Needlet thresholding methods in component separation

F. Oppizzi,[a,b,1] **A. Renzi,**[b] **M. Liguori,**[a,b,c] **F. K. Hansen**[d] **D. Marinucci**[e] **C. Baccigalupi**[f,g] **D. Bertacca**[a,b] **D. Poletti**[f]

[a]Dipartimento di Fisica e Astronomia "G. Galilei", Università degli Studi di Padova, Via Marzolo 8, 35131 Padova, Italy

[b]INFN, Sezione di Padova, via Marzolo 8, I-35131, Padova, Italy

[d]Institute of Theoretical Astrophysics, University of Oslo, Blindern, Oslo, Norway

[e]Dipartimento di Matematica, Università di Roma Tor Vergata

[f]SISSA - Scuola Internazionale Superiore di Studi Avanzati,Via Bonomea 265, 34136, Trieste, Italy

[g]INFN, Sezione di Trieste, Padriciano, 99, 34149 Trieste, Italy

E-mail: filippo.oppizzi@pd.infn.it, alessandro.renzi@pd.infn.it, michele.liguori@pd.infn.it, frodekh@astro.uio.no, marinucc@mat.uniroma2.it, carlo.baccigalupi@sissa.it, daniele.bertacca@pd.infn.it, davide.poletti@sissa.it

Abstract. Foreground components in the Cosmic Microwave Background (CMB) are sparse in a needlet representation, due to their specific morphological features (anisotropy, non-Gaussianity). This leads to the possibility of applying needlet thresholding procedures as a component separation tool. In this work, we develop algorithms based on different needlet-thresholding schemes and use them as extensions of existing, well-known component separation techniques, namely ILC and template-fitting. We test soft- and hard-thresholding schemes, using different procedures to set the optimal threshold level. We find that thresholding can be useful as a denoising tool for internal templates in experiments with few frequency channels, in conditions of low signal-to-noise. We also compare our method with other denoising techniques, showing that thresholding achieves the best performance in terms of reconstruction accuracy and data compression while preserving the map resolution. The best results in our tests are in particular obtained when considering template-fitting in an LSPE like experiment, especially for B-mode spectra.

---

[1]Corresponding author.

arXiv:1911.01298v1 [astro-ph.CO] 4 Nov 2019

Contents

1   Introduction

The search for primordial polarization B-modes is the main and most exciting challenge for both the current and coming generation of Cosmic Microwave Background (CMB) experiments. A detection of B-mode CMB polarization would carry huge implications for our understanding of the Early Universe, essentially allowing for a smoking-gun confirmation of Inflation, as well as for the measurement of its energy scale [1–4]. At the same time, the quest for B-modes presents formidable challenges. The polarization signal from the inflationary stochastic Gravitational Wave (GW) background is expected to be very faint. Its detection, if achievable, will require an exquisite and unprecedented level of accuracy in controlling systematic biases in the data. One of the major sources of systematic contamination, besides instrumental effects, is the astrophysical foreground [5–12].

A variety of methods have been so far designed, with the purpose to separate the CMB signal from foreground components. Of course, state-of-the-art component separation methods work very efficiently on the best available datasets, such as *Planck* [5, 13]. The target sensitivity for future B-mode surveys is, however, orders of magnitude below that of current experiments [14, 15]. This – together with the morphological complexity and incomplete understanding of polarized foregrounds – makes further study and advancements in this area still crucial.

Different component separation algorithms exploit characteristic features of foreground emission to disentangle them from the background radiation, (mostly, but not only, their non-blackbody spectrum) [16–22].

In this work, we present an investigation on a technique relying on the assumption that the foreground signal is "sparse" in a proper representation, *i.e.* the majority of the signal is concentrated in few expansion coefficients.

To this purpose, we will rely on a needlet expansion of the CMB map. Needlets are a special kind of spherical wavelets, directly defined in harmonic space and not relying on any tangent

plane approximation (see next section for details). They were developed as a functional analysis tool by [23] and applied to CMB analysis in various works [17, 24–27]. Like all wavelets, needlets display the property to be localized both in space (or time) and frequency, this is a key property to induce sparsity in the representation of coherent signals, such as foreground emission. In our analysis, we will exploit sparsity to either reconstruct foreground templates via thresholding procedures or for template denoising purposes. The idea is that these approaches can be combined with – and be used as preliminary steps in – different standard component separation methods: for example, initial template denoising in needlet space can be followed by standard template fitting; alternatively, needlet thresholding can be used for pre-cleaning single frequency channels, by exploiting morphological information (non-Gaussianity and anisotropy of foregrounds); the channels can then be combined with a common approach such as Internal Linear Combination (ILC). The idea is therefore not that of developing a component separation method per se, but rather to explore a range of applications in a general context. We will specify case by case what are the main novelties introduced in the different techniques which we are going to explore, and in which regime they find their best application. Our general finding is that the methods we explore here can be useful in situations characterized by limited frequency coverage, as it can be the case for current and forthcoming ground-based and balloon experiments. The paper is structured as follows: in section 2 we will review the main characteristic of the spherical needlet and we will introduce the notions of sparsity and thresholding. In section 3 we will show how these properties can be exploited for component separation and we will describe the techniques developed in this work. In section 4 we will check the performance of our thresholding methods on simulations of different CMB surveys, showing a comparison with alternative techniques for template denoising, as well as possible applications in which thresholding is used in synergy with other methods. Finally, section 5 is dedicated to the conclusions.

2   Needlet Regression

Spherical wavelets are usually constructed by relying on a local flat sky approximation. This means that the basis function is defined on a flat tangent plane and then implemented on the sphere. The needlet basis, instead, is defined directly in harmonic space, in terms of spherical harmonics. As we will show in the following, this is a great advantage for the exact computation of the needlet coefficients.

More specifically, needlets are defined starting from the window function $b(\ell, j)$ that sets the harmonic support for each needlet layer j. This function must satisfy three properties:

1. compact support: $b(\ell, j) > 0$ if $\ell_{min,j} \leq \ell \leq \ell_{max,j}$ and $b(\ell, j) = 0$ otherwise. This ensure that each needlet layers represents a fixed range of scales. Moreover, each layer j will have equal support in $\log(\ell)$.

2. partition to unity, so that $\sum_j b^2(\ell, j) = 1$.

3. smoothness, *i.e.* $b(\ell, j)$ is infinitely differentiable.

Given a window function with these properties (see [28] for a complete derivation), the spherical needlet basis function can be defined as:

$$\psi_{jk}(x) = \sqrt{\lambda_{jk}} \sum_\ell b(\ell B^{-j}) \sum_{m=-\ell}^{\ell} Y_\ell^m(\xi_{jk}) \overline{Y}_\ell^m(x), \qquad (2.1)$$

where j represents the needlets scale, λ_{jk} and ξ_{jk} are respectively the weights and the cubature points at the level j, and B is a parameter fixing the ℓ coverage of the needlets layers. As shown in [29], the support of $b(\ell B^{-j})$ extends over $\ell \in (B^{j-1}, B^{j+1})$, this is the multipoles window spanned by the needlets layer j. From Eq. 2.1 we can define the needlets coefficients β_{jk} of a square integrable function on the sphere $f(x)$ as:

$$\beta_{jk} = \int_{S^2} \mathrm{d}x \; f(x)\psi_{jk}(x) = \sqrt{\lambda_{jk}} \sum_\ell b(\ell B^{-j}) \sum_{m=-\ell}^{\ell} a_{\ell m} Y_\ell^m(\xi_{jk}), \qquad (2.2)$$

where we use:

$$a_{\ell m} = \int_{S^2} \mathrm{d}x \; f(x)\overline{Y}_\ell^m(x). \qquad (2.3)$$

Equation (2.2) is the direct needlet transform, and β_{jk} are the needlet coefficients at scale j in the position defined by the cubature point ξ_{jk}. The harmonic coefficients $a_{\ell m}$ can be computed numerically with a high level of precision [30], this makes the numerical implementation of needlets very convenient.

The inverse of equation (2.2) is:

$$f(x) = \sum_{jk} \beta_{jk}\psi_{jk}(x). \qquad (2.4)$$

With these preliminary definitions in hand, let us now review the main properties which make needlets a particularly suitable choice for CMB analysis. Needlets owe their name to their localization properties. The basis functions are localized quasi-exponentially around their centers, represented by the cubature points ξ_{jk}. In their seminal paper [23], Narcowich and collaborators proved this statement showing that for any point x on the sphere surface, there exists a constant c_M, such that:

$$|\psi_{jk}| \le \frac{c_M B^j}{(1 + B^j \arccos{(\xi_{jk}, x)})^M}. \qquad (2.5)$$

Note that the function arccos in the above formula represents the distance on the sphere; this states exactly that the function ψ_{jk} decreases faster than any power law. This property is of significant importance in CMB analysis, where the presence of missing observations poses a problem for the computation of harmonic coefficients due to the onset of spurious correlations between harmonic coefficients $a_{\ell m}$. These correlations represent a limitation for the evaluation of the power spectrum and the other cumulants and must be corrected for, generally with high computational costs. As proven in [29], needlet coefficients are instead much less sensitive to gaps in the map, therefore working in needlet space allows avoiding to correct for missing observation.

Needlets are also particularly well suited for the representation of random fields on the sphere, due to their uncorrelation properties. The fact that the window function $b(\ell B^{-j})$ has compact support in (B^{j-1}, B^{j+1}) indeed ensures that theoretical correlations between β_{jk} cancel if the difference in levels is greater than 2, so that if $j - j' > 2$ we have:

$$\beta_{jk}\beta_{j'k'} = \sqrt{\lambda_{jk}\lambda_{j'k'}} \sum_{\ell\ell'} b(\ell B^{-j})b(\ell' B^{-j'}) \sum_{mm'} a_{jk}a_{j'k'} Y_\ell^m(\xi_{jk}) Y\ell'^{m'}(\xi_{jk}) = 0, \qquad (2.6)$$

as the simple consequence of the fact that the supports of the two basis functions do not overlap. If we consider instead fixed scale correlations, it was proven in [24] that the needlet representation of a Gaussian random field with smooth power spectrum satisfies:

$$|Corr(\beta_{jk}\beta_{jk'})| = \left| \frac{\beta_{jk}\beta_{jk'}}{\sqrt{\beta_{jk}^2\beta_{jk'}^2}} \right| \leq \frac{c_M}{(1 + B^j \arccos(\xi_{jk}, \xi_{jk'}))}, \tag{2.7}$$

for any positive integer M and $c_M > 0$. This implies that, for growing scale j, needlets coefficients are asymptotically uncorrelated. Therefore, under Gaussianity, the small scales coefficients behave approximately as a sample of i.i.d. random variables.

2.1 Sparsity and Thresholding

A signal is said to be sparse if – in a given basis or frame – it can be reconstructed by using only a small amount of basis elements (see *e.g.* [31] for a complete review on the subject). Wavelets and needlets present several crucial features which allow for sparse representation of signals. First, they are not an orthonormal basis but instead a tight frame. This implies that the basis contains redundant elements. Furthermore, their tight space-frequency localization properties make them suitable to identify discontinuities in the signal and represent them with just a small number of modes. In such conditions, sparsity can always be achieved if the signal under study is smooth (although this is not a necessary condition). This makes sparsity a key element in wavelet regression methods [32], since it allows developing efficient techniques to separate a coherent signal (which is sparse, for the reasons just mentioned above) from a stochastic "noise" component. Note that in our setting, CMB itself can be viewed as part of this "noise" (for the purpose of foreground estimation). In general, stochastic fields do not admit a sparse representation, due to their lack of smoothness and, in the case of CMB, to its isotropy. Such separation can be essentially achieved by setting to zero all the coefficients under a certain threshold and eventually rescaling the remaining ones. This clearly filters the few large signal coefficients from the noise background. Such a procedure is referred to as thresholding and it is, in spirit, similar to principal component analysis, since it aims to reduce the complexity of a multidimensional data-set, by identifying the most significant modes.

The simplest thresholding scheme is called hard thresholding (HT); the effects of the hard thresholding operator on the needlet coefficients are simply:

$$HT(\beta_{jk}) = \begin{cases} 0 & \text{if} \quad |\beta_{jk}| < \lambda \\ \beta_{jk} & \text{if} \quad |\beta_{jk}| \geq \lambda, \end{cases} \tag{2.8}$$

where λ is a given threshold. In the case of a coherent signal, only few significant coefficients survive this operation, providing an optimal representation as well as efficient data compression.

A second option is soft thresholding (ST). In this case, the significant coefficients are rescaled, proportionally to the chosen threshold:

$$ST(\beta_{jk}) = \text{sgn}(\beta_{jk})(|\beta_{jk}| - \lambda)_+ \begin{cases} \beta_{jk} + \lambda & \text{if} \quad \beta_{jk} \leq \lambda \\ 0 & \text{if} \quad |\beta_{jk}| < \lambda \\ \beta_{jk} - \lambda & \text{if} \quad \beta_{jk} \geq \lambda, \end{cases} \tag{2.9}$$

where the operator $(*)_+$ stands for the positive part of the argument.

It is known that the soft thresholding solution can be interpreted – from a Bayesian perspective – as a maximum posterior estimator from a Gaussian Likelihood with a leptokurtic Laplace prior on the parameters, which in our case are the needlet coefficients. Let us briefly show this. Consider a dataset $x = \theta + n$, where θ is a signal that is sparse in some basis and n is a Gaussian noise with known variance σ^2. Assume also that the scale parameter $1/\lambda$ of the Laplace prior on θ is known and its mean is 0, so that we can write:

$$P(\theta|x) \propto \mathcal{L}(x|\theta)P(\theta) = N(x; \theta, \sigma)L(\theta; \lambda, 0), \tag{2.10}$$

$$-\log P(\theta|x) = \frac{(x-\theta)^2}{2\sigma^2} + \lambda|\theta| + const, \tag{2.11}$$

where we use $N(*; \mu, \sigma)$ and $L(*; \mu, \lambda)$ to define respectively the Normal and the Laplace distributions, that is:

$$L(d; \lambda, \mu) = \frac{\lambda}{2}e^{-\lambda|d-\mu|}. \tag{2.12}$$

The maximum posterior estimator (MPE) is obtained by minimizing equation (2.11). We start by taking the derivative with respect θ (that we denote with ∂_θ):

$$\partial_\theta(-\log P(\theta|x)) = -\frac{(x-\theta)}{\sigma^2} + \lambda\partial_\theta|\theta| = 0, \tag{2.13}$$

$$\theta = x - \sigma^2\lambda\partial_\theta|\theta|, \tag{2.14}$$

since the absolute value is not differentiable around zero (and equivalently the L1 norm $||\theta||$, if dealing with multidimensional data), we should take the subgradient, so that we have:

$$\partial_\theta||\theta|| = \begin{cases} 1 & \text{if } \theta > 0 \\ -1 & \text{if } \theta < 0 \\ [-1, 1] & \text{if } \theta = 0, \end{cases} \tag{2.15}$$

note that for $\theta = 0$ the subgradient is actually an interval of values. We can understand the soft thresholding solution applying the conditions (2.15) at equation (2.14). First notice that the term $\sigma^2\lambda\partial_\theta|\theta|$ can only take values in the interval $[-\sigma^2\lambda, \sigma^2\lambda]$. Therefore, considering the case $|x| > \sigma^2\lambda$ we must have $\theta \neq 0$ to satisfy condition (2.14). From the same equation we see that it must be $\text{sgn}(\theta) = \text{sgn}(x)$. In this case, the solution is given by $\theta = x - \text{sgn}(x)\sigma^2\lambda$. For $|x| = \lambda\sigma^2$ instead, $P(\theta|x)$ is maximum in $\theta = 0$ since $\lim_{x \to \sigma^2\lambda}(x - \text{sgn}(x)\sigma^2\lambda) = 0$ (remind that $P(\theta|x)$ is continuous). At last, if we have $|x| < \sigma^2\lambda$, the only admissible solution of (2.14) is $\theta = 0$ otherwise, due to condition (2.15), it would give $\theta < 0$ for $\theta > 0$ and vice versa. After these considerations it is clear that the solution coincides with the soft thresholding operator, that in this case is:

$$ST(x) = \text{sgn}(x)(|x| - \sigma^2\lambda)_+. \tag{2.16}$$

3 Methodology

The general idea behind this investigation is that foreground signals and the CMB fluctuations can be disentangled when the data are represented in a proper basis, frame or dictionary, and that needlets represent an ideal choice to this purpose. Foreground emission comes, in the
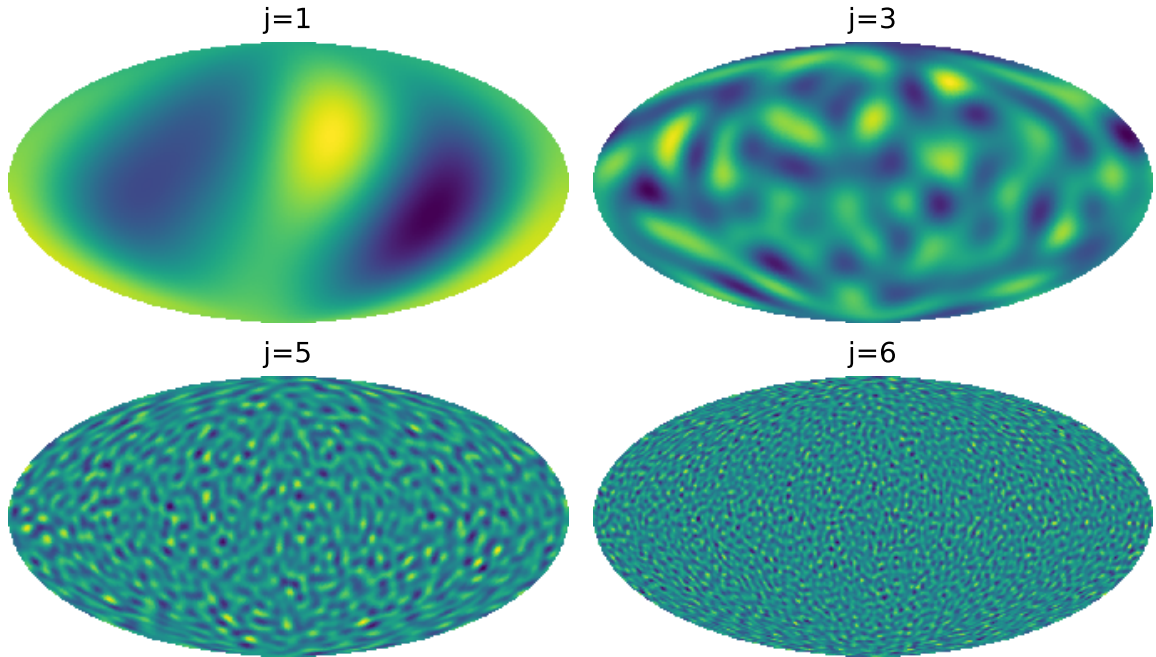
Figure 1. Needlet coefficients of a Gaussian random realization for some needlet scale j with scale parameter B=2, the corresponding multipole coverage is: $\ell = [1, 4]$ for $j = 1$, $\ell = [4, 16]$ for $j = 3$, $\ell = [16, 64]$ for $j = 5$, $\ell = [32, 128]$ for $j = 6$.

larger part, from coherent sources concentrated around the galactic plane and in few large structures that extend at higher galactic latitudes. As we saw in the previous sections, a space-frequency representation of coherent signals naturally tends to be sparse. We will thus expect that the contribution from the galactic foreground will be concentrated in few large coefficients that can be identified and fitted with a needlet thresholding technique. On the other hand, CMB has very different features, since it is a random isotropic field and not a coherent signal. Thus, unlike foregrounds, a needlet-space representation of the CMB signal will not be sparse. The reason is that the CMB mostly does not form coherent structures, but it is instead a homogeneous fluctuations field at all scales.

The different behaviour of the two components can be immediately appreciated by looking at the needlet decomposition of a CMB realization and a foreground template. We show in figure 1 the needlet decomposition of a random CMB realization: the signal is spread over all the coefficients, as expected, given its stochastic nature. Furthermore, it is easy to notice that only adjacent layers show some level of correlation. Needlets split a continuous field in several independent realizations, the layers, each one covering a limited range of frequencies. Figure 2 shows instead the corresponding decomposition of a thermal dust template. We see how the information is actually concentrated only in few coefficients located near the galactic plane. The lowest layers trace the diffuse emission while the higher frequency levels contain only few small scale corrections. Moreover, all scales are highly correlated at any distance in frequency level, and this is expected since the signal is coherent.

To get further insight into how the thresholding procedure operates in a component separation context, we now illustrate and justify it within a general Bayesian framework. As shown in e.g., [33], a Bayesian approach provides a way to describe different component
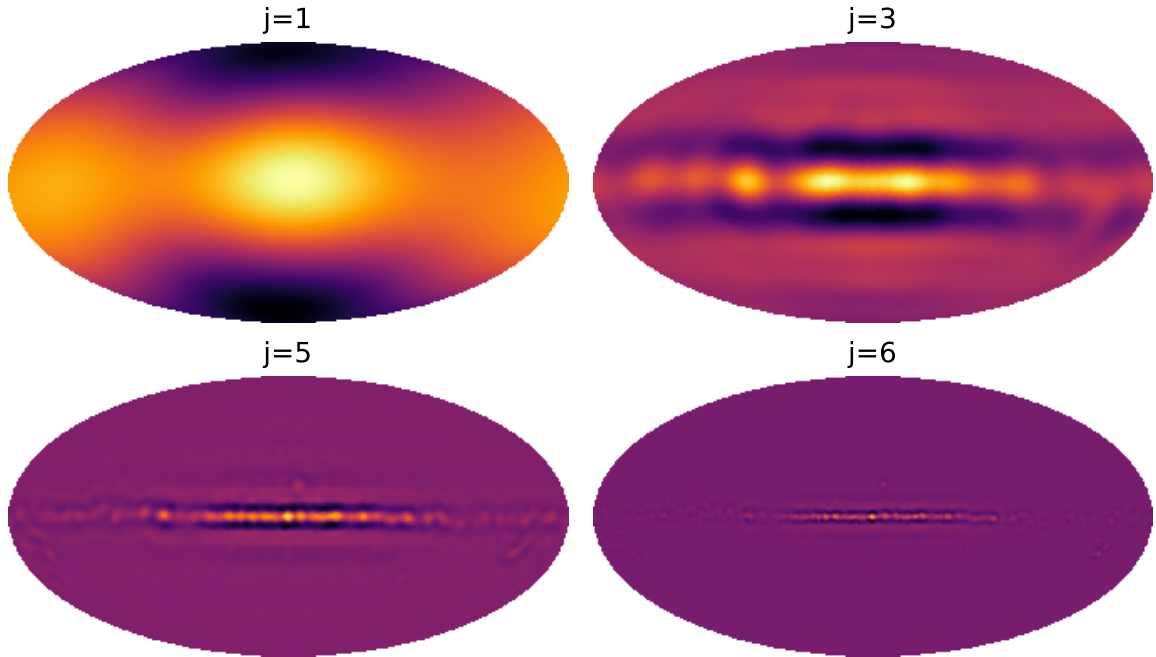
Figure 2. Needlet coefficients of thermal dust template for some needlet scale j with scale parameter B=2, the corresponding multipole coverage is: $\ell = [1, 4]$ for $j = 1$, $\ell = [4, 16]$ for $j = 3$, $\ell = [16, 64]$ for $j = 5$, $\ell = [32, 128]$ for $j = 6$.

separation techniques within a unified, general formalism. This approach shows in particular how different common component separation methods amount to different choices of priors and marginalized parameters.

We start as usual by assuming to have observations from K channels, with a mixture of N components and M elements (pixels) in each channel (by "pixel" we mean real space pixels, $a_{\ell m}$, needlet coefficients or the elements of whichever basis is adopted to represent the signal). We recall then the linear mixture model:

$$d_i = As_i + n_i, \tag{3.1}$$

where d_i is a vector of K elements representing the observations in a given pixel i, s_i is a vector of N elements representing the contribution of the components in the same pixel, A is the *mixing matrix* of dimension $N \times K$, which weighs the contributions of different components at different frequencies; finally, n_i is the noise in the pixel i.

The Bayesian formulation of the component separation problem aims to solve:

$$P(A, s|d) \propto \mathcal{L}(d|A, s)P(s), \tag{3.2}$$

as shown e.g., in [33]. With specific assumptions on the priors and eventually variables to marginalize over, the formulation above can be used to define typically adopted component separation techniques, such as ILC and SMICA [16, 34]. In our case, we want to introduce the sparsity assumption on the foreground templates. A similar hypothesis is also at the basis of the development of other algorithms as, for example, GMCA [20]. We will soon clarify the differences between the approach discussed here and GMCA.

As noticed in the previous section, the usual way to enforce sparsity in a Bayesian context is to assume leptokurtic priors, with the Laplace distribution as a common choice. We thus want to study the implementation of this kind of prior in a Bayesian component separation framework.

First, we must rewrite the linear mixture model in a form more suited to our scope, which is ultimately that of recovering the CMB signal. We assume that the foreground signal is sparse in the needlet domain, but the CMB is not, this is the main difference with the GMCA algorithm, which instead implicitly assumes sparsity also for the CMB component. In more detail, GMCA aims at the reconstruction of the full mixing matrix A and the signal s by exploiting the morphological diversity of the components in a given basis [35, 36]. In this work, instead, our purpose is to isolate the cosmological stochastic signal from the other emissions, avoiding the full reconstruction of the components.

We thus rewrite the linear mixture model assuming that components with a "Gaussian" and a "non-Gaussian" prior probability coexist in the data. Besides noise, we assume that the only "Gaussian" component is the CMB; thus we have, in the single pixel:

$$d_i = As_i + \mathbf{e}c_i + n_i = f_i + \mathbf{e}c_i + n_i, \tag{3.3}$$

where we explicitly separate the CMB from the other components, by denoting it with c_i, times the vector of ones $\mathbf{e}$ of length K (since the CMB signal is constant between channels). Furthermore $f_i = A's_i$, where A' is the mixing matrix with the row corresponding to the CMB set to zero. We rewrite the model in this way because at this stage we are not interested in the mixing matrix A. Therefore instead of explicitly estimating the underlying N templates, we consider the linear combination, f_i, of all of them, in each of the K channels. We assume that a Laplace distribution is a proper prior also for this combination.

We now split the problem as follows:

$$\begin{cases} P(f,c|d) = P(c|f,d)P(f|d), \\ P(f|d) \propto P(f) \int_\infty \mathrm{d}c\ \mathcal{L}(d|c,f)P(c), \end{cases} \tag{3.4}$$

$$\rightarrow P(f,c|d) \propto P(c|f,d)P(f) \int_\infty \mathrm{d}c\ \mathcal{L}(d|c,f)P(c). \tag{3.5}$$

We can then solve the problem of recovering the CMB component in two steps. We first find the value $\hat{f}$ which maximize 3.4, followed by replacing such value in 3.5 and maximizing again.

Since the noise component is uncorrelated between channels and since in our approach we are not interested in recovering explicitly the full mixing matrix, we can assume we will repeat our thresholding procedure independently at each frequency. Therefore, in our present derivation, we treat d, f and c as single-channel maps. The data likelihood takes the usual form (Gaussian noise):

$$\mathcal{L}(d|c,f) = N(d; c+f, C_n) \propto \exp\left[-\frac{1}{2}(d-f-c)^{\mathbf{T}}C_n^{-1}(d-f-c)\right] \tag{3.6}$$

$$P(c) = N(c; 0, C_c) \propto \exp\left[-\frac{1}{2}c^{\mathbf{T}}C_c^{-1}c\right], \tag{3.7}$$

where C_n and C_c are respectively the noise and CMB covariance matrices. Therefore we can write:

$$\mathcal{L}(d|c,f)P(c) \propto \exp\left[-\frac{1}{2}d^{\mathbf{T}}C_n^{-1}d + f^{\mathbf{T}}C_n^{-1}d - \frac{1}{2}f^{\mathbf{T}}C_n^{-1}f\right] \times$$

$$\times \exp\left[-\frac{1}{2}c^{\mathbf{T}}\left(C_n^{-1} + C_c^{-1}\right)c + c^{\mathbf{T}}\left(C_n^{-1}d - C_n^{-1}f\right)\right], \tag{3.8}$$

Marginalization over c can be carried out via standard Gaussian integration. We remind:

$$\int \exp\left[-\frac{1}{2}\vec{x}^T\mathbf{R}\vec{x} + \vec{B}^T\vec{x}\right]d^n x = \sqrt{\frac{(2\pi)^n}{\det R}}\exp\left[\frac{1}{2}\vec{B}^T\mathbf{R}^{-1}\vec{B}\right], \tag{3.9}$$

to obtain:

$$\mathcal{L}(d|f) = \int_\infty dc\,\mathcal{L}(d|f,c)P(c) \propto \exp\left[-\frac{1}{2}d^{\mathbf{T}}C_n^{-1}d + f^{\mathbf{T}}C_n^{-1}d - \frac{1}{2}f^{\mathbf{T}}C_n^{-1}f\right] \times$$

$$\times \exp\left[\frac{1}{2}\left(C_n^{-1}d - C_n^{-1}f\right)^{\mathbf{T}}\left(C_n^{-1} + C_c^{-1}\right)^{-1}\left(C_n^{-1}d - C_n^{-1}f\right)\right]. \tag{3.10}$$

We then implement the Laplacian prior, again assuming uncorrelation between the channels. Thus we have $P(f) \propto \exp -\lambda||f||$ where $||*||$ is the L1 norm. We add this to (3.10) define $R \equiv \left(C_n^{-1} + C_c^{-1}\right)$ for simplicity of notation. Finally, we differentiate posterior with respect to the foreground component, to find:

$$\partial_f(-\log(P(f,d))) = -C_n^{-1}d + C_n^{-1}f + C_n^{-1}R^{-1}C_n^{-1}d - C_n^{-1}R^{-1}C_n^{-1}f + \lambda\partial_f||f||, \quad (3.11)$$

which implies:

$$\hat{f} = d - \left(C_n^{-1} - C_n^{-1}R^{-1}C_n^{-1}\right)^{-1}\lambda\partial_f||f||. \tag{3.12}$$

As derived in section 2.1, the solution to this problem is the soft thresholding operator, with threshold $\left(C_n^{-1} - C_n^{-1}R^{-1}C_n^{-1}\right)^{-1}\lambda$. Following the above derivation, the central idea of our study is therefore that of using thresholding as a preliminary tool for foreground cleaning or foreground template reconstruction, in single channels. This captures morphological information in the foreground spatial distribution, at any fixed frequency. Channels can then be combined and the overall cleaning procedure further refined by applying standard algorithms that exploit CMB and foreground spectral properties. To this purpose, in the following, we combine thresholding with template fitting and Internal Linear Combination. Before concluding this section, let us stress here that our thresholding operator is applied to reconstruct foregrounds and not the CMB component. Reconstructing the CMB via thresholding in specific representations could break isotropy and Gaussianity, and we explicitly avoid this potential issue with our approach [37, 38].

3.1 Template Fitting

Template fitting provides estimates of the amplitudes of each component from the fit of known templates to the data of interest. The results are then subtracted from the data to remove the spurious signal.

In the linear mixture model (3.1), the distribution of the components over the data is stored in the matrix s. The templates used should then reproduce the elements of s, other than the CMB, or a linear combination thereof. Assuming to have the exact templates, the

linear fit would provide the entries of the mixing matrix A corresponding to the given source and channel. The result for a collection of N_{temp} known templates T fitted to a map y, is the standard linear regression solution. Calling $\boldsymbol{\alpha}$ the vectors of estimated amplitudes we have:

$$\boldsymbol{\alpha} = \left(T^{\mathrm{T}}C^{-1}T\right)^{-1}\left(T^{\mathrm{T}}C^{-1}y\right), \tag{3.13}$$

where C is the $N_{\text{pix}} \times N_{\text{pix}}$ covariance matrix of the map, which depends on the noise and the CMB, and T is the $N_{\text{pix}} \times N_{\text{temp}}$ matrix representing the template. The estimation and the inversion of C is a major limitation in template fitting since, given the high number of data points collected by modern surveys, it is very large and computationally expensive.

The choice of templates is, obviously, the crucial part of any template fitting technique. A possibility is that of resorting to external templates from previous experiments or theoretical modeling. This approach requires a lot of a-priori information on the emissions of interest, which can be unavailable or unreliable. Relying on external data-sets also runs into the issue of having to deal with additional systematic effects, cross-calibration problems and so on. For this reason, most of the modern CMB surveys have been designed with several channels at foreground dominated frequencies, allowing us to track spurious contaminant components without having to rely on external information. These templates obtained directly from the data are called internal templates. The most straightforward approach would be just to use these foreground dominated maps as templates, for example, a high-frequency channel as a dust template and a low frequency as synchrotron. However, these maps contain also the CMB contribution, which would be removed together with the contaminants, so that a correction factor must be introduced in formula (3.13). Another widely implemented solution is thus to build linear combinations of different channels so that the constant component (the CMB, providing that the observations are calibrated to its black body spectrum) vanishes [21, 39]. These combinations usually are computed as the subtraction of adjacent frequency channels. The major drawback of this approach is that the noise in the internal templates is enhanced compared with the single channel. *A natural application of the thresholding techniques just described is thus the denoising of the internal difference templates.*

In other words, our goal is to clean a "target" map with a given number of noisy foreground templates; a needlet thresholding algorithm is then used to obtain optimal templates in which the noise component is minimized while preserving the resolution of the starting template. In the next section, we will discuss in more detail the "optimality" of the thresholding solution for this purpose, and we will show a comparison with other viable estimators. In our procedure, these templates are then fitted to the target channel, and the fitting coefficients are then used to combine the *original* templates. The map obtained is an estimate of the foreground contribution in the target channel, which is then cleaned by subtracting it. We start by decomposing the data (the target channel and the internal templates) in needlet coefficients. For each needlet scale j, the templates are then thresholded with threshold λ_i and fitted to the data to obtain the amplitude coefficients α_i (where i runs over the different templates). The optimal thresholds are selected in recursive way to maximize the goodness of fit of the templates with the target channel; namely, we iteratively threshold and fit the internal templates so that:

$$\chi^2 = \sum_k \frac{\left(\beta_{jk}^{map} - \sum_i \alpha_i \beta_{jk}^{T_i}(\lambda_i)\right)^2}{\sigma_{jk}^2}, \tag{3.14}$$

is minimized. Here, β_{jk}^{map} represents the target channel at the needlet scale j (With k running over pixels), $\beta_{jk}^{T_i}(\lambda_i)$ are the templates thresholded with threshold λ_i and the coefficients α_i are obtained with the standard template fitting solution. We will discuss other threshold selection method in the next section.

The templates are thus linearly combined with weights α_i and subtracted to the target map as a standard template fitting procedure. We stress the fact that the cleaned template are used only to obtain the fitting coefficient, while the full templates are used to clean the map. This allows us to preserve the linearity of the template fitting procedure. In section 4.3.1, we will show the results of the application of this algorithm to different simulated data-sets.

3.2   ILC

As a further case study and an example of the versatility of our approach, we merge our thresholding algorithm with an Internal Linear Combination cleaning procedure. The general idea is as follows: a thresholding algorithm is "inverted" to remove from the map the most contaminated coefficients. This is followed by combining the channels, following the usual ILC prescription. Since we work in needlet space, such method can be straightforwardly implemented in needlet-based ILC pipelines, such as NILC. However, this is by no means mandatory, and any other signal representation domain can be chosen in the ILC step.

The overall rationale of the approach is as usual that thresholding exploits complementary foreground information, compared to ILC, since it allows us to minimize the foreground contribution in *single channels*, on the basis of spatial – rather than spectral – features. Merging the two methods could therefore in principle lead to useful improvements, especially in experiments where a limited number of frequency channels are available. In the section 4.3.2 we will show the results of the application of our ILC with thresholding procedure, using simulated data-set and comparing it with the standard approach.

The algorithm developed in these tests is structured as follows. First of all, each channel map is decomposed in needlet coefficients, and we treat different scales separately. In other words, we exploit the flexibility of the needlet representation to identify and remove large, spurious coefficients scale by scale. Note that, to avoid distortions in the spectral energy distribution of foregrounds, which would compromise the ILC step, in this analysis the masked coefficients after thresholding must be *the same in each channel*. In practice, we construct an initial linear combination of the channels (for example with a preliminary j ILC) to identify the foreground dominated regions as the isotropic residuals in this co-added map. We then go back to single-channel maps and remove these regions at each frequency. The criterion to choose the threshold is again recursive and based on the minimization of the anisotropy of the residual coefficients. More specifically, in the thresholding step, we minimize:

$$\Delta_j = \frac{1}{N_j} \sum_k \left[\frac{\left(\beta_{jk} - \beta_{jk}^{T_{(\lambda)}}\right)^2}{\sigma_j^2} - 1 \right]^2 , \tag{3.15}$$

where, as usual, j set the scale, β_{jk} represents the map in needlet space and $\beta_{jk}^{T\lambda}$ is the map thresholded with threshold λ, k is the position index (the HEALPix pixels, in our algorithm), N_j is the number of coefficients at the given layer j and the variance σ_j can be estimated from the coefficients themself. After this "pre-cleaning" via thresholding, the different channels are then combined again with an ILC algorithm to produce a final CMB map. The overall method can be essentially interpreted as a needlet space masking, where the masked area varies with

the scale, followed by Internal Linear Combination. At the end of this procedure, we extract the power spectrum from the cleaned maps, correcting for the missing coefficients with the standard MASTER technique [40]. The final power spectrum is used as a figure of merit to assess the performance of both our template fitting and ILC algorithms. The performance of our algorithms is discussed in the next section.

4 Results

The initial goal of our analysis is that of verifying how well we can reconstruct noisy foreground templates at high resolution, using needlet thresholding. We then focus on the issue of combining our thresholding procedure with standard component separation methods - namely ILC and template fitting - with the aim of improving their performance. However, before moving to these points, we start by quantitatively checking the foreground sparsity assumption, on which the entire procedure is based.

4.1 Measuring sparsity

Our main goal is to separate the CMB and foreground components, by exploiting the sparsity of the latter. It is, therefore, useful to start our analysis by defining a metric to quantify sparsity and verify in detail whether and to which extent this assumption holds in our context of interest.

It has been proven that a good measure of sparsity is the so-called Gini index [41, 42], defined as:

$$G_P = 1 - 2 \int_0^1 \mathrm{d}C(x) \; \frac{\int_0^x \mathrm{d}t \; tP(t)}{\int_0^\infty \mathrm{d}t \; tP(t)}, \tag{4.1}$$

where $P(x)$ is a positive valued probability distribution, with cumulative distribution $C(x)$.

For an ordered data set, $d = \{[d_1, ..., d_i, ..., d_N] : d_i < d_{i+1} \forall \; i < N\}$, the Gini index can be estimated with the formula:

$$\hat{G}(d) = 1 - 2 \sum_{i=1}^N \frac{d_i}{\sum_{k=1}^N d_k} \left(\frac{N - i + 0.5}{N} \right). \tag{4.2}$$

Originally, the Gini index was introduced in social economics to measure the degree of inequality of the income distribution of a population. It runs from 0 (perfect equality, every person has the same wealth) to 1 (perfect inequality, one person owns all the goods) [43]. It is immediate to notice that the notion of inequality as defined above corresponds to the notion of sparsity in signal processing. To check the sparsity assumption, we thus measure the Gini index of different foreground templates and we compare it to a Gaussian realization. Since the Gini index is defined for positive valued distributions, we use the square of the needlet coefficients β_{jk}.

Table 1 shows the results for intensity templates of different components, at different scales j, setting the needlet scaling parameter $B = 2$ (we remind that each scale j cover the multipole interval $[B^{j-1}, B^{j+1}]$, and the corresponding window function peaks at B^j). We report the values computed both on the full sky and outside the galactic plane, with a galactic mask covering the 20% of the sky. The first row refers to a random CMB realization. The following rows refer instead the principal sources of foreground emission. Since they roughly follow the shape of the Galaxy, the results are similar between components and are less sparse outside the galactic plane, where the diffuse emission dominates. We generate the templates

j	2	3	4	5	6	7	8	9	10	11
ℓ_{peak}	4	8	16	32	64	128	256	512	1024	2048
CMB	0.669	0.652	0.633	0.638	0.637	0.637	0.637	0.637	0.637	0.637
dust	0.599	0.788	0.906	0.957	0.982	0.991	0.993	0.995	0.997	0.997
$f_{sky} = 0.8$	0.692	0.693	0.809	0.886	0.938	0.969	0.985	0.993	0.997	0.997
syncrothron	0.688	0.844	0.933	0.969	0.986	0.988	0.979	0.989	0.994	0.995
$f_{sky} = 0.8$	0.719	0.754	0.812	0.870	0.895	0.873	0.844	0.842	0.849	0.868
AME	0.631	0.823	0.940	0.988	0.998	1.000	1.000	1.000	1.000	1.000
$f_{sky} = 0.8$	0.687	0.694	0.830	0.918	0.954	0.977	0.995	0.999	0.999	0.997
free-free	0.739	0.868	0.941	0.974	0.989	0.995	0.997	0.997	0.997	0.998
$f_{sky} = 0.8$	0.711	0.852	0.935	0.969	0.983	0.991	0.995	0.997	0.997	0.995

Table 1. Gini index of the square of needlet coefficients β_{jk}^2 of various components computed scale by scale, both full-sky and outside the galactic plane ($f_{sky} = 0.8$). Values approaching 1 correspond to a more sparse representation. 0.637 is the expected value for the χ^2 distribution.

with the PySM software, based on the Planck sky model [44, 45].

If we start by looking at the Gini index of the square of the CMB coefficients, we notice that it remains substantially constant between scales and it converges to 0.637 for high j. This is the expected value for the χ^2 distribution, *i.e.* the distribution of the squares of a Gaussian random variable. This behaviour reflects what stated in section 2, namely that the needlet coefficients are asymptotically uncorrelated/independent at higher frequencies. If we now focus on the foreground components, the first thing we see is that lower frequencies display lower values of the Gini index. This is due to the fact that lower frequencies represent diffuse emission, which is not sparse. Layers corresponding to higher multipoles are instead significantly more sparse, with a high degree of sparsity achieved at $j \geq 3$ or $\ell \gtrsim 10$ in the corresponding harmonic representation. This reflects the fact that, at smaller scales, we track single structures rather than a homogeneous fluctuations field. The small scale portion of the signal can thus be recovered keeping only the coefficients corresponding exactly to the size and positions of these structures, which are few compared to the total number of cubature points.

In light of these observations, we, therefore, expect our thresholding-based methods to achieve significantly better performance at $\ell \gtrsim 10$.

Let us point out that the fact that very large scales are not sparse does not refute in any way the overall sparsity assumption. This is of course due to the fact that the cardinality of the needlet coefficients at a given frequency scales as B^{2j}, thus the number of low j β_{jk} is negligible compared to the total. Summarizing, we have just verified that foreground templates can be faithfully represented taking almost all the (few) large scale needlet coefficients and a minimal portion of the high-frequency ones, selected with a thresholding algorithm. This highlights how thresholding can also achieve a remarkable level of data compression.

4.2 Template reconstruction

After this preliminary investigation, we now move on to show the actual performance of thresholding on the recovery of a signal template from noisy data. In this analysis, we use as benchmark a map obtained from a foreground template at 200 GHz and an isotropic Gaussian noise realization, with HEALPix nside 512. In order to highlight the denoising properties of thresholding, we set a very high noise level, so that a large portion of the available scales are
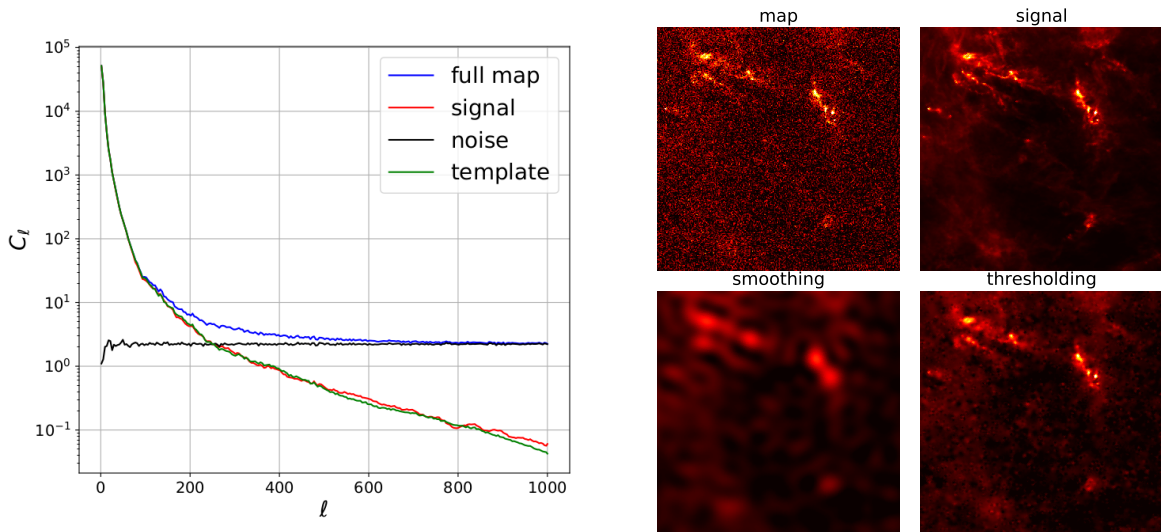
Figure 3. Left Panel: Power spectra of the input map (blue), of the noise (black), of the signal (red) and of the template recovered via thresholding. Right panel: a $30° \times 30°$ patch extracted from the input map. Clockwise from the top left: full input foreground signal + noise, input foreground signal only, reconstructed foreground signal after thresholding, patch smoothed at signal dominated resolution.

noise dominated. We run a soft thresholding algorithm on this synthetic map: the threshold is selected following the method introduced in [46], based on the minimization of the Stein Unbiased Estimate of Risk (SURE), defined as, given a map x and a thresholding operator $ST_\lambda(x)$:

$$\text{SURE}(x) = N\sigma^2 + ||ST_\lambda(x) - x||^2 + 2\sigma^2 \sum_{i=1}^{N} \frac{\partial}{\partial x_i}(ST_\lambda(x) - x)_i, \tag{4.3}$$

where N is the dimension of the map x, i runs over the pixels indices and, in the specific case of soft thresholding, the last summation corresponds simply to minus the number of thresholded coefficients.

As expected, thresholding is very effective in recovering the power of the input signal, deep into the noise dominated region. This is shown in the left panel of figure 3, where the angular power spectrum of the recovered template (in green) follows faithfully the one of the input signal (in red) at all scales and well below the noise level (in black). The excellent performance of thresholding is directly related to what discussed in the previous section. At high multipoles, corresponding to the noise dominated region in this example, the signal power is concentrated in a tiny number of very large β_{jk}, whereas the noise power is spread among all the coefficients. In the right panel, we show the real space reconstruction of signal structures, compared to the noisy map and the input signal. We also show the map smoothed at the signal-dominated scale. The comparison with the thresholding results shows clearly how the latter removes a large part of the noise while also preserving the smallest structures. In summary, this test shows that thresholding out small coefficients produces an almost complete suppression of the noise with only marginal loss of the signal power while maintaining the full resolution of the starting template. Furthermore, as said before, the dimension of the data set is greatly reduced: the reconstruction presented here use only 7% of the total
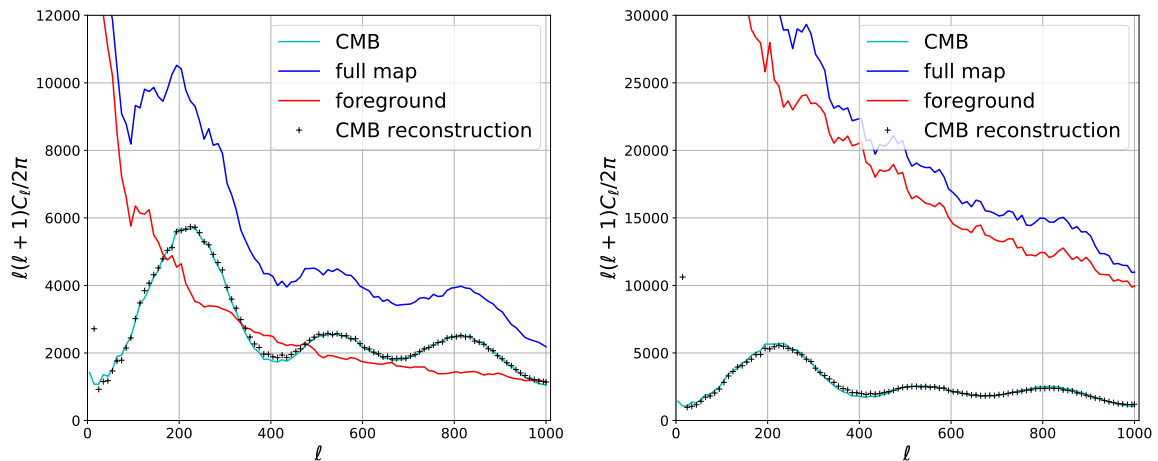
Figure 4. Power spectra of the full map (blue) of the foreground (red), of the input CMB signal (cyan) and of the CMB reconstruction from the residuals of thresholding (black cross) for simulations at 143 GHz (left panel) and 217 GHz (right panel).

number of coefficients.

Moreover, the denoising of a foreground template is not the only applications of these methods. Needlet properties also allow us to separate the coherent (foreground) and stochastic (CMB, noise) components. Since we have just reconstructed a foreground template, using only 7% of the map coefficients, it is natural to assume that the residual coefficients provide a reconstruction of the stochastic component, being it the noise (as in our previous example) or the CMB. To check for this, we build a map from a foreground template and a CMB realization. We are interested in the recovery of specific features of the CMB power spectrum, *e.g.* the acoustic peaks, thus, for the purpose of this test, we do not need to include noise in the map. We show the results of this analysis in figure 4, considering two realizations at 143 and 217 GHz respectively. In this case, we used a hard thresholding algorithm implemented with a threshold selection criterion based on the minimization of the difference between the Gini index of the residuals with the expected value for a Gaussian field. In both cases we remove a small part of the galactic plane ($f_{sky} = 90\%$). Note that the size of the mask does not affect the results of thresholding. We see that the power spectrum of the residuals (black cross) remarkably follows the CMB one (cyan line) for $\ell \gtrsim 20$ irrespective of the relative amplitude of the foreground component (red line). On the other hand, at low multipoles the reconstruction completely fails. As we commented before, these multipoles represent the diffuse signal, that is not sparse and cannot be separated from the stochastic background with this technique.

As a final interesting result we show that, in consequence of the sparsity of foregrounds, thresholding (and in particular soft thresholding) dominates all linear estimators in terms of mean square error, when we attempt to reconstruct a foreground template from a noisy realization. The problem we have dealt with so far is that of estimating a foreground template f from a single noisy observation $d = f + n$, where the noise is a Gaussian realization $n = N(0, \mathbf{I}\sigma)$, so that $d = N(f, \mathbf{I}\sigma)$. The usual least square solution would be simply $\hat{f} = d$. In contexts of this type, it however proven that the asymptotically optimal estimator is not least square but it is the so-called James-Stein (JS) estimator. More specifically, if we focus

method		smoothing	James-Stein			Thresholding	
			real space	needlets	harmonic	hard	soft
rms ($\sigma_n = 27.10$mK)		$0.020\sigma_n$	$0.026\sigma_n$	$0.018\sigma_n$	$0.023\sigma_n$	$0.020\sigma_n$	$0.016\sigma_n$
rms ($\sigma_n = 0.75$mK)		$0.295\sigma_n$	$0.735\sigma_n$	$0.259\sigma_n$	$0.259\sigma_n$	$0.189\sigma_n$	$0.170\sigma_n$
rms ($\sigma_n = 0.28$mK)		$0.493\sigma_n$	$0.925\sigma_n$	$0.429\sigma_n$	$0.426\sigma_n$	$0.315\sigma_n$	$0.286\sigma_n$

Table 2. Root mean square errors of the real space template in unit of the noise standard deviation per pixel. The noise levels are set so that the power spectrum match the signal one at $\ell = 20, 250, 600$ (top-down).

on linear estimators, and the dimension of the data-set is greater than 3, the JS estimator is known to provide be the lowest mean square error, always dominating the least square solution [47, 48]. We will therefore focus here on this class of estimators to set our performance benchmark, which we will use to assess thresholding results later on.

Following the notation just defined, the JS estimator of the template f is defined as:

$$\hat{f}^{JS} = \left(1 - \frac{(N-2)\sigma^2}{\sum_{i=1}^{N} d_i^2} \right)^+ d \qquad (4.4)$$

where i runs over the mode indices in the chosen representation, N is the dimension of the data-set and the apex + stand for the positive part. We develop a simple implementation of this estimator and test it in needlet, harmonic and real space. In real space, the application is straightforward: d is the map, σ is the noise standard deviation, and N is the number of pixels. For needlet (harmonic) space we apply the estimator adaptively, scale by scale and multipole by multipole, so that – for each scale j' (multipole ℓ) – d in formula 4.4 corresponds to $\beta_{j'k}$ ($a_{\ell'm}$), while σ^2 corresponds to the needlet noise variance $\sigma_{j'}^2$ (noise power spectrum $C_{\ell'}$). A general implementation would require the computation of the full noise covariance. However, in our idealized situation, where the noise is white and uncorrelated both in space and frequency, the solution we provide here is exact. Note that, under the assumption that signal and noise are independent Gaussian random fields, the harmonic James-Stein estimator just described would be formally analogous to the Wiener filter solution. Basically, in its scale/multipole dependent implementation, this estimator suppresses the noise dominated scales.

As mentioned above, after applying the JS estimator to our case of interest, we compare its performance to thresholding. In particular, we compare two thresholding algorithms: a soft thresholding where the threshold is selected minimizing SURE, and a hard thresholding based on the so-called universal threshold defined as:

$$\lambda_j = \sigma_j \sqrt{2 \log N}, \qquad (4.5)$$

where σ_j is the noise variance at the scale j and N in the number of needlet coefficients. We find that, in the case of sparse signals, thresholding dominates (*i.e.* provides lower mean square error) even this estimator (and therefore any other linear estimator). We show this in Table 2, where we list the root mean square errors of the real space template with respect to the input signal, in units of σ_{noise} for different estimators and noise levels for a 200GHz foreground template. The central row corresponds to the configuration used in the analysis shown in figure 3. As said earlier, we compare the results obtained with thresholding with the JS estimates in real, harmonic and needlet space, and with a simple smoothing at signal dominated scales (respectively $\ell = 20, 250, 600$ for the three configurations top-down).

Channel	140	220	240
Beam (arcmin)	110	110	110
σ (μK_{CMB}arcmin)	30	40	80

Table 3. SWIPE specifications

Besides optimality issues, another advantage of needlet methods is that they are completely blind (under the assumption of white noise), whereas the other approaches assume knowledge of either the noise Power Spectrum or σ_{pixel}. For the smoothing case, also some knowledge about the signal Power Spectrum is necessary, in order to set the scale of the low-pass filter. On the contrary, in the needlet methods, the noise standard deviation (that appear both in SURE and in the JS shrinking factor) is computed directly from the data using the median absolute deviation (MAD) [1] of the highest frequency layer, knowing that this is noise dominated. As expected, thresholding always provides the lowest rms in these tests. In particular, soft thresholding with SURE based threshold selection achieves the best results in all cases. The hard thresholding estimator, based on the universal, threshold also provides good results, being slightly outperformed by the needlet JS estimator only in the noisiest case. The improvement is always higher for higher level of noise for all, while the differences between them are more pronounced for lower noise.

Summarizing, our analyses so far have shown that thresholding achieves the best results in the extraction of foreground templates from noisy realizations. In the next section, we discuss some examples of applications of these techniques, in synergy with other component separation methods.

4.3 Synergy with other methods

In this section, we present the results obtained by applying the techniques described in the previous sections to different simulated data-sets. We generate two mock datasets using the PySM software [45], which mimic respectively observations of the *Planck* mission and of the SWIPE instrument of the forthcoming LSPE mission [49–51].

LSPE (Large Scale Polarization Explorer) is a forthcoming ASI mission aimed at the measurement of large scale CMB polarization fluctuations. It will scan a large region of the sky (20%), with two instruments, SWIPE and STRIP. In this work, we concentrate on the former, SWIPE (Short Wavelength Instrument for the Polarization Explorer). SWIPE will measure CMB polarization in three frequency channels from a stratospheric balloon flying long-duration in the northern polar region during the winter night.

Due to its configurations, SWIPE represents a good benchmark for our methods: since it observes in just three different frequency channels, needlet thresholding can in principle significantly improve the results of blind component separation techniques as internal templates fitting and ILC. Planck simulations, instead, are useful to test the performance of the algorithms on the current state of the art CMB maps and, more in general, on full-sky data-sets with significant frequency coverage.

4.3.1 Template Fitting

In this section, we test the performance of our internal template fitting pipeline, equipped with the needlet thresholding method described before. Our aim here is not that of present-

---

[1]The MAD times a given factor provides a robust estimator of the standard deviation in presence of outliers.
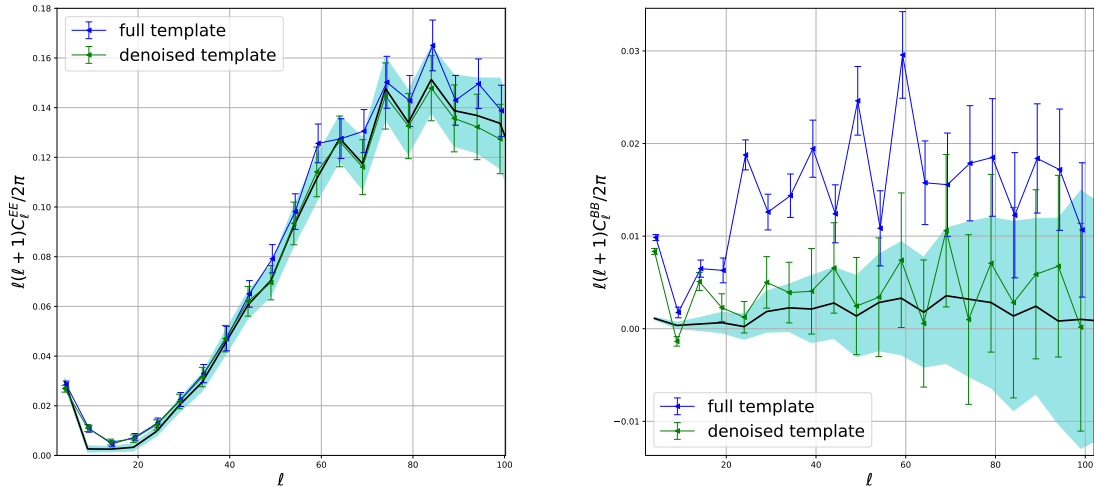
Figure 5. Power spectrum reconstruction with a template fitting algorithm in needlet space – with and without thresholding – for SWIPE simulations, corrected for the average noise contribution. The black line represents the average of the power spectra of the maps and the shaded blue area is its standard error. Green and blue lines represent the mean of the power spectra of the cleaned maps with and without thresholding respectively. The error bars are the respective standard errors.

ing a full component separation pipeline, but rather that of quantifying the impact of the application of thresholding techniques on simple template fitting procedures.

As a first test, we produce a set of 100 simulations of the three SWIPE channels, assuming the Planck cosmology and tensor-scalar ratio $r = 0.1$, while the foregrounds are generated following the Planck sky model (see [45] for reference). Given the coarse angular resolution of the experiment under exam, simulations with HEALPix nside=128 are sufficient. The noise is assumed to be white, Gaussian and isotropic.

With just three frequency bands, only one internal template can be used. This regime provides therefore an ideal benchmark to verify the improvement from the pre-processing of the templates with our denoising algorithm. The internal template is obtained from the difference between the 240 GHz and the 220 GHz channels, and it is used to clean the 140 GHz cosmological channel. As a probe of the flexibility of this technique, we implement two different pipelines. The former strictly follows the steps described in section 3. The latter still uses the same procedure to obtain the thresholded template, but the final fit is now performed in real space.

The overall approach to assess the performance of our method is quite straightforward: we decompose each set of polarization maps in E and B modes, followed by applying our template fitting algorithms with and without thresholding. We then measure EE and BB power spectra in all cases and use them as figure-of-merit, by comparing cleaned spectra with the input ones.

We show the results of this analysis in figure 5 and 6. We find that, for both techniques (needlet and real space fit), pre-cleaning the templates with the needlet thresholding provide noticeable improvements. Especially in the case of B-mode reconstruction, where the noise dominate the templates, we found that preliminary denoising allows us to successfully recover
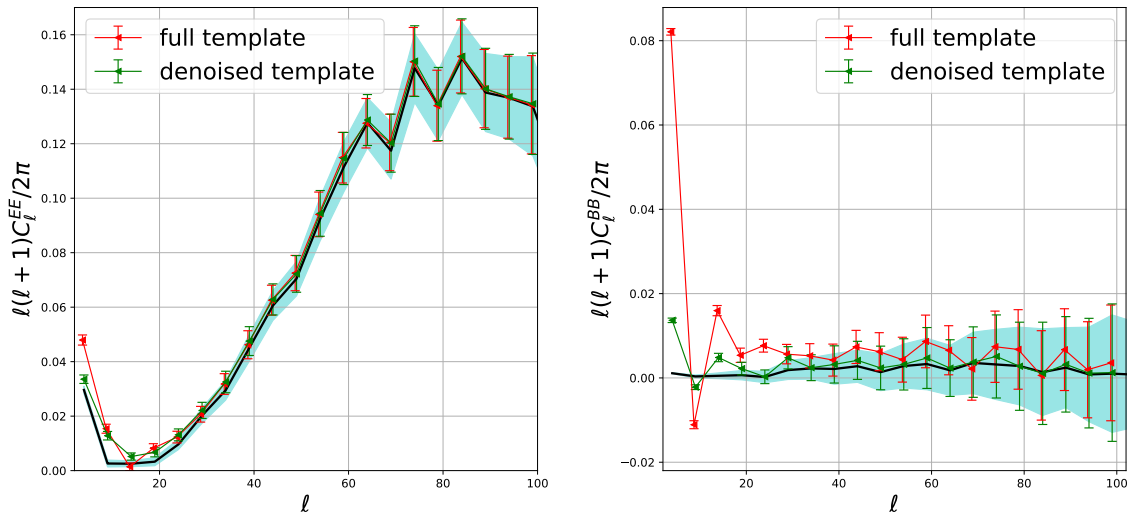
Figure 6. Power spectrum reconstruction using a template fitting algorithm in real space – with and without thresholding – for SWIPE simulations, corrected for the average noise contribution. The black line represents the average of the maps' power spectra and the shaded blue area is its standard error. Green and the red lines represent the mean of the power spectra of the cleaned maps with and without thresholding respectively; the error bars are the respective standard errors.

the input power spectrum in a large portion of the multipole space under examination. We repeat the same kind of analysis on simulations of Planck data. Planck, with 9 frequency channels, has a much larger frequency coverage than SWIPE, therefore we expect the effects of template thresholding to be less relevant in this case. Given the preliminary level of this investigation, we do not produce full resolution simulations but we limit the maps to nside=256 and we restrict the analysis to lmax=500. Since we are looking at diffuse foregrounds on large scales, this does not alter significantly our final performance assessment. For consistency, all the template are smoothed to match the Planck channel with lowest resolution, i.e. the 30 GHz channel with a 30 arcmin beam. Our internal templates are built following the SEVEM pipeline described in [13]. In figure 7 we show results for the cleaning of the 147 GHz channel. Following the Planck team, we use as internal templates the difference between the channels: (30-44)GHz, (217-100)GHz, (353-217)GHz. The first traces the synchrotron, while the last two will trace the dust emission.

Our results are obtained on a set of 50 simulation; as expected in this case, the impact of thresholding is much less relevant than in the LSPE case. However, some improvement can still be noticed in the B-mode spectrum reconstruction, where the signal to noise is very low. Let us note here that, in principle, other denoising techniques can be applied to this problem, with a good performance in terms of the mean square error. For example, if we focus on the B-mode maps, a simple smoothing of the template at very low resolution can still produce an accurate fit. This is because - using the foreground model of our simulations (based on PySM)- the result is mostly driven by the first few multipoles, where the foreground amplitude is well above the template noise. However, even in cases where the error improvement by using thresholding is only marginal, we stress the fact that soft thresholding does maintain a number of important advantages. The main advantage is that thresholding preserves the
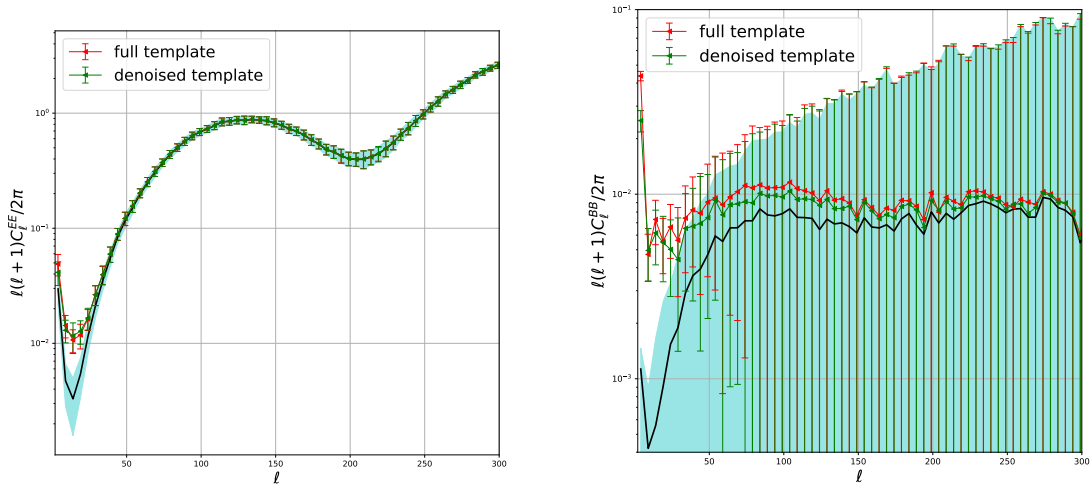
Figure 7. Same results as in figure 5, but for *Planck*-like simulations.

original resolution of the maps after denoising, thus allowing us to reconstruct and study the actual structure of the template on smaller scales in real observations. All this can be achieved at the same time with a huge amount of data compression, as discussed and tested in detail in the previous section. This last property is of particular interest in the context of template fitting, because a smaller data-set requires to invert a smaller covariance matrix, with a remarkable computational gain. No other technique among those we analyzed allows us to achieve this combined result, i.e., optimality, no loss of resolution and data compression. If we focus for example on the denoising method described in our previous section, we already commented how smoothing the maps has the obvious drawback of removing potentially relevant information at small scales. The JS estimator, on the other hand, preserve resolution, but does not allow for data compression.

4.3.2 ILC

Here we present a similar analysis as in the previous section but we focus on ILC techniques rather than template fitting. We work on the same set of simulations described before. We consider a needlet space ILC approach, implemented via the algorithm described in section 3, where the map needlet coefficients are thresholded before combining the channels. We compare this technique with a needlet space ILC where the weights are left free to vary between scales. Before applying this "standard" needlet ILC with no thresholding, we mask the galactic plane with the Planck component separation common mask in polarization. On the other hand, our method does not need any masking since the thresholding automatically remove the most contaminated regions. We want to verify how our blind threshold selection criterion based on the minimization of anisotropy performs with respect to a standard, "a priori" masking procedure.

Figures 8 and 9 show the results for LSPE and Planck respectively. We find that the results are comparable, proving that our technique performs well in selecting the foreground contaminated regions. We point out here that our currently implemented ILC technique can be improved in several ways, *e.g.* changing the weight in different areas of the sky. However,
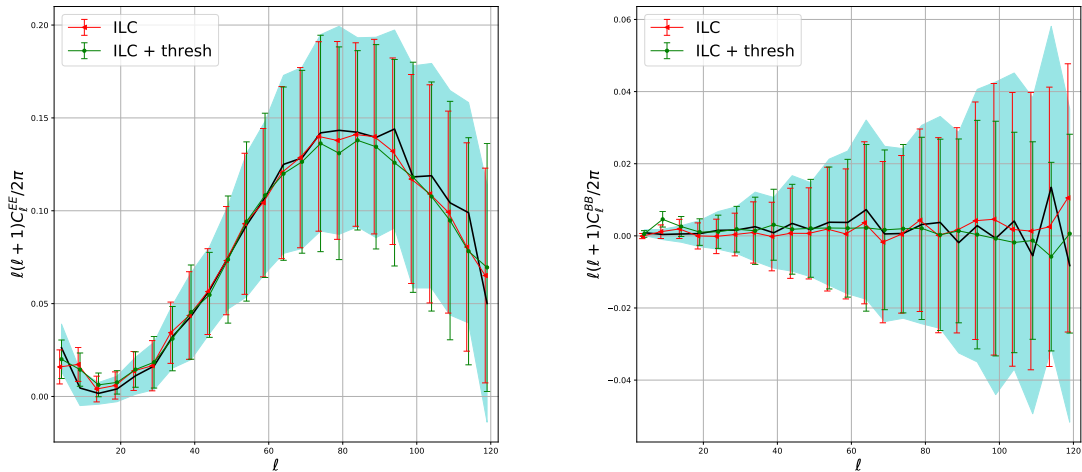
Figure 8. Power spectrum reconstruction with ILC with and without thresholding for SWIPE simulations, corrected for the average noise contribution. The conventions are the same as in previous figures: the average power spectrum from input simulations is represented in black, with the shaded blue area showing its standard error. Green and red lines show the average power spectra of the cleaned map with and without thresholding respectively.

we did not introduce these additional refinements since – as for the previous template-fitting analysis – they are not required for what we are strictly interested about in this work, namely the specific impact of thresholding on the cleaning procedure.

These results show how this technique can act as an automatic, blind masking method, with similar performance as the standard real space mask. It is relevant to notice that this approach allows recovering the input power spectrum without external assumptions on the contaminated areas of the sky (i.e., the masked part of the sky is recovered internally via thresholding, and no externally generated galactic mask is required). We also stress again that the role of thresholding, in this case, is conceptually completely different from the denoising performed in the template fitting algorithm. This is an additional probe of the large flexibility of this method.

5   Conclusions

In this paper, we showed several applications of needlet thresholding techniques to the problem of CMB component separation.

The unifying idea behind our study is that of exploiting sparsity of foreground components in the needlet representation, as a tool to separate foregrounds from the stochastic background by exploiting their peculiar morphological features (anisotropy, non-Gaussianity).

In the first part of our analysis, we tested explicitly on simulations that our needlet expansion of foreground templates is sparse, by using Gini coefficients as a measure of sparsity. We then showed how thresholding allows reconstructing a noisy template with high accuracy, up to small scales, well below the noise level. We also made a comparison between different denoising techniques, showing that for our purposes, needlet thresholding has the best
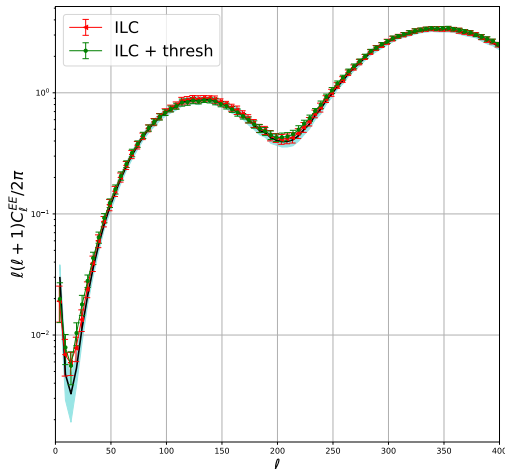
Figure 9. Same results as shown in figure 8, but for EE Power Spectrum of *Planck*-like simulations

performance in terms of reconstruction accuracy, while preserving the full resolution of the templates and at the same time achieving strong data compression.

After this investigation, in the second part of our study we implemented specific needlet thresholding procedures as extensions of existing component separation techniques. We then verified whether and in which situations this could improve the final CMB reconstruction. To this purpose, we focused on two well-known component separation procedures, namely ILC and template-fitting, considering simulated data sets and using as figure of merit the reconstruction of the input CMB polarization power spectrum. We compared bot soft- and hard-thresholding schemes and developed different procedures to set the optimal threshold level.

In the case of ILC, the role of thresholding is that of "pre-cleaning" single channels, before combining all the frequencies. This captures information on the foreground spatial distribution, which complements spectral frequency information and can in principle lead to a more accurate foreground cleaning procedure. In the case of template fitting, we have instead already discussed how thresholding is a powerful denoising method for internal templates.

After applying our algorithms to realistic simulations of different experimental setups, we found in practice that thresholding can be useful in experiments with few frequency channels, in conditions of low signal-to-noise. This is logical, since in these cases the original internal foreground templates are very noisy and the small frequency coverage reduces the accuracy of the standard approaches.

The best performance of thresholding in our tests are in particular found when considering a template fitting technique in an LSPE like experiment, especially for B-mode. Our ILC-thresholded algorithm, where we set the threshold level to maximize isotropy, gives instead similar results to standard ILC. Similarly, as anticipated, only marginal improvements are obtained in both cases for a Planck-like experiment with many frequency bands.

After the preliminary exploration discussed in this paper, we will therefore focus in a future work on developing in detail a full thresholding-based, needlet template-fitting pipeline. We will also explore the performance of this approach in a different context, namely for

foreground cleaning and template reconstruction in intensity mapping experiments.

Acknowledgments

References

[1] M. Kamionkowski, A. Kosowsky, and A. Stebbins, *Statistics of cosmic microwave background polarization*, Phys. Rev. **D55** (1997) 7368–7388, [`astro-ph/9611125`].

[2] M. Zaldarriaga and U. Seljak, *An all sky analysis of polarization in the microwave background*, Phys. Rev. **D55** (1997) 1830–1840, [`astro-ph/9609170`].

[3] U. Seljak and M. Zaldarriaga, *Signature of gravity waves in polarization of the microwave background*, Phys. Rev. Lett. **78** (1997) 2054–2057, [`astro-ph/9609169`].

[4] M. Kamionkowski, A. Kosowsky, and A. Stebbins, *A Probe of primordial gravity waves and vorticity*, Phys. Rev. Lett. **78** (1997) 2058–2061, [`astro-ph/9609132`].

[5] Planck Collaboration, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Band ay, R. B. Barreiro, N. Bartolo, S. Basak, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, A. Bracco, M. Bucher, C. Burigana, E. Calabrese, J. F. Cardoso, J. Carron, H. C. Chiang, C. Combet, B. P. Crill, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, C. Dickinson, J. M. Diego, A. Ducout, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, E. Falgarone, Y. Fantaye, K. Ferrière, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, T. Ghosh, J. González-Nuevo, K. M. Górski, A. Gruppuso, J. E. Gudmundsson, V. Guillet, W. Handley, F. K. Hansen, D. Herranz, Z. Huang, A. H. Jaffe, W. C. Jones, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, J. M. Lamarre, A. Lasenby, M. Le Jeune, F. Levrier, M. Liguori, P. B. Lilje, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, P. G. Martin, E. Martínez-González, S. Matarrese, J. D. McEwen, P. R. Meinhold, A. Melchiorri, M. Migliaccio, M. A. Miville-Deschênes, D. Molinari, A. Moneti, L. Montier, G. Morgante, P. Natoli, L. Pagano, D. Paoletti, V. Pettorino, F. Piacentini, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, J. D. Soler, L. D. Spencer, J. A. Tauber, D. Tavagnacco, L. Toffolatti, M. Tomasi, T. Trombetti, J. Valiviita, F. Vansyngel, F. Van Tent, P. Vielva, F. Villa, N. Vittorio, I. K. Wehus, A. Zacchei, and A. Zonca, *Planck 2018 results. XI. Polarized dust foregrounds*, arXiv e-prints (Jan, 2018) arXiv:1801.04945, [`arXiv:1801.04945`].

[6] J. Dunkley, A. Amblard, C. Baccigalupi, M. Betoule, D. Chuss, A. Cooray, J. Delabrouille, C. Dickinson, G. Dobler, J. Dotson, H. K. Eriksen, D. Finkbeiner, D. Fixsen, P. Fosalba, A. Fraisse, C. Hirata, A. Kogut, J. Kristiansen, C. Lawrence, A. M. MagalhÃces, M. A. MivilleâĂŘDeschenes, S. Meyer, A. Miller, S. K. Naess, L. Page, H. V. Peiris, N. Phillips, E. Pierpaoli, G. Rocha, J. E. Vaillancourt, and L. Verde, *Prospects for polarized foreground*

removal, *AIP Conference Proceedings* **1141** (2009), no. 1 222–264,
[https://aip.scitation.org/doi/pdf/10.1063/1.3160888].

[7] **BICEP2/Keck and Planck Collaborations** Collaboration, P. A. R. Ade, N. Aghanim, Z. Ahmed, R. W. Aikin, K. D. Alexander, M. Arnaud, J. Aumont, C. Baccigalupi, A. J. Banday, D. Barkats, R. B. Barreiro, J. G. Bartlett, N. Bartolo, E. Battaner, K. Benabed, A. Benoît, A. Benoit-Lévy, S. J. Benton, J.-P. Bernard, M. Bersanelli, P. Bielewicz, C. A. Bischoff, J. J. Bock, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, J. A. Brevik, M. Bucher, I. Buder, E. Bullock, C. Burigana, R. C. Butler, V. Buza, E. Calabrese, J.-F. Cardoso, A. Catalano, A. Challinor, R.-R. Chary, H. C. Chiang, P. R. Christensen, L. P. L. Colombo, C. Combet, J. Connors, F. Couchot, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, J.-M. Delouis, F.-X. Désert, C. Dickinson, J. M. Diego, H. Dole, S. Donzelli, O. Doré, M. Douspis, C. D. Dowell, L. Duband, A. Ducout, J. Dunkley, X. Dupac, C. Dvorkin, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, E. Falgarone, J. P. Filippini, F. Finelli, S. Fliescher, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frejsel, S. Galeotta, S. Galli, K. Ganga, T. Ghosh, M. Giard, E. Gjerløw, S. R. Golwala, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, M. Halpern, F. K. Hansen, D. Hanson, D. L. Harrison, M. Hasselfield, G. Helou, S. Henrot-Versillé, D. Herranz, S. R. Hildebrandt, G. C. Hilton, E. Hivon, M. Hobson, W. A. Holmes, W. Hovest, V. V. Hristov, K. M. Huffenberger, H. Hui, G. Hurier, K. D. Irwin, A. H. Jaffe, T. R. Jaffe, J. Jewell, W. C. Jones, M. Juvela, A. Karakci, K. S. Karkare, J. P. Kaufman, B. G. Keating, S. Kefeli, E. Keihänen, S. A. Kernasovskiy, R. Keskitalo, T. S. Kisner, R. Kneissl, J. Knoche, L. Knox, J. M. Kovac, N. Krachmalnicoff, M. Kunz, C. L. Kuo, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J.-M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, E. M. Leitch, R. Leonardi, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, M. Lueker, J. F. Macías-Pérez, B. Maffei, D. Maino, N. Mandolesi, A. Mangilli, M. Maris, P. G. Martin, E. Martínez-González, S. Masi, P. Mason, S. Matarrese, K. G. Megerian, P. R. Meinhold, A. Melchiorri, L. Mendes, A. Mennella, M. Migliaccio, S. Mitra, M.-A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, C. B. Netterfield, H. T. Nguyen, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, R. O'Brient, R. W. Ogburn, A. Orlando, L. Pagano, F. Pajot, R. Paladini, D. Paoletti, B. Partridge, F. Pasian, G. Patanchon, T. J. Pearson, O. Perdereau, L. Perotto, V. Pettorino, F. Piacentini, M. Piat, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, N. Ponthieu, G. W. Pratt, S. Prunet, C. Pryke, J.-L. Puget, J. P. Rachen, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, A. Renzi, S. Richter, I. Ristorcelli, G. Rocha, M. Rossetti, G. Roudier, M. Rowan-Robinson, J. A. Rubiño Martín, B. Rusholme, M. Sandri, D. Santos, M. Savelainen, G. Savini, R. Schwarz, D. Scott, M. D. Seiffert, C. D. Sheehy, L. D. Spencer, Z. K. Staniszewski, V. Stolyarov, R. Sudiwala, R. Sunyaev, D. Sutton, A.-S. Suur-Uski, J.-F. Sygnet, J. A. Tauber, G. P. Teply, L. Terenzi, K. L. Thompson, L. Toffolatti, J. E. Tolan, M. Tomasi, M. Tristram, M. Tucci, A. D. Turner, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, A. G. Vieregg, F. Villa, L. A. Wade, B. D. Wandelt, R. Watson, A. C. Weber, I. K. Wehus, M. White, S. D. M. White, J. Willmert, C. L. Wong, K. W. Yoon, D. Yvon, A. Zacchei, and A. Zonca, *Joint analysis of bicep2/keck array and planck data*, *Phys. Rev. Lett.* **114** (Mar, 2015) 101301.

[8] J. Errard, S. M. Feeney, H. V. Peiris, and A. H. Jaffe, *Robust forecasts on fundamental physics from the foreground-obscured, gravitationally-lensed CMB polarization*, *"J. Cosmology Astropart. Phys."* **2016** (Mar, 2016) 052, [arXiv:1509.06770].

[9] M. Remazeilles, A. J. Banday, C. Baccigalupi, S. Basak, A. Bonaldi, G. De Zotti, J. Delabrouille, C. Dickinson, H. K. Eriksen, J. Errard, R. Fernandez-Cobos, U. Fuskeland, C. Hervías-Caimapo, M. López-Caniego, E. Martinez-González, M. Roman, P. Vielva, I. Wehus, A. Achucarro, P. Ade, R. Allison, M. Ashdown, M. Ballardini, R. Banerji, J. Bartlett,

N. Bartolo, D. Baumann, M. Bersanelli, M. Bonato, J. Borrill, F. Bouchet, F. Boulanger, T. Brinckmann, M. Bucher, C. Burigana, A. Buzzelli, Z. Y. Cai, M. Calvo, C. S. Carvalho, G. Castellano, A. Challinor, J. Chluba, S. Clesse, I. Colantoni, A. Coppolecchia, M. Crook, G. D'Alessandro, P. de Bernardis, G. de Gasperis, J. M. Diego, E. Di Valentino, S. Feeney, S. Ferraro, F. Finelli, F. Forastieri, S. Galli, R. Genova-Santos, M. Gerbino, J. González-Nuevo, S. Grandis, J. Greenslade, S. Hagstotz, S. Hanany, W. Handley, C. Hernandez-Monteagudo, M. Hills, E. Hivon, K. Kiiveri, T. Kisner, T. Kitching, M. Kunz, H. Kurki-Suonio, L. Lamagna, A. Lasenby, M. Lattanzi, J. Lesgourgues, A. Lewis, M. Liguori, V. Lindholm, G. Luzzi, B. Maffei, C. J. A. P. Martins, S. Masi, S. Matarrese, D. McCarthy, J. B. Melin, A. Melchiorri, D. Molinari, A. Monfardini, P. Natoli, M. Negrello, A. Notari, A. Paiella, D. Paoletti, G. Patanchon, M. Piat, G. Pisano, L. Polastri, G. Polenta, A. Pollo, V. Poulin, M. Quartin, J. A. Rubino-Martin, L. Salvati, A. Tartari, M. Tomasi, D. Tramonte, N. Trappe, T. Trombetti, C. Tucker, J. Valiviita, R. Van de Weijgaert, B. van Tent, V. Vennin, N. Vittorio, K. Young, and M. Zannoni, *Exploring cosmic origins with CORE: B-mode component separation*, *"J. Cosmology Astropart. Phys."* **2018** (Apr, 2018) 023, [arXiv:1704.04501].

[10] B. S. Hensley and P. Bull, *Mitigating Complex Dust Foregrounds in Future Cosmic Microwave Background Polarization Experiments*, *"Astrophys. J."* **853** (Feb, 2018) 127, [arXiv:1709.07897].

[11] R. Stompor, J. Errard, and D. Poletti, *Forecasting performance of CMB experiments in the presence of complex foreground contaminations*, *Phys.Rev.D* **94** (Oct, 2016) 083526, [arXiv:1609.03807].

[12] J. Delabrouille and J.-F. Cardoso, *Diffuse Source Separation in CMB Observations*, pp. 159–205. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[13] Planck Collaboration, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Band ay, R. B. Barreiro, N. Bartolo, S. Basak, K. Benabed, M. Bersanelli, P. Bielewicz, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, E. Calabrese, J. F. Cardoso, J. Carron, B. Casaponsa, A. Challinor, L. P. L. Colombo, C. Combet, B. P. Crill, F. Cuttaia, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, C. Dickinson, J. M. Diego, S. Donzelli, O. Doré, A. Ducout, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, E. Falgarone, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, W. Hand ley, F. K. Hansen, G. Helou, D. Herranz, Z. Huang, A. H. Jaffe, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, F. Levrier, M. Liguori, P. B. Lilje, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, P. G. Martin, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. A. Miville-Deschênes, D. Molinari, A. Moneti, L. Montier, G. Morgante, P. Natoli, F. Oppizzi, L. Pagano, D. Paoletti, B. Partridge, M. Peel, V. Pettorino, F. Piacentini, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, D. S. Seljebotn, C. Sirignano, L. D. Spencer, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, H. Thommesen, L. Toffolatti, M. Tomasi, T. Trombetti, J. Valiviita, B. Van Tent, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, A. Zacchei, and A. Zonca, *Planck 2018 results. IV. Diffuse component separation*, *arXiv e-prints* (Jul, 2018) arXiv:1807.06208, [arXiv:1807.06208].

[14] M. Hazumi, P. A. R. Ade, Y. Akiba, D. Alonso, K. Arnold, J. Aumont, C. Baccigalupi, D. Barron, S. Basak, S. Beckman, J. Borrill, F. Boulanger, M. Bucher, E. Calabrese, Y. Chinone, S. Cho, A. Cukierman, D. W. Curtis, T. de Haan, M. Dobbs, A. Dominjon,

T. Dotani, L. Duband, A. Ducout, J. Dunkley, J. M. Duval, T. Elleflot, H. K. Eriksen, J. Errard, J. Fischer, T. Fujino, T. Funaki, U. Fuskeland, K. Ganga, N. Goeckner-Wald, J. Grain, N. W. Halverson, T. Hamada, T. Hasebe, M. Hasegawa, K. Hattori, M. Hattori, L. Hayes, N. Hidehira, C. A. Hill, G. Hilton, J. Hubmayr, K. Ichiki, T. Iida, H. Imada, M. Inoue, Y. Inoue, K. D. Irwin, H. Ishino, O. Jeong, H. Kanai, D. Kaneko, S. Kashima, N. Katayama, T. Kawasaki, S. A. Kernasovskiy, R. Keskitalo, A. Kibayashi, Y. Kida, K. Kimura, T. Kisner, K. Kohri, E. Komatsu, K. Komatsu, C. L. Kuo, N. A. Kurinsky, A. Kusaka, A. Lazarian, A. T. Lee, D. Li, E. Linder, B. Maffei, A. Mangilli, M. Maki, T. Matsumura, S. Matsuura, D. Meilhan, S. Mima, Y. Minami, K. Mitsuda, L. Montier, M. Nagai, T. Nagasaki, R. Nagata, M. Nakajima, S. Nakamura, T. Namikawa, M. Naruse, H. Nishino, T. Nitta, T. Noguchi, H. Ogawa, S. Oguri, N. Okada, A. Okamoto, T. Okamura, C. Otani, G. Patanchon, G. Pisano, G. Rebeiz, M. Remazeilles, P. L. Richards, S. Sakai, Y. Sakurai, Y. Sato, N. Sato, M. Sawada, Y. Segawa, Y. Sekimoto, U. Seljak, B. D. Sherwin, T. Shimizu, K. Shinozaki, R. Stompor, H. Sugai, H. Sugita, A. Suzuki, J. Suzuki, O. Tajima, S. Takada, R. Takaku, S. Takakura, S. Takatori, D. Tanabe, E. Taylor, K. L. Thompson, B. Thorne, T. Tomaru, T. Tomida, N. Tomita, M. Tristram, C. Tucker, P. Turin, M. Tsujimoto, S. Uozumi, S. Utsunomiya, Y. Uzawa, F. Vansyngel, I. K. Wehus, B. Westbrook, M. Willer, N. Whitehorn, Y. Yamada, R. Yamamoto, N. Yamasaki, T. Yamashita, and M. Yoshida, *LiteBIRD: A Satellite for the Studies of B-Mode Polarization and Inflation from Cosmic Background Radiation Detection*, *Journal of Low Temperature Physics* **194** (Mar, 2019) 443–452.

[15] K. Abazajian, G. Addison, P. Adshead, Z. Ahmed, S. W. Allen, D. Alonso, M. Alvarez, A. Anderson, K. S. Arnold, C. Baccigalupi, K. Bailey, D. Barkats, D. Barron, P. S. Barry, J. G. Bartlett, R. Basu Thakur, N. Battaglia, E. Baxter, R. Bean, C. Bebek, A. N. Bender, B. A. Benson, E. Berger, S. Bhimani, C. A. Bischoff, L. Bleem, S. Bocquet, K. Boddy, M. Bonato, J. R. Bond, J. Borrill, F. R. Bouchet, M. L. Brown, S. Bryan, B. Burkhart, V. Buza, K. Byrum, E. Calabrese, V. Calafut, R. Caldwell, J. E. Carlstrom, J. Carron, T. Cecil, A. Challinor, C. L. Chang, Y. Chinone, H.-M. S. Cho, A. Cooray, T. M. Crawford, A. Crites, A. Cukierman, F.-Y. Cyr-Racine, T. de Haan, G. de Zotti, J. Delabrouille, M. Demarteau, M. Devlin, E. Di Valentino, M. Dobbs, S. Duff, A. Duivenvoorden, C. Dvorkin, W. Edwards, J. Eimer, J. Errard, T. Essinger-Hileman, G. Fabbian, C. Feng, S. Ferraro, J. P. Filippini, R. Flauger, B. Flaugher, A. A. Fraisse, A. Frolov, N. Galitzki, S. Galli, K. Ganga, M. Gerbino, M. Gilchriese, V. Gluscevic, D. Green, D. Grin, E. Grohs, R. Gualtieri, V. Guarino, J. E. Gudmundsson, S. Habib, G. Haller, M. Halpern, N. W. Halverson, S. Hanany, K. Harrington, M. Hasegawa, M. Hasselfield, M. Hazumi, K. Heitmann, S. Henderson, J. W. Henning, J. C. Hill, R. Hlozek, G. Holder, W. Holzapfel, J. Hubmayr, K. M. Huffenberger, M. Huffer, H. Hui, K. Irwin, B. R. Johnson, D. Johnstone, W. C. Jones, K. Karkare, N. Katayama, J. Kerby, S. Kernovsky, R. Keskitalo, T. Kisner, L. Knox, A. Kosowsky, J. Kovac, E. D. Kovetz, S. Kuhlmann, C.-l. Kuo, N. Kurita, A. Kusaka, A. Lahteenmaki, C. R. Lawrence, A. T. Lee, A. Lewis, D. Li, E. Linder, M. Loverde, A. Lowitz, M. S. Madhavacheril, A. Mantz, F. Matsuda, P. Mauskopf, J. McMahon, M. McQuinn, P. D. Meerburg, J.-B. Melin, J. Meyers, M. Millea, J. Mohr, L. Moncelsi, T. Mroczkowski, S. Mukherjee, M. Münchmeyer, D. Nagai, J. Nagy, T. Namikawa, F. Nati, T. Natoli, M. Negrello, L. Newburgh, M. D. Niemack, H. Nishino, M. Nordby, V. Novosad, P. O'Connor, G. Obied, S. Padin, S. Pandey, B. Partridge, E. Pierpaoli, L. Pogosian, C. Pryke, G. Puglisi, B. Racine, S. Raghunathan, A. Rahlin, S. Rajagopalan, M. Raveri, M. Reichanadter, C. L. Reichardt, M. Remazeilles, G. Rocha, N. A. Roe, A. Roy, J. Ruhl, M. Salatino, B. Saliwanchik, E. Schaan, A. Schillaci, M. M. Schmittfull, D. Scott, N. Sehgal, S. Shandera, C. Sheehy, B. D. Sherwin, E. Shirokoff, S. M. Simon, A. Slosar, R. Somerville, D. Spergel, S. T. Staggs, A. Stark, R. Stompor, K. T. Story, C. Stoughton, A. Suzuki, O. Tajima, G. P. Teply, K. Thompson, P. Timbie, M. Tomasi, J. I. Treu, M. Tristram, G. Tucker, C. Umiltà, A. e. van Engelen, J. D. Vieira, A. G. Vieregg, M. Vogelsberger, G. Wang, S. Watson, M. White, N. Whitehorn, E. J. Wollack, W. L. Kimmy Wu, Z. Xu, S. Yasini, J. Yeck, K. W. Yoon,

E. Young, and A. Zonca, *CMB-S4 Science Case, Reference Design, and Project Plan*, *arXiv e-prints* (Jul, 2019) arXiv:1907.04473, [`arXiv:1907.04473`].

[16] J.-F. Cardoso, M. Martin, J. Delabrouille, M. Betoule, and G. Patanchon, *Component separation with flexible models. Application to the separation of astrophysical emissions*, *ArXiv e-prints* (Mar., 2008) [`arXiv:0803.1814`].

[17] S. Basak and J. Delabrouille, *A needlet internal linear combination analysis of WMAP 7-year data: estimation of CMB temperature map and power spectrum*, *"Mon. Not. Roy. Astron. Soc."* **419** (Jan., 2012) 1163–1175, [`arXiv:1106.5383`].

[18] H. K. Eriksen, I. J. O'Dwyer, J. B. Jewell, B. D. Wandelt, D. L. Larson, K. M. Górski, S. Levin, A. J. Banday, and P. B. Lilje, *Power Spectrum Estimation from High-Resolution Maps by Gibbs Sampling*, *"Astrophys. J. S"* **155** (Dec., 2004) 227–241, [`astro-ph/0407028`].

[19] H. K. Eriksen, J. B. Jewell, C. Dickinson, A. J. Banday, K. M. GÃ§rski, and C. R. Lawrence, *Joint bayesian component separation and cmb power spectrum estimation*, *The Astrophysical Journal* **676** (2008), no. 1 10.

[20] J. Bobin, Y. Moudden, J. L. Starck, J. Fadili, and N. Aghanim, *SZ and CMB reconstruction using Generalized Morphological Component Analysis*, *Stat. Meth.* **5** (2008) 307, [`arXiv:0712.0588`].

[21] R. Fernández-Cobos, P. Vielva, R. B. Barreiro, and E. Martínez-González, *Multiresolution internal template cleaning: an application to the Wilkinson Microwave Anisotropy Probe 7-yr polarization data*, *"Mon. Not. Roy. Astron. Soc."* **420** (Mar., 2012) 2162–2169, [`arXiv:1106.2016`].

[22] C. Baccigalupi et al., *Neural networks and separation of cosmic microwave background and astrophysical signals in sky maps*, *Mon. Not. Roy. Astron. Soc.* **318** (2000) 769, [`astro-ph/0002257`].

[23] F. J. Narcowich, P. Petrushev, and J. D. Ward, *Localized tight frames on spheres*, *SIAM J. Math. Analysis* **38** (2006) 574–594.

[24] P. Baldi, G. Kerkyacharian, D. Marinucci, and D. Picard, *Asymptotics for spherical needlets*, *Ann. Statist.* **37** (06, 2009) 1150–1171.

[25] P. Baldi, G. Kerkyacharian, D. Marinucci, and D. Picard, *Subsampling needlet coefficients on the sphere*, *Bernoulli* **15** (05, 2009) 438–463.

[26] D. Pietrobon, A. Balbi, and D. Marinucci, *Integrated Sachs-Wolfe effect from the cross correlation of WMAP 3year and the NRAO VLA sky survey data: New results and constraints on dark energy*, *Phys. Rev. D* **74** (Aug., 2006) 043524, [`astro-ph/0606475`].

[27] S. Basak and J. Delabrouille, *A needlet ILC analysis of WMAP 9-year polarization data: CMB polarization power spectra*, *"Mon. Not. Roy. Astron. Soc."* **435** (Oct., 2013) 18–29, [`arXiv:1204.0292`].

[28] D. Marinucci and G. Peccati, *Random Fields on the Sphere: Representation, Limit Theorems and Cosmological Applications*. London Mathematical Society Lecture Note Series. Cambridge University Press, 2011.

[29] D. Marinucci, D. Pietrobon, A. Balbi, P. Baldi, P. Cabella, G. Kerkyacharian, P. Natoli, D. Picard, and N. Vittorio, *Spherical needlets for cosmic microwave background data analysis*, *Mon. Not. Roy. Astron. Soc.* **383** (Jan., 2008) 539–545, [`arXiv:0707.0844`].

[30] K. M. Gorski, B. D. Wandelt, F. K. Hansen, E. Hivon, and A. J. Banday, *The HEALPix Primer*, *arXiv e-prints* (May, 1999) astro–ph/9905275, [`astro-ph/9905275`].

[31] J.-L. Starck, F. Murtagh, and J. Fadili, *Sparse Image and Signal Processing: Wavelets and Related Geometric Multiscale Analysis*. Cambridge University Press, New York, NY, USA, 2nd ed., 2015.

[32] D. Donoho and J. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, *Biometrika* **81** (1994), no. 3 425–455.

[33] F. Vansyngel, B. D. Wandelt, J.-F. Cardoso, and K. Benabed, *Semi-blind Bayesian inference of CMB map and power spectrum*, *A&A* **588** (Apr., 2016) A113, [arXiv:1409.0858].

[34] Y. Moudden, J.-F. Cardoso, J.-L. Starck, and J. Delabrouille, *Blind Component Separation in Wavelet Space: Application to CMB Analysis*, *EURASIP Journal on Applied Signal Processing* **2005** (Dec., 2005) 484606, [astro-ph/0407053].

[35] Bobin, J., Sureau, F., Starck, J.-L., Rassat, A., and Paykari, P., *Joint planck and wmap cmb map reconstruction*, *A&A* **563** (2014) A105.

[36] J. M. Fadili, J.-L. Starck, J. Bobin, and Y. Moudden, *Image decomposition and separation using sparse representations: an overview*, *Proceedings of the IEEE, Special Issue: Applications of Sparse Representation* **98** (2010), no. 6 983–994.

[37] V. Cammarota and D. Marinucci, *The Stochastic Properties of ℓ^1-Regularized Spherical Gaussian Fields*, *Analysis* **38** (2015), no. 2 262–283, [arXiv:1309.3459].

[38] S. M. Feeney, D. Marinucci, J. D. McEwen, H. V. Peiris, B. D. Wandelt, and V. Cammarota, *Sparse Inpainting and Isotropy*, *J. Cosmology Astropart. Phys.* **1401** (2014) 050, [arXiv:1308.0602].

[39] F. K. Hansen, A. J. Banday, H. K. Eriksen, K. M. Gorski, and P. B. Lilje, *Foreground subtraction of cosmic microwave background maps using wi-fit (wavelet based high resolution fitting of internal templates)*, *Astrophys. J.* **648** (2006) 784–796, [astro-ph/0603308].

[40] E. Hivon, K. M. Gorski, C. B. Netterfield, B. P. Crill, S. Prunet, and F. Hansen, *Master of the cosmic microwave background anisotropy power spectrum: a fast method for statistical analysis of large and complex cosmic microwave background data sets*, *Astrophys. J.* **567** (2002) 2, [astro-ph/0105302].

[41] N. Hurley and S. Rickard, *Comparing measures of sparsity*, *IEEE Transactions on Information Theory* **55** (Oct, 2009) 4723–4741.

[42] S. Rickard, *Sparse sources are separated sources*, in *2006 14th European Signal Processing Conference*, pp. 1–5, Sep., 2006.

[43] R. Aaberge, *Axiomatic characterization of the gini coefficient and lorenz curve orderings*, *Journal of Economic Theory* **101** (2001), no. 1 115 – 132.

[44] Delabrouille, J., Betoule, M., Melin, J.-B., Miville-Deschênes, M.-A., Gonzalez-Nuevo, J., Le Jeune, M., Castex, G., de Zotti, G., Basak, S., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Bernard, J.-P., Bouchet, F. R., Clements, D. L., da Silva, A., Dickinson, C., Dodu, F., Dolag, K., Elsner, F., Fauvet, L., Faÿ, G., Giardino, G., Leach, S., Lesgourgues, J., Liguori, M., Macías-Pérez, J. F., Massardi, M., Matarrese, S., Mazzotta, P., Montier, L., Mottet, S., Paladini, R., Partridge, B., Piffaretti, R., Prezeau, G., Prunet, S., Ricciardi, S., Roman, M., Schaefer, B., and Toffolatti, L., *The pre-launch planck sky model: a model of sky emission at submillimetre to centimetre wavelengths*, *A&A* **553** (2013) A96.

[45] B. Thorne, J. Dunkley, D. Alonso, and S. Næss, *The Python Sky Model: software for simulating the Galactic microwave sky*, *"Mon. Not. Roy. Astron. Soc."* **469** (Aug., 2017) 2821–2833, [arXiv:1608.02841].

[46] D. L. Donoho and I. M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, *Journal of the American Statistical Association* **90** (1995), no. 432 1200–1224, [https://www.tandfonline.com/doi/pdf/10.1080/01621459.1995.10476626].

[47] W. James and C. Stein, *Estimation with quadratic loss*, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, (Berkeley, Calif.), pp. 361–379, University of California Press, 1961.

[48] C. Stein, *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*, in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, (Berkeley, Calif.), pp. 197–206, University of California Press, 1956.

[49] The LSPE collaboration, S. Aiola, G. Amico, P. Battaglia, E. Battistelli, A. Baù, P. de Bernardis, M. Bersanelli, A. Boscaleri, F. Cavaliere, A. Coppolecchia, A. Cruciani, F. Cuttaia, A. D' Addabbo, G. D'Alessandro, S. De Gregori, F. Del Torto, M. De Petris, L. Fiorineschi, C. Franceschet, E. Franceschi, M. Gervasi, D. Goldie, A. Gregorio, V. Haynes, N. Krachmalnicoff, L. Lamagna, B. Maffei, D. Maino, S. Masi, A. Mennella, N. M. Wah, G. Morgante, F. Nati, L. Pagano, A. Passerini, O. Peverini, F. Piacentini, L. Piccirillo, G. Pisano, S. Ricciardi, P. Rissone, G. Romeo, M. Salatino, M. Sandri, A. Schillaci, L. Stringhetti, A. Tartari, R. Tascone, L. Terenzi, M. Tomasi, E. Tommasi, F. Villa, G. Virone, S. Withington, A. Zacchei, and M. Zannoni, *The Large-Scale Polarization Explorer (LSPE)*, *arXiv e-prints* (Aug, 2012) arXiv:1208.0281, [`arXiv:1208.0281`].

[50] P. de Bernardis, S. Aiola, G. Amico, E. Battistelli, A. Coppolecchia, A. Cruciani, A. D'Addabbo, G. D'Alessandro, S. De Gregori, M. De Petris, D. Goldie, R. Gualtieri, V. Haynes, L. Lamagna, B. Maffei, S. Masi, F. Nati, M. W. Ng, L. Pagano, F. Piacentini, L. Piccirillo, G. Pisano, G. Romeo, M. Salatino, A. Schillaci, E. Tommasi, and S. Withington, *SWIPE: a bolometric polarimeter for the Large-Scale Polarization Explorer*, in *"Proc. SPIE"*, vol. 8452 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, p. 84523F, Sep, 2012. `arXiv:1208.0282`.

[51] R. Gualtieri, E. S. Battistelli, A. Cruciani, P. de Bernardis, M. Biasotti, D. Corsini, F. Gatti, L. Lamagna, and S. Masi, *Multi-mode TES Bolometer Optimization for the LSPE-SWIPE Instrument*, *Journal of Low Temperature Physics* **184** (Aug, 2016) 527–533, [`arXiv:1602.07744`].