaningful insights.

This part of our algorithm now proceeds as follows: for the given foreground models, true and assumed, and the noise power spectrum, we maximize Eq. (35) using its first derivatives, Eq. (36), to find the maximum likelihood-like values of the cosmological parameters, which can be however biased by the presence of the residual foregrounds. We then use the ensemble average Hessian of the likelihood in Eq. (35), with the true data matrix set to $\mathbf{E}$, to assign uncertainties to these estimates.

The proposed procedure is well-defined in full- or nearly full- sky coverage cases. If only a limited sky area is available, the procedure can be adapted to produce some meaningful estimates. This involves the usual steps of introducing a low-ℓ cut-off corresponding to the largest mode, which can be still well-constrained by the cut-sky data, and of multiplying the derived, full-sky Hessian by the observed sky fraction, f_{sky}, to reflect the overall loss of the independent modes in the available data. Were we projecting out all the sky modes potentially contaminated by the foreground residuals, what would correspond to $\mathbf{\Sigma} \to \infty$, this later step would suffice. However, in our case $\mathbf{\Sigma}$ is finite and in fact hoped to be small so the spectral parameters are well-determined. Moreover, as it is estimated on the component separation step, it already incorporates the information about the observed sky fraction, as roughly $\mathbf{\Sigma} \propto 1/n_{pix} \propto 1/f_{sky}$. To account on that, as an input to the cosmological parameter estimation procedure, we use the rescaled matrix of spectral parameter errors, $\mathbf{\Sigma'} \equiv f_{sky}\mathbf{\Sigma}$. This rescaled error matrix roughly reflects the full sky errors and we use it in our algorithm to compute the bias and the Hessian for this case as described earlier. Once this is done we then rescale the Hessian by f_{sky}, so the statistical errors on the cosmological parameters are amplified by $\sqrt{f_{sky}}$.

E. Algorithm

The approach proposed here involves three main steps,

STEP I: estimation of the spectral parameters, used to parametrize the frequency scaling laws of the components and their uncertainty.

STEP II: estimation of the cross-spectra of the pixel-domain objects, $\mathbf{y}$, $\mathbf{Y}^{(1)}$, $\mathbf{Y}^{(2)}$, $\mathbf{z}$, characterizing the systematic and statistical residuals.

STEP III: estimation of the cosmological parameters and their uncertainty.

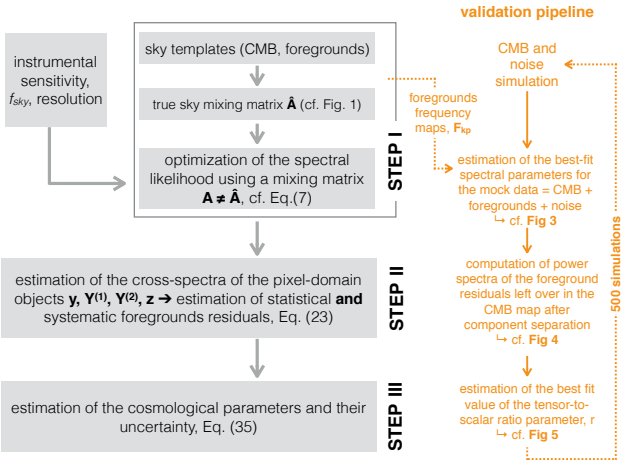

FIG. 1: A flow chart of the main steps of the proposed approach and of its validation pipeline implemented in this work.

As the products computed on each of these steps provide the inputs for the next ones, the necessary initial inputs consist of those required for the first step computations. In general these are,

1. the multi-frequency, noiseless true sky signal maps, $\hat{\mathbf{d}}$, split into its CMB, $\hat{\mathbf{s}}$, and foreground, $\hat{\mathbf{f}}$, parts;

2. the noise covariance matrices for the frequency channels, $\mathbf{N}$;

3. assumed, parametrizable scaling laws for all considered sky components, $\mathbf{A}$.

These are in principle sufficient to perform all the steps of the proposed approach. If available, the algorithm proceeds as follows,

STEP I: The best-fit spectra parameters are found by a direct maximization of the ensemble average likelihood in Eq. (7). This is implemented using a minimization routine from the PYTHON's SCIPY library implementing the truncated Newton constrained (TNC) solver. It capitalizes on the analytic derivatives of the likelihood with respect to the spectral parameters given by Eq. (A4). The statistical uncertainty is then computed using Eq. (A5).

STEP II: Given the best-fit values of the spectral parameters and their statistical uncertainty estimated on STEP I, we first estimate the pixel-domain objects,

TABLE II: Instruments specifications

frequency [GHz]	40	50	60	70	80	90	100	120	140	165	200	235	280	340	400
sensitivity [μK-arcmin]	42	26	20	15	12	19	12	10	7	7	5	6	19	10	19
FWHM [arcmin]	108	86	72	63	55	49	43	36	31	26	22	18	37	31	26

$\mathbf{y}$, $\mathbf{Y}^{(1)}$, $\mathbf{Y}^{(2)}$, $\mathbf{z}$, using Eqs. (17) and (25), and calculate their spherical harmonic decomposition and cross-spectra as needed;

STEP III: The forecasted values of the cosmological parameters are then computed by directly maximizing the likelihood in Eq. (35) with the first derivative computed with help of Eqs. (C13), (C18) and (C20). The likelihood itself is calculated using Eqs. (C8)-(C12), These computations use the cross-spectra computed on STEP II and is performed, as before, with help of a SCIPY minimization routine implementing the truncated Newton constrained (TNC) solver. However, this is now only performed after a rough grid-based search needed to ensure a reasonable starting point. The statistical uncertainty is then computed numerically as the curvature of the ensemble average cosmological parameter likelihood, which is an output of the routine.
Alternately, whenever the number of the sought-after cosmological parameters is very limited, typically $\lesssim 2$, the proposed formalism permits a full investigation of the likelihood function by a direct evaluation of Eq. (35) on a grid of the parameters and using the analytic results from Appendix C, Eqs. (C8)-(C12).

We note that in many cases of interest not all these calculations have to be actually performed and instead can be supplemented by some additional or alternative inputs. For instance, if neither the frequency scaling laws for the assumed sky model nor the pixel-domain noise depend on the pixel position on the sky, the sufficient information about the foregrounds can be provided by the foreground component-component covariance matrix and the component-component cross-spectra, Appendix D. In such a case, the cross-spectra of $\mathbf{y}$, $\mathbf{Y}^{(1)}$, $\mathbf{Y}^{(2)}$, $\mathbf{z}$ are directly related to those, simplifying and accelerating the calculations on required on STEP II. Also, the estimation of the spectral parameters, STEP I, can be then performed using only the component-component covariance matrix of the foregrounds as the input.

A simplified flowchart of the method is shown in Fig. 1.

III. VALIDATION AND DEMONSTRATION

We validate our approach using simulated, multi-frequency data sets of a putative CMB observation. Below we describe in turn: the assumed experimental set-up, the assumed foreground models, and the adopted validation methodology.

A. Set-up

1. Observation

We assume here a nearly full-sky observation, which have produced a set of multi-frequency maps of 70% of the entire sky suitable for cosmological component separation. The frequency bands, their assumed resolutions and sensitivites are all listed in Table II and have been selected following loosely the example of a contemporary satellite mission concept, [30]. Top-hat band passes are used throughout with their widths set to 30% of the central frequency for each band. We assume the duration of the observation to be 3 years, the noise to be white in map domain, with the homogeneous sky coverage, and no correlation between the maps at different frequencies included. In the following analyses we always include harmonic modes with ℓ ranging from 2 to 500, with the high-end cut-off set by the assumed resolutions. We emphasize that our goal here is to demonstrate and validate the proposed method and not to provide a performance evaluation of any specific instrument, what is left to the future work.

The noise spectrum, $C_\ell^{\rm noise}$, characterizing the noise of the final recovered CMB map, is computed using Eqs. (32) and (33) with the parameters listed in Table II.

2. Sky models

We adopt, as the true sky model, a model composed of two diffuse foreground components, one dust-like and the other synchrotron-like. We represent them as templates at 150GHz and scale them following the scalings laws as described below to all the frequencies of interest. For the templates we used the so-called COMMANDER dust and synchrotron maps, scaled to 150GHz using the Planck's fiducial scaling laws, as included and described in the Planck's latest polarized data release [31]. For simplicity, but also in agreement with recent findings of [26], hereafter we restrict ourselves to the case with no spatial variability of the scaling laws. We consider only 70% of the sky leaving out the Galactic plane as well as some other high foreground regions as defined by the sky mask provided by the Planck collaboration [32].

We model dust frequency scaling law as a sum of multiple grey-body terms each computed with a different set of grey-body parameters: the power law index, β_d, and temperature, T_d. Similarly, we model synchrotron frequency scaling as a sum of power laws with different

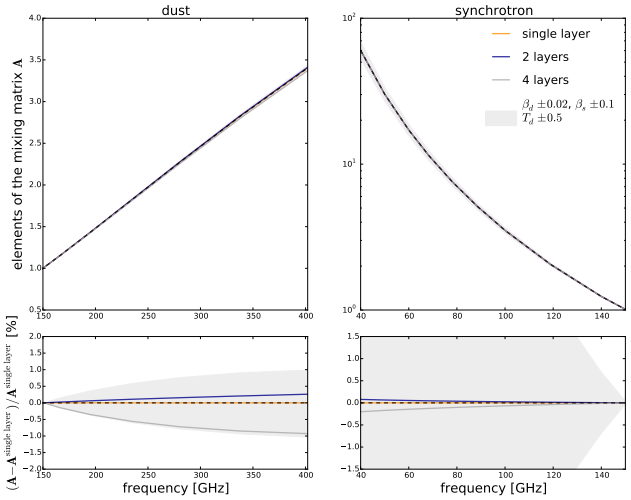

FIG. 2: Frequency scaling laws used in this work to demonstrate our method. The top panels show the scaling laws for dust, left, and synchrotron, right, for three different cases involving mixtures of 1, 2 and 4, grey-bodies, for dust, and power-laws, for synchrotron. The bottom panels show the same lines but relatively to the single term case. Shaded areas show a rough, 1-σ, uncertainty on these scaling laws consistent with the Planck data [27].

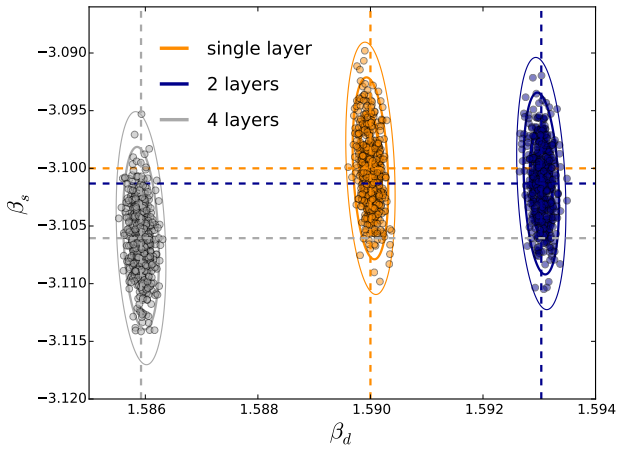

FIG. 3: Constraints on spectral parameters, power-law indices for dust, β_d, and synchrotron, β_s, forecasted for the assumed multi-frequency observation using the approach presented here. These are shown as solid ovals corresponding to $1, 2$, and 3- σ contours of the spectral parameter likelihood and computed for the three different foreground models as discussed in the text and shown in Fig. 2. Each model results are shown in different colors as indicated in the legend. The filled circles show the results of a direct maximization of the spectral likelihood performed for 500 independent realizations of the considered data set. The thin dashed lines show the position of the likelihood peaks as determined by the semi-analytic approach. These are virtually indistinguishable from the average values derived from simulations.

power law indices, β_s. Though these models are very simple, they allow us to investigate a range of different cases from the simple scaling law models, involving only a single term, to progressively more complex ones based on two and more terms. Specifically, in the following we will focus on three cases of the true sky scaling laws: one involving only one term, with the parameters set to be $T_d = 19.6$K, $\beta_d = 1.59$, and $\beta_s = -3.1$, and two more complex scalings based on a combination of 2 and 4 terms respectively for both dust and synchrotron. The effective scalings are shown in Fig. 2, where the top panels show the absolute scaling laws and the bottom ones – these laws relative to the single-term case. Clearly, the two-term scaling leads to the departures of up to $\sim 0.5\%$ in the dust scaling within the consider range of frequencies, while the 4-term cases admits deviations as big as $\sim 1\%$ for both the dust and synchrotron. The specific parameters used for the calculations of all the terms have been tuned to ensure that the adopted scaling laws are consistent with the Planck constraints [27], but also to allow us to validate and demonstrate our method in qualitatively different regimes. We discuss this in more detail in the next Section.

In contrast, while performing the component separation step on the simulated data sets, we always assume single-term-only scalings for both dust and synchrotron, fixing the dust temperature, T_d, to the actual value for the first term, i.e., $T_d = 19.6$K, and let the data determine values of two spectral indices, β_d and β_s.

B. Validation procedure

We validate our approach using simulations. These assume the foreground model described above. Each foreground template is scaled to each considered frequency band using one of the three scaling laws and integrated over the frequency band passes. These are coadded with 500 simulated CMB maps produced as a random realization of the Gaussian process with power spectra defined by the standard cosmological model with parameters set to the best fit Planck values. All the signal maps are downgraded to $N_{\rm side} = 32$. We thus obtain three sets of 500 simulated, single-frequency maps, each set implementing a different foreground scaling law. In addition, for each set we generate 500 independent realizations of the instrumental noise, which combined together with the signal maps create mock data sets. These noise realizations are drawn for each frequency band separately and modelled as a Gaussian process with variance as given in Table II and zero mean. There are no other systematic effects included in the simulations.

As part of the validation procedure, we analyze each of the simulations as we would the actual data, performing first the pixel-domain, maximum-likelihood, parametric component separation followed by the pixel-domain, maximum-likelihood cosmological parameter fitting. At

the conclusion of each of these two steps we compare the simulated results with the corresponding results obtained with the proposed method. Specific comparisons, performed on each step of the proposed algorithm, are as follows, cf. Fig. 1:

STEP I: we estimate the best-fit spectral parameters for the mock data by explicitly maximizing the spectral likelihood in Eq. (6). We do so for each of the three sets of the 500 simulations. We then compare the results with the expected distribution of the measured spectral parameters derived semi-analytically using Eqs. (7) and (9) and assuming a Gaussian approximation.

STEP II: to validate this step we compute power spectra of the foreground residuals left over in the CMB maps after the component separation step. We note that these are higher level objects, which are neither explicitly derived on this step of the processing nor needed for the subsequent stages of the procedure. However, they combine the same information as the direct products and have well-defined physical interpretation. They therefore provide a meaningful and intuitive comparison metric.
We implement this comparison as follows. For the simulated data, we first compute the cleaned CMB map using expression in Eq. (10) assuming the spectral parameters as derived on STEP I. The map domain residual is then derived by subtracting from this map the true CMB map used to simulate the input data. We then calculate the power spectrum of the residuals and compare it with the semi-analytic results derived via Eqs. (23) and (28).

STEP III: we estimate the best fit value of the tensor-to-scalar ratio parameter, r, by explicitly maximizing full cosmological parameter likelihood, Eq. (34), we bin the results and compare the histogram with the Gaussian distribution with an average and dispersion derived using the proposed approach, Section II D. In addition, we explicitly compute the averaged likelihood as a function of r, using Eq. (35). In both these latter cases the computations are performed directly in the harmonic domain using the harmonic space representation of the cleaned CMB maps obtained on STEP II. This is done with help of a standard spherical harmonic transform and thus neglects the effects due to cut sky. This is expected to lead to some power loss at the low-ℓ end and thus may affect the level of bias in the estimated values of r. This should be taken care of in the actual forecasting process however is irrelevant for the formalism demonstration purposes.

Due to computational-time limitation we perform the end-to-end analysis of the actual simulation on underpixelized maps with HEALPIX $n_{side} = 32$. This is what restricts the analysis presented hereafter to $\ell_{max} = 64$ as the semi-analytic approach can be easily applied for much higher cut-offs.

C. Results

We present here results obtained for each of the three steps defined earlier. They are visualized in Figs. 3, 4, and 5.

In Fig. 3 we show the spectral laws parameters and their $1, 2, 3-\sigma$ confidence levels, shown as contours, obtained as a Gaussian approximation to our spectral likelihood, Eq. (7). This likelihood is averaged over possible noise realizations. The thick points show the results of the simulations and therefore each point shows the values of the spectral parameters derived by a direct maximization of the spectral likelihood, Eq. (6). We note that each set of the mock multifrequency data contains different realization of the noise and the CMB signal, however as discussed earlier on, the latter does not impact the spectral parameter estimation, see Eq. (15), and the scatter of the spectral parameters values from the simulations seen in the figure is driven only by the noise. Overall, we see very good agreement between the semi-analytic contours and the overall distribution of the results derived from the simulations, which as expected cover the areas delineated by the contours aggregating around the expected peaks of the likelihood. More quantitively, the average spectral parameters derived from the simulations are the dashed lines and show a good agreement with the semi-analytic values.

We note that the position of the peak of the likelihood is different for each of the cases. Indeed, only in the case of the single-term model the true and assumed sky agree and the recovered values of the spectral parameters agree with those used in the sky simulations. For the multi-term scaling laws, the assumed model does not provide an accurate description of the true scaling laws and the derived values of the spectral parameters do not carry any more any physical meaning, but rather they are some effective values that lead to the scaling laws matching the true ones most accurately.

In Fig. 4 we show the comparison performed after STEP II. The grey lines show power spectra of the foreground residuals computed for each of the 500 realizations of the noise and the thick dashed lines show their average. The solid orange lines depict the average residual spectra and the orange shaded areas show 2σ uncertainty computed with help of the proposed method. We see a very good overall agreement for all three choices of the scaling laws.

The residuals shown in the left panel are merely due to the statistical scatter in the estimates of the spectral parameters and this case corresponds to those studied in [14]. Though in this case no bias is expected on the map level, this is not so on the power spectrum level. Indeed, the average power spectrum of the residuals does not vanish as it is indeed shown in the figure by both semi-analytic and simulated results.

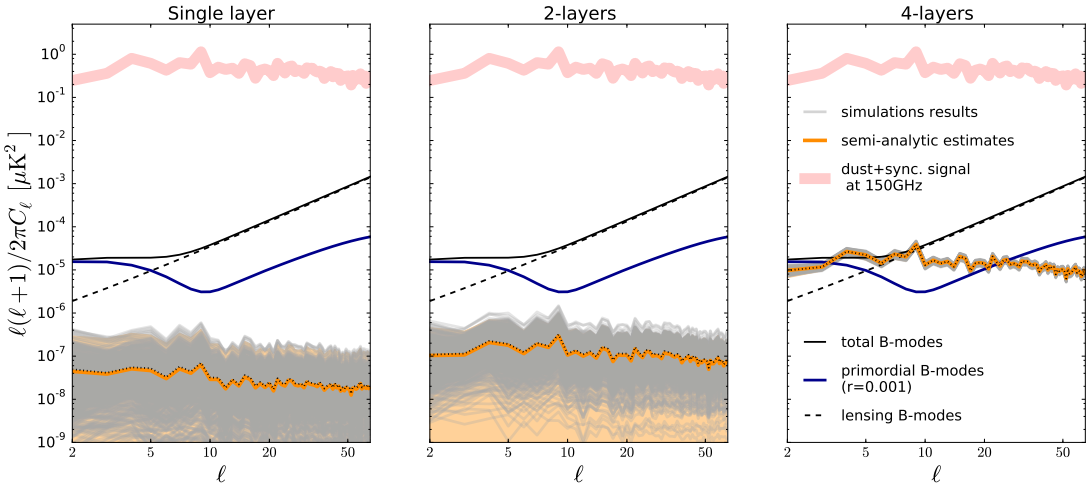

FIG. 4: Foreground residuals present in the cleaned CMB maps after the parametric component separation. The panels corresponds to the three different scaling laws considered in this work. The results shown in orange are obtained using the semi-analytic method proposed in this work with the solid orange lines showing the ensemble-averaged residuals and the shaded orange areas depicting the 2σ scatter. The grey lines show the residuals computed case-by-case for each of the simulations and the thick, dashed lines show their averages. The theoretical B-mode spectra showing the primordial ($r = 10^{-3}$), dark blue line, lensing, dashed black, and total, thick black line, are also shown for reference. The three models represent cases with dominant statistical residuals, left panel, dominant systematic residuals, right panel, and comparable statistical and systematic residuals, middle panel. For comparison the red-shaded band shows the total, dust plus synchrotron, foreground signal at 150GHz. Only the low-ℓ residuals are shown here.

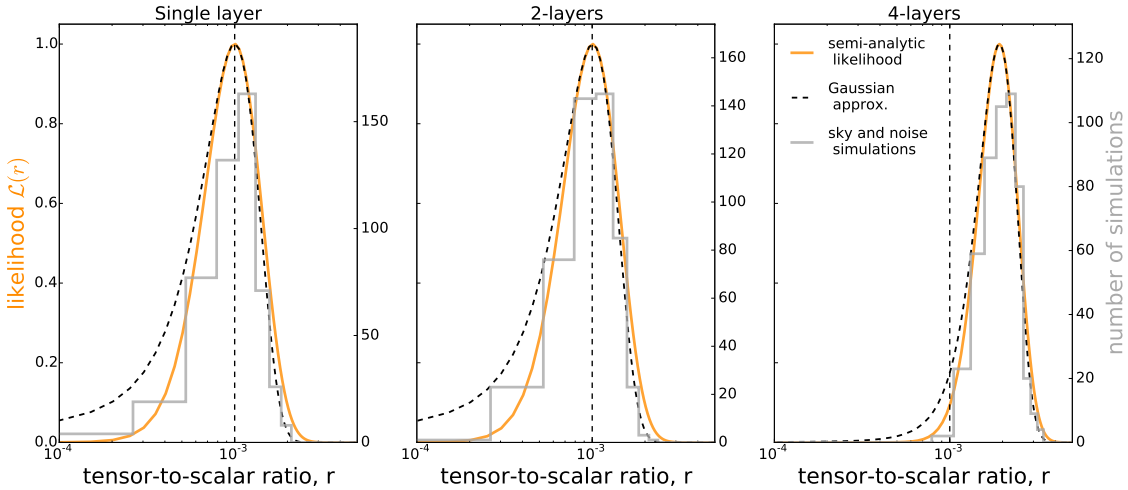

FIG. 5: Forecasted constraints on the tensor-to-scalar ratio parameter, r, derived from foreground-cleaned CMB maps derived for the multi-frequency observation studied here. The orange solid lines show the likelihoods on r averaged over the statistical ensemble of noise and CMB realizations estimated using Eq. (35). The black, thick, dashed lines show the results of the Gaussian approximation with the average and variance computed semi-analytically as described in this work. The histograms show results of an end-to-end analysis of the putative, simulated data sets involving random realizations of the instrumental noise and CMB signal and the foreground contributions as used in the semi-analytic approach. The results in the left panel show no bias in the estimated value of r. The bias in the middle panel though formally non-zero is negligible as compared to the statistical error. In contrast, the bias seen in the rightmost panel is significant as compared to the true value of r and the estimated statistical uncertainty. These results demonstrate an excellent agreement between our semi-analytic approach and the full computation with the Gaussian approximation being, however, more permissive as far as low values of r are concerned.

In the middle and right panels, given the scaling laws mismatch the residuals are due to both the systematic and the statistical errors. In the middle panel, both these errors are comparable and non-negligible. While in the right panel, the systematic contribution is by far dominant. These results show that the systematic residuals due to the mismatch in the scaling laws can quickly dominate over the statistical residuals even for rather minor levels of the mismatch (0.5% and 1% in the cases shown in the middle and right panels) at least as long as the freedom of introducing more spectral parameters is not capitalized on.

The residuals may potentially affect the values of the cosmological parameters determined from the CMB maps cleaned with the parametric component separation approach. This is not the case for the statistical residuals as in the case shown in the left panel of Fig. 4. This is because our parameter likelihood, Eq. (34), is written in the map-domain, where the foreground residual averaged over ensemble of the noise realization vanishes, and because it accounts for the extra statistical uncertainty. In this case we thus expect only the extra uncertainty but no bias for the estimated cosmological parameters. This would not have been the case, were our likelihood written in the power spectrum domain. The bias of the cosmological parameters is expected in our approach once the systematic foreground residuals are present. In such cases, the cosmological parameter estimates will be affected by both biases and extra uncertainty.

Fig. 5 demonstrates all these general considerations in the context of a determination of parameter r. In this figure, the smooth orange lines show the predictions obtained from our semi-analytic approach, while the histograms are obtained by performing simplified maximum likelihood parameter fitting applied to the CMB maps contaminated with the foreground residuals. Again we find a very good qualitative and quantitive agreement in all three cases. The peak value of the likelihoods and histograms shifts progressively away from the true value, taken here to be $r = 10^{-3}$, when the scaling law mismatch and therefore the systematic residual is getting bigger. The case shown in the left panel and thus affected only by the statistical residual does not lead to any bias in r. In the intermediate case, middle panel, the bias is marginal and negligible when compared to the statistical scatter, however, in the case shown in the right panel the bias is already statistically important. However, even in this case the bias on r is not as large as one may have expected from the level of the residuals in the power spectrum domain as seen in Fig. 4. This merely reflects the fact that power spectra of the foregrounds and CMB signals are sufficiently different that the effects of the former are minimized in the parameter fitting procedure. Last but not least, we note that the Gaussian approximation tends to underestimate the actual significance of the detection due to the long tail for values of r going to 0.

With regard to the methodology, in the studied cases we have found that the terms containing either vectors $\mathbf{z}$ or $\mathbf{Y}^{(\mathbf{2})}$, Eqs. (25) and (17), and which thus arise due to our inclusion of the second order terms in the expansion of the residuals with respect to the spectral parameter deviations, Eq. (18), tend to be subdominant and can probably be safely discarded, what would simplify the numerical implementation. On the other hand, neglecting the off-diagonal terms in the covariance matrices, $\mathbf{C}$ and $\mathbf{E}$, Eqs. (41) and (40), is more consequential as it can potentially cause mis-estimation of the bias in the estimated parameters by as much as 100%. At the same time, this will not however lead to a spurious bias if such is absent in the cleaned CMB map.

IV. CONCLUSIONS AND PROSPECTS

In this work we have proposed a semi-analytic approach suitable for realistic forecasting of constraints, which can be set on the cosmological parameters by multi-frequency CMB experiments in the presence of complex foreground contaminations. The derived constraints are averaged over the instrumental noise and CMB realizations and consist of the estimates of the most likely values of the parameters as well as of their dispersion.

The method assumes that the foregrounds are cleaned using a pixel-based, parametric, maximum likelihood component separation approach, however it does not require that the parametric model assumed for the separation process matches the true one for any set of parameter values. If the mismatch is indeed present, the estimated scaling laws will differ from the actual sky ones in a systematic way. This leads to foreground residuals, both systematic and statistical, which will be present in the cleaned CMB map. In our approach we first estimate both these residuals and subsequently incorporate them in the pixel-based cosmological parameter likelihood, which we use to set constraints on cosmological parameters. The constraints derived in this way therefore include both biases as well as statistical uncertainty. In this sense our method generalizes previous efforts of the similar kind [14, 18, 19]. We have validated the method in the case of pixel-independent scaling laws and white pixel-domain noise, however, the presented algebraic framework is flexible enough to allow for spatial variation of the foreground scaling for both the true and modelled signals as well as some other real life effects. Furthermore, we also note that the proposed formalism permits incorporating any uncertainties in the foreground modelling in the final forecasts. This could be a potentially very handy feature if broad families of the foreground models need to be investigated. We leave detailed studies of those cases for future work.

In the cases studied in this work, we have found that even a rather minor mismatch, say of $\sim 1\%$, between the true and assumed scaling laws over a broad range of frequencies can lead to substantial biases of the esti-

mated value of the tensor-to-ratio parameter, r, if its true value is as low as 10^{-3}. This emphasizes two things: (1) importance of accurate and realistic modelling of the underlying foreground signals in ensuring that the obtained forecasts are realistic; (2) importance of suitably chosen, parametric scaling models. In this work, for the demonstration purposes, we have adopted rather simple models in both these instances. In particular, we have employed a simple, two-parameter scaling model for the separation stage and thus have not explored all the constraining power of the considered observation, which allows for a significantly larger number of the spectral parameters. For these reasons the results shown here should not be seen as a fair evaluation of the performance of the assumed instrumental set-up but rather merely as indicative of more qualitative effects and dependences one may expect in such circumstances. Again we leave exhaustive explorations of this kind to future work.

Our approach, though clearly more involved and complex than that of [14, 18, 19], retains the speed and efficiency of these previous, simplified techniques, while permitting to attain a higher level of realism. Indeed, all the numerical computations scale linearly with the high-ℓ cut-off, ℓ_{max}, allowing the calculation to be conducted efficiently even for high-resolution experimental set-ups. Consequently, the method is very well-suited for optimizations of experimental set-ups and forecasting their performance, in particular whenever large parameter space of experimental characteristics needs to be considered. Equally importantly, this approach also allows for a direct exploration of a large number of viable foreground models, thus enabling investigations of robustness of the predictions with respect to details of the foreground modelling – a key feature given our present ignorance about the polarized foreground emissions in the microwave band and the impact of the assumed foreground models on the derived predictions. In all these aspects, the proposed approach is complementary to a more thorough but also more resource demanding, fully-fledged, end-to-end analysis of the realistic simulations.

Acknowledgments

We acknowledge use of the HEALPix [28] package, as well as the CMB4cast code [33]. The work of JE was performed within the Labex ILP (reference ANR-10-LABX-63) part of the Idex SUPER, and received financial state aid managed by the Agence Nationale de la Recherche, as part of the programme Investissements d'avenir under the reference ANR-11-IDEX-0004-02.

Appendix A: Spectral likelihood derivatives

We calculate here the first derivative of the likelihood averaged over the statistical ensemble of noise. Without losing the generality, for the derivation's sake, we assume only one pixel, and thus drop subscript p and take the noise covariance, $\mathbf{N}$, to be the identity. We then present the fully general expressions only at the very end.

We start from Eq, (6) and rewrite it as

$$\mathcal{S}_{spec} = \mathrm{tr}\left[(\mathbf{1}-\mathbf{P})\,\mathbf{d}\mathbf{d}^t\right]. \tag{A1}$$

Consequently all we need to do is to compute $\mathbf{P}_{,\beta}$ and $\mathbf{P}_{,\beta\beta'}$. On defining, $\mathbf{M} \equiv (\mathbf{A}^t\mathbf{A})^{-1}$, these can be written as,

$$\mathbf{P}_{,\beta} \;=\; -\mathbf{A}\mathbf{M}\mathbf{A}^t_{,\beta}\mathbf{P} + \mathrm{transpose}, \tag{A2}$$

and,

$$\begin{aligned}\mathbf{P}_{,\beta\beta'} \;=\; & -\mathbf{P}\mathbf{A}_{,\beta'}\mathbf{M}\mathbf{A}^t_{,\beta}\mathbf{P} + \mathbf{A}\mathbf{M}\mathbf{A}^t_{,\beta'}\mathbf{A}\mathbf{M}\mathbf{A}^t_{,\beta}\mathbf{P} - \mathbf{A}\mathbf{M}\mathbf{A}^t_{,\beta\beta'}\mathbf{P} \\ & + \mathbf{A}\mathbf{M}\mathbf{A}^t_{,\beta}\mathbf{A}\mathbf{M}\mathbf{A}^t_{,\beta'}\mathbf{P} + \mathbf{A}\mathbf{M}\mathbf{A}^t_{,\beta}\mathbf{P}\mathbf{A}_{,\beta'}\mathbf{M}\mathbf{A}^t + \mathrm{transpose}.\end{aligned} \tag{A3}$$

We can now combine all these terms to form the derivatives of the spectral likelihood, obtaining,

$$\left\langle \frac{\partial \mathcal{S}_{spec}}{\partial\beta} \right\rangle = \sum_p \mathrm{tr}\left[\mathbf{N}_p^{-1}\mathbf{A}_p(\mathbf{A}_p^t\mathbf{N}_p^{-1}\mathbf{A}_p)^{-1}\mathbf{A}^t_{p,\beta}\mathbf{P}_p\langle\mathbf{d}_p\mathbf{d}_p^t\rangle\right], \tag{A4}$$

and

$$\begin{aligned}\left\langle \frac{\partial^2 \mathcal{S}_{spec}}{\partial\beta\partial\beta'} \right\rangle \;=\; & \sum_p \mathrm{tr}\Big\{\Big[\mathbf{P}_p\mathbf{A}_{p,\beta'}(\mathbf{A}_p^t\mathbf{N}_p^{-1}\mathbf{A}_p)^{-1}\mathbf{A}^t_{p,\beta}\mathbf{P}_p + \mathbf{N}_p^{-1}\mathbf{A}_p(\mathbf{A}_p^t\mathbf{N}_p^{-1}\mathbf{A}_p)^{-1}\mathbf{A}^t_{p,\beta\beta'}\mathbf{P}_p \\ & - \mathbf{N}_p^{-1}\mathbf{A}_p(\mathbf{A}_p^t\mathbf{N}_p^{-1}\mathbf{A}_p)^{-1}\mathbf{A}^t_{p,\beta'}\mathbf{N}_p^{-1}\mathbf{A}_p(\mathbf{A}_p^t\mathbf{N}_p^{-1}\mathbf{A}_p)^{-1}\mathbf{A}^t_{p,\beta}\mathbf{P}_p \\ & - \mathbf{N}_p^{-1}\mathbf{A}_p(\mathbf{A}_p^t\mathbf{N}_p^{-1}\mathbf{A}_p)^{-1}\mathbf{A}^t_{p,\beta}\mathbf{N}_p^{-1}\mathbf{A}_p(\mathbf{A}_p^t\mathbf{N}_p^{-1}\mathbf{A}_p)^{-1}\mathbf{A}^t_{p,\beta'}\mathbf{P}_p \\ & - \mathbf{N}_p^{-1}\mathbf{A}_p(\mathbf{A}_p^t\mathbf{N}_p^{-1}\mathbf{A}_p)^{-1}\mathbf{A}^t_{p,\beta}\mathbf{P}_p\mathbf{A}_{p,\beta'}(\mathbf{A}_p^t\mathbf{N}_p^{-1}\mathbf{A}_p)^{-1}\mathbf{A}_p^t\mathbf{N}_p^{-1}\Big]\langle\mathbf{d}_p\mathbf{d}_p^t\rangle\Big\},\end{aligned} \tag{A5}$$

where we have used the fact that the trace of a product of a symmetric matrix and an arbitrary matrix is the same as that of the symmetric matrix and the transpose of the arbitrary matrix. We note that these equations agree with Eqs. (4) and (A9) of [14]. We also note that as in the case studied in this latter work neither derivative depends on the specific CMB sky signal included in the data, $\hat{\mathbf{d}}_p = \hat{\mathbf{A}}\hat{\mathbf{s}}$, as long as the CMB frequency scaling is assumed to be known. This can be seen on observing that the sky signal, $\hat{\mathbf{s}}$, in the expressions for the first and second derivative of the likelihood is processed either by operator,

$$\mathbf{A}_{p,\beta}\,\mathbf{A}_p^t\mathbf{N}_p^{-1}\mathbf{A}_p)^{-1}\mathbf{A}_p^t\mathbf{N}_p^{-1}\hat{\mathbf{A}}_p, \tag{A6}$$

or the projection operator, $\mathbf{P}_p$. However, by assumption we have, $\mathbf{A}_{i0} = \hat{\mathbf{A}}_{i0} = 1$, where subscript 0 denotes the column corresponding to the CMB and we have adopted the thermodynamical units, and therefore (see also Eq (14)),

$$\Big[(\mathbf{A}_p^t\mathbf{N}_p^{-1}\mathbf{A}_p)^{-1}\mathbf{A}_p^t\mathbf{N}_p^{-1}\hat{\mathbf{A}}_p\Big]_{i0} = \delta_{i0}. \tag{A7}$$

Moreover, given that, $[\mathbf{A}_{p,\beta}]_{0i} = 0$,

$$\Big[\mathbf{A}_{p,\beta}(\mathbf{A}_p^t\mathbf{N}_p^{-1}\mathbf{A}_p)^{-1}\mathbf{A}_p^t\mathbf{N}_p^{-1}\hat{\mathbf{A}}\Big]_{0i} = 0 \tag{A8}$$

and the operator removes all the CMB signal present in the input data vector, $\hat{\mathbf{s}}_p$. Similarly, the projection operator, $\mathbf{P_p}$, projects out the CMB signal in its entirety as,

$$\Big[\mathbf{P}_p\hat{\mathbf{A}}\Big]_{i0} = \big[\mathbf{N}_p^{-1}\big]_{ij}\,\Big[\hat{\mathbf{A}}_p\,-\,\mathbf{A}_p(\mathbf{A}_p^t\mathbf{N}_p^{-1}\mathbf{A}_p)^{-1}\mathbf{A}_p\mathbf{N}_p^{-1}\hat{\mathbf{A}}_p\Big]_{j0} \,=\, \big[\mathbf{N}_p^{-1}\big]_{ij}\Big(\hat{\mathbf{A}}_{j0} - \mathbf{A}_{j0}\Big) = 0. \tag{A9}$$

Consequently, the CMB signal affects neither the best-fit values of the spectral parameters, β, nor their uncertainties.

Appendix B: Residuals power spectrum and its variance.

We calculate the power spectrum of the residuals up to the second order in $\delta\beta$. From Eq. (21) we get,

$$\begin{aligned} C_l^{\rm res} \;&\simeq\; \frac{1}{2l+1}\sum_m \langle\tilde{\mathbf{r}}_j^{\rm cmb,\,\dagger}\,\tilde{\mathbf{r}}_j^{\rm cmb}\rangle = \frac{1}{2l+1}\sum_m\Big[\tilde{\mathbf{y}}_j^\dagger\tilde{\mathbf{y}}_j + {\rm tr}\Big[\boldsymbol{\Sigma}\,\tilde{\mathbf{Y}}_j^{(1)\,\dagger}\,\tilde{\mathbf{Y}}_j^{(1)}\Big] + \tilde{\mathbf{y}}_j^\dagger{\rm tr}\Big[\tilde{\mathbf{Y}}_j^{(2)}\boldsymbol{\Sigma}\Big] + {\rm tr}\Big[\tilde{\mathbf{Y}}_j^{(2)\,\dagger}\,\boldsymbol{\Sigma}\Big]\tilde{\mathbf{y}}_j\Big] \\ &=\; \frac{1}{2l+1}\Big[\sum_m\tilde{\mathbf{y}}_j^\dagger\tilde{\mathbf{y}}_j + {\rm tr}\Big[\boldsymbol{\Sigma}\sum_m\tilde{\mathbf{Y}}_j^{(1)\,\dagger}\,\tilde{\mathbf{Y}}_j^{(1)}\Big] + \sum_m\Big(\tilde{\mathbf{y}}_j^\dagger\,\tilde{\mathbf{z}}_j + \tilde{\mathbf{z}}_j^\dagger\,\tilde{\mathbf{y}}_j\Big)\Big], \end{aligned} \tag{B1}$$

where $j = \ell^2 + \ell + m$. From this, Eq. (23) follows.

To calculate the expression for the variance we first compute, (here j and j' correspond to the same multipole, ℓ, and two different values of m),

$$\begin{aligned} \sum_{m,m'}\langle\tilde{\mathbf{r}}_j^{\rm cmb\,\dagger}\,\tilde{\mathbf{r}}_j^{\rm cmb}\,\tilde{\mathbf{r}}_{j'}^{\rm cmb\,\dagger}\,\tilde{\mathbf{r}}_{j'}^{\rm cmb}\rangle = \sum_{m,m'}\Big[&{\rm tr}\left[\tilde{\mathbf{Y}}_j^{(1)}\boldsymbol{\Sigma}\tilde{\mathbf{Y}}_j^{(1)\dagger}\right]\,{\rm tr}\left[\tilde{\mathbf{Y}}_{j'}^{(1)}\boldsymbol{\Sigma}\tilde{\mathbf{Y}}_{j'}^{(1)\dagger}\right] + 2\,{\rm tr}\left[\tilde{\mathbf{Y}}_j^{(1)\dagger}\,\boldsymbol{\Sigma}\tilde{\mathbf{Y}}_{j'}^{(1)}\right]\,{\rm tr}\left[\tilde{\mathbf{Y}}_{j'}^{(1)\dagger}\,\boldsymbol{\Sigma}\tilde{\mathbf{Y}}_j^{(1)}\right] \\ &+\;\tilde{\mathbf{y}}_j^\dagger\tilde{\mathbf{y}}_j\,\tilde{\mathbf{y}}_{j'}^\dagger\tilde{\mathbf{y}}_{j'} + \tilde{\mathbf{y}}_j^\dagger\tilde{\mathbf{y}}_j{\rm tr}\Big[\boldsymbol{\Sigma}\,\tilde{\mathbf{Y}}_{j'}^{(1)\dagger}\tilde{\mathbf{Y}}_{j'}^{(1)}\Big] + \tilde{\mathbf{y}}_j^\dagger\tilde{\mathbf{y}}_{j'}{\rm tr}\Big[\boldsymbol{\Sigma}\,\tilde{\mathbf{Y}}_{j'}^{(1)\dagger}\,\tilde{\mathbf{Y}}_j^{(1)}\Big] \\ &+\;\tilde{\mathbf{y}}_{j'}^\dagger\tilde{\mathbf{y}}_j{\rm tr}\Big[\boldsymbol{\Sigma}\,\tilde{\mathbf{Y}}_j^{(1)\dagger}\tilde{\mathbf{Y}}_{j'}^{(1)}\Big] + \tilde{\mathbf{y}}_{j'}^\dagger\tilde{\mathbf{y}}_{j'}{\rm tr}\Big[\boldsymbol{\Sigma}\,\tilde{\mathbf{Y}}_j^{(1)\dagger}\tilde{\mathbf{Y}}_j^{(1)}\Big] + \tilde{\mathbf{y}}_j^\dagger\tilde{\mathbf{y}}_j\tilde{\mathbf{y}}_{j'}^\dagger{\rm tr}\Big[\tilde{\mathbf{Y}}_{j'}^{(2)}\boldsymbol{\Sigma}\Big] \\ &+\;\tilde{\mathbf{y}}_j^\dagger\tilde{\mathbf{y}}_j\tilde{\mathbf{y}}_{j'}{\rm tr}\Big[\tilde{\mathbf{Y}}_{j'}^{(2)}\boldsymbol{\Sigma}\Big] + \tilde{\mathbf{y}}_{j'}^\dagger\tilde{\mathbf{y}}_{j'}\tilde{\mathbf{y}}_j^\dagger{\rm tr}\Big[\tilde{\mathbf{Y}}_j^{(2)}\boldsymbol{\Sigma}\Big] + \tilde{\mathbf{y}}_{j'}^\dagger\tilde{\mathbf{y}}_{j'}\tilde{\mathbf{y}}_j{\rm tr}\Big[\tilde{\mathbf{Y}}_j^{(2)}\boldsymbol{\Sigma}\Big]\Big], \end{aligned} \tag{B2}$$

where we have retained contributions the lowest order in $\boldsymbol{\delta}$ separately for the terms related to the bias and the bias-free ones. From this we now have,

$$\begin{aligned} {\rm Var}(C_\ell^{\rm res}) \;&=\; \frac{1}{(2\ell+1)^2}\sum_{m,m'}\langle\tilde{\mathbf{r}}_j^{\rm cmb\,\dagger}\,\tilde{\mathbf{r}}_j^{\rm cmb}\,\tilde{\mathbf{r}}_{j'}^{\rm cmb\,\dagger}\,\tilde{\mathbf{r}}_{j'}^{\rm cmb}\rangle - (C_\ell^{\rm res})^2 \\ &=\; \frac{1}{(2\ell+1)^2}\left\{2\left[\sum_m{\rm tr}\left(\tilde{\mathbf{Y}}_j^{(1)}\boldsymbol{\Sigma}\tilde{\mathbf{Y}}_j^{(1)\dagger}\right)\right]^2 + \sum_{m,m'}\left[\tilde{\mathbf{y}}_j^\dagger\tilde{\mathbf{y}}_{j'}{\rm tr}\Big[\boldsymbol{\Sigma}\,\tilde{\mathbf{Y}}_{j'}^{(1)\dagger}\,\tilde{\mathbf{Y}}_j^{(1)}\Big] + \tilde{\mathbf{y}}_{j'}^\dagger\tilde{\mathbf{y}}_j{\rm tr}\Big[\boldsymbol{\Sigma}\,\tilde{\mathbf{Y}}_j^{(1)\dagger}\tilde{\mathbf{Y}}_{j'}^{(1)}\Big]\right]\right\}, \end{aligned} \tag{B3}$$

which is equivalent to Eq. (28).

Appendix C: Cosmological parameter likelihood

Here, we derive explicit equations for the calculation of first order derivatives of the cosmological parameter likelihood. We note that the same type of calculations are in principle involved in the computation of the matrix of the second derivatives of the likelihood. Importantly, also these derivatives can be expressed solely via cross-spectra of the pixel-domain objects defined in Sect. II C 2. However, the number of terms of the analytic expression grows rapidly and we have found that in practice computing this matrix numerically is more efficient and thus useful, while ensuring precision sufficient for the purpose. For this reason we do not include explicitly the corresponding derivations.

1. Preliminaries

Let us first compute the inverse of the covariance matrix, $\mathbf{C}$. This can be done with help of the Sherman-Morrison-Woodbury formula (e.g., [29]),

$$\mathbf{C}^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\tilde{\mathbf{Y}}^{(1)} \underbrace{(\boldsymbol{\Sigma}^{-1} + \tilde{\mathbf{Y}}^{(1)\dagger}\mathbf{D}^{-1}\tilde{\mathbf{Y}}^{(1)})^{-1}}_{\equiv \mathbf{U}} \tilde{\mathbf{Y}}^{(1)\dagger}\mathbf{D}^{-1}, \tag{C1}$$

where $\mathbf{U}$ is a square, real, symmetric matrix of rank n_β. On noting that,

$$[\tilde{\mathbf{Y}}^{(1)\dagger}\mathbf{D}^{-1}\tilde{\mathbf{Y}}^{(1)}]_{\beta\beta'} = \sum_{j,j'} \tilde{\mathbf{Y}}^{(1)\dagger}_{j\beta}(\mathbf{D}^{-1})_{jj'}\tilde{\mathbf{Y}}^{(1)}_{j'\beta'} = \sum_\ell C_\ell^{-1}(2\ell+1)\otimes_\ell(\tilde{\mathbf{Y}}^{(1)}_\beta, \tilde{\mathbf{Y}}^{(1)}_{\beta'}), \tag{C2}$$

where $j = \ell^2 + \ell + m$, the inverse of $\mathbf{U}$ can be expressed as,

$$\mathbf{U^{-1}} = \boldsymbol{\Sigma}^{-1} + \tilde{\mathbf{Y}}^{(1)\,\dagger}\mathbf{D}^{-1}\tilde{\mathbf{Y}}^{(1)} = \boldsymbol{\Sigma}^{-1} + \sum_\ell C_\ell^{-1}(2\ell+1)\otimes_\ell(\tilde{\mathbf{Y}}^{(1)}, \tilde{\mathbf{Y}}^{(1)}). \tag{C3}$$

Typically, this matrix will be dense and its inversion has to be then calculated numerically. However, given the typically limited number of spectral parameters, this does not pose any computational problems and matrix $\mathbf{U}$ can be readily derived.

We can now write down the explicit expression for $(\mathbf{C}^{-1})_{jj'}$, which is given by,

$$(\mathbf{C}^{-1})_{jj'} \;=\; C_\ell^{-1}\delta_{jj'} - \sum_{\substack{j'',j'''\\ \beta,\beta'}} C_\ell^{-1}\delta_{jj''}\tilde{\mathbf{Y}}^{(1)}_{j''\beta}\,\mathbf{U}_{\beta\beta'}\tilde{\mathbf{Y}}^{(1)\dagger}_{j'''\beta'}\,C_{\ell'}^{-1}\delta_{j'''j'} = C_\ell^{-1}\delta_{jj'} - C_\ell^{-1}C_{\ell'}^{-1}\sum_{\beta,\beta'}\tilde{\mathbf{Y}}^{(1)}_{j\beta}\,\mathbf{U}_{\beta\beta'}\tilde{\mathbf{Y}}^{(1)\dagger}_{j'\beta'}, \tag{C4}$$

where ℓ is related to j and ℓ' to j'. Given this we can write,

$$(\mathbf{C}^{-1})^2_{jj'} \;=\; C_\ell^{-2}\delta_{jj'} - 2C_\ell^{-3}\delta_{jj'}\sum_{\beta,\beta'}\tilde{\mathbf{Y}}^{(1)}_{j\beta}\mathbf{U}_{\beta\beta'}\tilde{\mathbf{Y}}^{(1)\dagger}_{j'\beta'} + C_\ell^{-2}C_{\ell'}^{-2}\sum_{\substack{\beta,\beta'\\ \alpha,\alpha'}}\tilde{\mathbf{Y}}^{(1)}_{j\beta}\tilde{\mathbf{Y}}^{(1)}_{j\alpha}\mathbf{U}_{\beta\beta'}\mathbf{U}_{\alpha\alpha'}\tilde{\mathbf{Y}}^{(1)\dagger}_{j'\beta'}\tilde{\mathbf{Y}}^{(1)\dagger}_{j'\alpha'}, \tag{C5}$$

which will be found useful later on.

Similarly, for an arbitrary harmonic space vector, $\tilde{\mathbf{x}}$, we have on using Eq. (C4),

$$(\mathbf{C}^{-1}\tilde{\mathbf{x}})_j \;=\; \equiv C_\ell^{-1}\Big(\tilde{\mathbf{x}}_j - \sum_\beta \tilde{\mathbf{Y}}^{(1)}_{j\beta}\,\mathbf{u}_\beta[\tilde{\mathbf{x}}]\Big), \tag{C6}$$

where,

$$\mathbf{u}_\beta[\tilde{\mathbf{x}}] \equiv \sum_{\beta'}\mathbf{U}_{\beta\beta'}\sum_{\ell'}\frac{2\ell'+1}{C_{\ell'}}\otimes_{\ell'}(\tilde{\mathbf{Y}}^{(1)}_{\beta'}, \tilde{\mathbf{x}}). \tag{C7}$$

2. Likelihood

The 'χ^2-term' of the likelihood can be represented as follows, Eq. (35),

$$\mathrm{tr}\,\mathbf{C}^{-1}\mathbf{E} = \mathrm{tr}\,\mathbf{C}^{-1}\hat{\mathbf{C}} + \mathrm{tr}\,\mathbf{C}^{-1}\left(\mathbf{E} - \hat{\mathbf{C}}\right), \tag{C8}$$

where $\hat{\mathbf{C}}$ stands for the true signal covariance matrix. We can write the first term on the rhs of this equation as,

$$\begin{aligned}\mathrm{tr}\,\mathbf{C}^{-1}\hat{\mathbf{C}} = \sum_{j,j'}(\mathbf{C}^{-1})_{jj'}\hat{\mathbf{C}}_{j'j} &= \sum_{j,j'}\Big(C_\ell^{-1}\delta_{jj'} - C_\ell^{-1}C_{\ell'}^{-1}\sum_{\beta,\beta'}\tilde{\mathbf{Y}}^{(1)}_{j\beta}\,\mathbf{U}_{\beta\beta'}\tilde{\mathbf{Y}}^{(1)\dagger}_{j'\beta'}\Big)\Big(\hat{C}_{\ell'}\delta_{jj'} + \sum_{\beta,\beta'}\tilde{\mathbf{Y}}^{(1)}_{j\beta}\mathbf{\Sigma}_{\beta\beta'}\tilde{\mathbf{Y}}^{(1)\dagger}_{j'\beta'}\Big)\\ &= \sum_\ell\left[(2\ell+1)\frac{\hat{C}_\ell}{C_\ell}\Big(1 - C_\ell^{-1}\,\mathrm{tr}\,[\mathbf{U}\otimes_\ell(\tilde{\mathbf{Y}}^{(1)},\tilde{\mathbf{Y}}^{(1)})]\Big) + \frac{(2\ell+1)}{C_\ell}\,\mathrm{tr}\,[\mathbf{\Sigma}\otimes_\ell(\tilde{\mathbf{Y}}^{(1)},\tilde{\mathbf{Y}}^{(1)})]\right]\\ &- \sum_{\ell,\ell'}\frac{(2\ell+1)}{C_\ell}\,\frac{(2\ell'+1)}{C_{\ell'}}\mathrm{tr}\,[\mathbf{U}\otimes_{\ell'}(\tilde{\mathbf{Y}}^{(1)},\tilde{\mathbf{Y}}^{(1)})\,\mathbf{\Sigma}\otimes_\ell(\tilde{\mathbf{Y}}^{(1)},\tilde{\mathbf{Y}}^{(1)})],\end{aligned} \tag{C9}$$

where for the computational reasons it is better to first performs the sum over ℓ and ℓ' before calculating traces. In particular the last term becomes then linear in ℓ_{max}. Similarly, we can now write the second term of the χ^2,

$$\begin{aligned}\mathrm{tr}\,\mathbf{C}^{-1}\left(\mathbf{E}-\hat{\mathbf{C}}\right) &= \sum_{j,j'}(\mathbf{C}^{-1})_{jj'}(\tilde{\mathbf{y}}_{j'}\tilde{\mathbf{y}}_j^\dagger + \tilde{\mathbf{z}}_{j'}\tilde{\mathbf{y}}_j^\dagger + \tilde{\mathbf{y}}_{j'}\tilde{\mathbf{z}}_j^\dagger) = \sum_\ell\frac{(2\ell+1)}{C_\ell}(\otimes_\ell(\tilde{\mathbf{y}},\tilde{\mathbf{y}}) + \otimes_\ell(\tilde{\mathbf{z}},\tilde{\mathbf{y}}) + \otimes_\ell(\tilde{\mathbf{y}},\tilde{\mathbf{z}}))\\ &- \sum_{\ell,\ell'}\frac{(2\ell+1)}{C_\ell}\,\frac{(2\ell'+1)}{C_{\ell'}}\,\mathrm{tr}\Big[\mathbf{U}\Big(\otimes_{\ell'}(\tilde{\mathbf{Y}}^{(1)},\tilde{\mathbf{y}})\,\otimes_\ell(\tilde{\mathbf{y}},\tilde{\mathbf{Y}}^{(1)}) + \otimes_{\ell'}(\tilde{\mathbf{Y}}^{(1)},\tilde{\mathbf{y}})\,\otimes_\ell(\tilde{\mathbf{z}},\tilde{\mathbf{Y}}^{(1)})\\ &\qquad + \otimes_{\ell'}(\tilde{\mathbf{Y}}^{(1)},\tilde{\mathbf{z}})\,\otimes_\ell(\tilde{\mathbf{y}},\tilde{\mathbf{Y}}^{(1)})\Big)\Big],\end{aligned} \tag{C10}$$

and again in a numerical implementation it is better to first perform sums over the multipoles and only later take the traces.

The determinant of the assumed covariance matrix, $\mathbf{C}$, can then be efficiently calculated by noting that,

$$\det\left(\begin{bmatrix}\mathbf{D} & -\tilde{\mathbf{Y}}^{(1)}\\ \tilde{\mathbf{Y}}^{(1)\,\dagger} & \mathbf{\Sigma}^{-1}\end{bmatrix}\right) = \det\mathbf{D}\,\det\left(\mathbf{\Sigma}^{-1} + \tilde{\mathbf{Y}}^{(1)\,\dagger}\,\mathbf{D}^{-1}\,\tilde{\mathbf{Y}}^{(1)}\right) = \det\mathbf{\Sigma}^{-1}\,\det\left(\mathbf{D} + \tilde{\mathbf{Y}}^{(1)}\,\mathbf{\Sigma}\,\tilde{\mathbf{Y}}^{(1)\,\dagger}\right), \tag{C11}$$

and thus

$$\det\mathbf{C} = \det\mathbf{D}\,\frac{\det\mathbf{\Sigma}}{\det\mathbf{U}}. \tag{C12}$$

Given that $\mathbf{D}$ is diagonal and $\mathbf{U}$ and $\mathbf{\Sigma}$ – small, the computation of the determinant of the covariance does not pose typically any problem. Note that as these two latter matrices are typically dense their determinants need to be calculated numerically using standard techniques.

3. First derivatives of the likelihood

We can now compute the likelihood gradient. From Eq. (36) we can write,

$$\langle\mathcal{S}^{par}_{,i}\rangle = \mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,i} - \mathbf{C}^{-1}\mathbf{C}_{,i}\,\mathbf{C}^{-1}\,\mathbf{E}\right] = \mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,i} - \mathbf{C}^{-1}\mathbf{C}_{,i}\,\mathbf{C}^{-1}\,\hat{\mathbf{C}}\right] - \mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,i}\,\mathbf{C}^{-1}(\mathbf{E}-\hat{\mathbf{C}})\right], \tag{C13}$$

where the second part contains all the extra corrections from the model mismatches, while the former vanishes when the true and estimated parameters are the same, i.e, $\hat{\mathbf{C}} = \mathbf{C}$.

We compute each of these two terms separately below.

a. $\mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,i} - \mathbf{C}^{-1}\mathbf{C}_{,i}\,\mathbf{C}^{-1}\,\hat{\mathbf{C}}\right]$:

We first observe that,

$$\hat{\mathbf{C}} - \mathbf{C} = \hat{\mathbf{D}} - \mathbf{D}, \tag{C14}$$

where $\hat{\mathbf{D}}$ stands for $\mathbf{D}$ computed for the true values of the cosmological parameters. Hence, we can represent $\hat{\mathbf{C}}$ as,

$$\hat{\mathbf{C}} = \mathbf{C} + \hat{\mathbf{D}} - \mathbf{D} \equiv \mathbf{C} + \Delta\mathbf{D}, \tag{C15}$$

where,

$$\Delta\mathbf{D}_{jj'} = (\hat{C}_\ell - C_\ell)\,\delta_{jj'} \equiv \Delta C_\ell^{\rm cmb}\,\delta_{jj'}. \tag{C16}$$

On using Eq. (C15) we can rewrite the complete term as,

$$\mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,i} - \mathbf{C}^{-1}\mathbf{C}_{,i}\,\mathbf{C}^{-1}\,\hat{\mathbf{C}}\right] = -\mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,i}\,\mathbf{C}^{-1}\,\Delta\mathbf{D}\right], \tag{C17}$$

and then,

$$\begin{aligned}
\mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,i}\,\mathbf{C}^{-1}\,\Delta\mathbf{D}\right] &= \sum_{j,j',j'',j'''} (\mathbf{C}^{-1})_{jj'}\,(\mathbf{C}_{,i})_{j'j''}\,(\mathbf{C}^{-1})_{j''j'''}\,\Delta\mathbf{D}_{j'''j} = \sum_{j,j'} (\mathbf{C}^{-1})^2_{jj'}\,\frac{\partial C_{\ell'}^{\rm cmb}}{\partial p_i}\,\Delta C_\ell^{\rm cmb} \\
&= \sum_{j,j'}\Big(C_\ell^{-2}\delta_{jj'} - 2C_\ell^{-3}\delta_{jj'}\sum_{\beta,\beta'}\tilde{\mathbf{Y}}^{(1)}_{j\beta}\,\mathbf{U}_{\beta\beta'}\tilde{\mathbf{Y}}^{(1)\dagger}_{j'\beta'} + C_\ell^{-2}C_{\ell'}^{-2}\sum_{\substack{\beta,\beta'\\ \alpha,\alpha'}}\tilde{\mathbf{Y}}^{(1)}_{j\beta}\,\tilde{\mathbf{Y}}^{(1)}_{j\alpha}\,\mathbf{U}_{\beta\beta'}\,\mathbf{U}_{\alpha\alpha'}\,\tilde{\mathbf{Y}}^{(1)\dagger}_{j'\beta'}\,\tilde{\mathbf{Y}}^{(1)\dagger}_{j'\alpha'}\Big)\frac{\partial C_{\ell'}^{\rm cmb}}{\partial p_i}\Delta C_\ell, \\
&= \sum_\ell \frac{\partial C_\ell^{\rm cmb}}{\partial p_i}\,\Delta C_\ell\frac{(2\ell+1)}{C_\ell^2} - \mathrm{tr}\Big\{\mathbf{U}\sum_\ell\Big[\frac{2(2\ell+1)}{C_\ell^3}\,\frac{\partial C_\ell^{\rm cmb}}{\partial p_i}\,\Delta C_\ell\,\otimes_\ell(\tilde{\mathbf{Y}}^{(1)},\tilde{\mathbf{Y}}^{(1)})\Big]\Big\} + \\
&\quad + \mathrm{tr}\Big\{\mathbf{U}\sum_\ell\Big[\frac{(2\ell+1)}{C_\ell^2}\frac{\partial C_\ell^{\rm cmb}}{\partial p_i}\otimes_\ell(\tilde{\mathbf{Y}}^{(1)\dagger},\tilde{\mathbf{Y}}^{(1)})\Big]\mathbf{U}\sum_\ell\Big[\frac{(2\ell+1)}{C_\ell^2}\,\Delta C_\ell\,\otimes_\ell(\tilde{\mathbf{Y}}^{(1)},\tilde{\mathbf{Y}}^{(1)\dagger})\Big]\Big\}
\end{aligned} \tag{C18}$$

where p_i is the parameter we estimate and we have used Eq. (C5) derived earlier and the fact that $\mathbf{U}$ is positive and symmetric. We note that the expression on the right hand side above is manifestly real as it should.

b. $\mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,i}\,\mathbf{C}^{-1}(\mathbf{E}-\hat{\mathbf{C}})\right]$:

On using Eq. (40) and the fact that the covariance, $\mathbf{C}$, is symmetric and real, we can rewrite this term as follows,

$$\begin{aligned}
\mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,i}\,\mathbf{C}^{-1}(\mathbf{E}-\hat{\mathbf{C}})\right] &= \mathrm{tr}\Big\{\mathbf{C}_{,i}\left[(\mathbf{C}^{-1}\tilde{\mathbf{y}})(\mathbf{C}^{-1}\tilde{\mathbf{y}})^\dagger + (\mathbf{C}^{-1}\tilde{\mathbf{y}})(\mathbf{C}^{-1}\tilde{\mathbf{z}})^\dagger + (\mathbf{C}^{-1}\tilde{\mathbf{z}})(\mathbf{C}^{-1}\tilde{\mathbf{y}})^\dagger\right]\Big\} \\
&= \sum_j\Big\{\frac{\partial C_\ell}{\partial p_i}\Big[(\mathbf{C}^{-1}\tilde{\mathbf{y}})_j(\mathbf{C}^{-1}\tilde{\mathbf{y}})^\dagger_j + (\mathbf{C}^{-1}\tilde{\mathbf{z}})_j(\mathbf{C}^{-1}\tilde{\mathbf{y}})^\dagger_j + (\mathbf{C}^{-1}\tilde{\mathbf{y}})_j(\mathbf{C}^{-1}\tilde{\mathbf{z}})^\dagger_j\Big]\Big\}.
\end{aligned} \tag{C19}$$

where as usual $j = \ell^2+\ell+m$. Therefore, on using Eq. (C6),

$$\begin{aligned}
\mathrm{tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,i}\,\mathbf{C}^{-1}(\mathbf{E}-\hat{\mathbf{C}})\right] &= \sum_\ell\Big\{\frac{2\ell+1}{C_\ell^2}\,\frac{\partial C_\ell}{\partial p_i}\Big[\otimes_\ell(\tilde{\mathbf{y}},\tilde{\mathbf{y}}) - 2\,\Re\sum_\beta\otimes_\ell(\tilde{\mathbf{y}},\tilde{\mathbf{Y}}^{(1)}_\beta)\,\tilde{\mathbf{u}}_\beta[\tilde{\mathbf{y}}] + \sum_{\beta,\beta'}\otimes_\ell(\tilde{\mathbf{Y}}^{(1)}_\beta,\tilde{\mathbf{Y}}^{(1)}_{\beta'})\,\tilde{\mathbf{u}}^\dagger_\beta[\tilde{\mathbf{y}}]\,\tilde{\mathbf{u}}_{\beta'}[\tilde{\mathbf{y}}] \\
&\quad + 2\,\Re\Big(\otimes_\ell(\tilde{\mathbf{y}},\tilde{\mathbf{z}}) - \sum_\beta\otimes_\ell(\tilde{\mathbf{y}},\tilde{\mathbf{Y}}^{(1)}_\beta)\,\tilde{\mathbf{u}}_\beta[\tilde{\mathbf{z}}] - \sum_\beta\otimes_\ell(\tilde{\mathbf{z}},\tilde{\mathbf{Y}}^{(1)}_\beta)\,\tilde{\mathbf{u}}_\beta[\tilde{\mathbf{y}}] + \sum_{\beta,\beta'}\otimes_\ell(\tilde{\mathbf{Y}}^{(1)}_\beta,\tilde{\mathbf{Y}}^{(1)}_{\beta'})\,\tilde{\mathbf{u}}^\dagger_\beta[\tilde{\mathbf{y}}]\,\tilde{\mathbf{u}}_{\beta'}[\tilde{\mathbf{z}}]\Big)\Big].
\end{aligned} \tag{C20}$$

Appendix D: Special case: homogeneous noise and pixel-independent scaling laws.

Let us assume that the noise of the frequency maps is homogeneous and that we use global scaling laws for all considered components and all considered pixels. Then, matrix $\mathbf{W}_p^{0k}(\beta)$ and its derivatives with respect to the spectra parameters are all the same for all considered pixels, p. We can then drop the pixel subscripts and introduce following pixel-independent objects,

$$\mathbf{w}_k \equiv \mathbf{W}^{0k}(\bar{\beta}), \quad \partial\mathbf{W}_{k\beta} \equiv \frac{\partial\mathbf{W}^{0k}}{\partial\beta}\Big|_{\bar{\beta}}, \quad \boldsymbol{v}_k \equiv \sum_{\beta,\beta'}\frac{\partial^2\mathbf{W}^{0k}}{\partial\beta\,\partial\beta'}\Big|_{\bar{\beta}}\boldsymbol{\Sigma}_{\beta'\beta}, \tag{D1}$$

so then, we can rewrite Eqs. (17) and (25) as,

$$\begin{aligned}
\tilde{\mathbf{y}} &= \tilde{\mathbf{F}}\,\mathbf{w} \\
\tilde{\mathbf{Y}}^{(1)} &= \tilde{\mathbf{F}}\,\partial\mathbf{W} \\
\tilde{\mathbf{z}} &= \tilde{\mathbf{F}}\,\boldsymbol{v}.
\end{aligned} \tag{D2}$$

Subsequently, we can relate all the cross-correlations appearing in the earlier equations to the cross-correlation matrix of the foreground signal in different frequency bands, $\mathcal{F}_\ell^{\rm fore}$, defined in Eq. (27),

$$\begin{aligned}
\otimes_\ell(\tilde{\mathbf{y}},\tilde{\mathbf{y}}) &= \mathbf{w}^t\mathcal{F}_\ell^{\rm fore}\mathbf{w},\\
\otimes_\ell(\tilde{\mathbf{Y}}^{(1)},\tilde{\mathbf{Y}}^{(1)}) &= \partial\mathbf{W}^t\,\mathcal{F}_\ell^{\rm fore}\,\partial\mathbf{W},\\
\otimes_\ell(\tilde{\mathbf{y}},\tilde{\mathbf{z}}) &= \mathbf{w}^t\mathcal{F}_\ell^{\rm fore}\boldsymbol{v},\\
\otimes_\ell(\tilde{\mathbf{Y}}^{(1)},\tilde{\mathbf{y}}) &= \partial\mathbf{W}^t\,\mathcal{F}_\ell^{\rm fore}\,\mathbf{w},\\
\otimes_\ell(\tilde{\mathbf{Y}}^{(1)},\tilde{\mathbf{z}}) &= \partial\mathbf{W}^t\,\mathcal{F}_\ell^{\rm fore}\,\boldsymbol{v}.
\end{aligned} \tag{D3}$$

[1] B. Gold, N. Odegard, J. L. Weiland, R. S. Hill, A. Kogut, C. L. Bennett, G. Hinshaw, X. Chen, J. Dunkley, M. Halpern, et al., ApJS **192**, 15 (2011), 1001.4555.

[2] Planck Collaboration, P. A. R. Ade, N. Aghanim, D. Alina, M. I. R. Alves, G. Aniano, C. Armitage-Caplan, M. Arnaud, D. Arzoumanian, M. Ashdown, et al., A&A **576**, A105 (2015), 1405.0872.

[3] N. Krachmalnicoff, C. Baccigalupi, J. Aumont, M. Bersanelli, and A. Mennella, A&A **588**, A65 (2016).

[4] K. N. Abazajian, K. Arnold, J. Austermann, B. A. Benson, C. Bischoff, J. Bock, J. R. Bond, J. Borrill, I. Buder, D. L. Burke, et al., Astroparticle Physics **63**, 55 (2015), 1309.5381.

[5] L. Verde, H. V. Peiris, and R. Jimenez, J. Cosmology Astropart. Phys. **1**, 019 (2006), astro-ph/0506036.

[6] A. Amblard, A. Cooray, and M. Kaplinghat, Phys. Rev. D **75**, 083508 (2007), astro-ph/0610829.

[7] J. Dunkley, A. Amblard, C. Baccigalupi, M. Betoule, D. Chuss, A. Cooray, J. Delabrouille, C. Dickinson, G. Dobler, J. Dotson, et al., AIP Conf. Proceedings **1141**, 222 (2009), 0811.3915.

[8] A. Bonaldi and S. Ricciardi, MNRAS **414**, 615 (2011), 1101.4876.

[9] P. Creminelli, D. López Nacir, M. Simonović, G. Trevisan, and M. Zaldarriaga, J. Cosmology Astropart. Phys. **11**, 031 (2015), 1502.01983.

[10] A. Kogut and D. J. Fixsen, ApJ **826**, 101 (2016), 1607.02150.

[11] N. Katayama and E. Komatsu, ApJ **737**, 78 (2011), 1101.5210.

[12] M. Remazeilles, C. Dickinson, H. K. K. Eriksen, and I. K. Wehus, MNRAS **458**, 2032 (2016), 1509.04714.

[13] D. Alonso, J. Dunkley, S. Naess, and B. Thorne, ArXiv e-prints (2016), 1608.00551.

[14] J. Errard, F. Stivoli, and R. Stompor, Phys. Rev. D **84**, 063005 (2011).

[15] W. N. Brandt, C. R. Lawrence, A. C. S. Readhead, J. N. Pakianathan, and T. M. Fiola, ApJ **424**, 1 (1994).

[16] H. K. Eriksen, C. Dickinson, C. R. Lawrence, C. Baccigalupi, A. J. Banday, K. M. Górski, F. K. Hansen, P. B. Lilje, E. Pierpaoli, M. D. Seiffert, et al., ApJ **641**, 665 (2006), arXiv:astro-ph/0508268.

[17] R. Stompor, S. Leach, F. Stivoli, and C. Baccigalupi, MNRAS **392**, 216 (2009), 0804.2645.

[18] J. Errard and R. Stompor, Phys. Rev. D **85**, 083006 (2012), 1203.5285.

[19] J. Errard, S. M. Feeney, H. V. Peiris, and A. H. Jaffe, J. Cosmology Astropart. Phys. **3**, 052 (2016), 1509.06770.

[20] J. Poh and S. Dodelson, ArXiv e-prints (2016), 1606.08922.

[21] R. R. Caldwell, C. Hirata, and M. Kamionkowski, ArXiv e-prints (2016), 1608.08138.

[22] F. Stivoli, J. Grain, S. M. Leach, M. Tristram, C. Baccigalupi, and R. Stompor, MNRAS **408**, 2319 (2010), 1004.4756.

[23] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, and et al., A&A **536**, A1 (2011), 1101.2022.

[24] C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, et al., ApJS **148**, 1 (2003), astro-ph/0302207.

[25] M. Tegmark, A. N. Taylor, and A. F. Heavens, ApJ **480**, 22 (1997), astro-ph/9603021.

[26] Planck Collaboration, N. Aghanim, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, et al., ArXiv e-prints (2016), 1606.07335.

[27] Planck Collaboration, P. A. R. Ade, M. I. R. Alves, G. Aniano, C. Armitage-Caplan, M. Arnaud, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, et al., A&A **576**, A107 (2015), 1405.0874.

[28] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, ApJ **622**, 759 (2005), arXiv:astro-ph/0409513.

[29] G. H. Golub and C. F. van Loan, Matrix computations (Johns Hopkins University Press, 1996).

[30] `http://ltd16.grenoble.cnrs.fr/IMG/UserFiles/Images/09_TMatsumura_20150720_LTD_v18.pdf`

[31] `http://pla.esac.esa.int/pla`

[32] `http://pla.esac.esa.int/pla`

[33] `http://portal.nersc.gov/project/mp107/index.html`