

A statistical analysis of the “internal linear combination” method in problems of signal separation as in cosmic microwave background observations

R. Vio¹ and P. Andreani^{2,3}

¹ Chip Computers Consulting s.r.l., Viale Don L. Sturzo 82, S.Liberale di Marcon, 30020 Venice, Italy
e-mail: robertovio@tin.it,

² ESO, Karl Schwarzschild strasse 2, 85748 Garching, Germany

³ INAF - Osservatorio Astronomico di Trieste, via Tiepolo 11, 34143 Trieste, Italy
e-mail: pandrean@eso.org

Received 2 January 2008 / Accepted 26 May 2008

ABSTRACT

Aims. The separation of foreground contamination from cosmic microwave background (CMB) observations is one of the most challenging and important problems of digital signal processing in Cosmology. In the literature, various techniques have been presented but no general consensus about their real performance and properties has been reached. This is due to the characteristics of these techniques that have been studied essentially through numerical simulations based on semi-empirical models of the CMB and the Galactic foreground. Such models often have different levels of sophistication and/or are based on different physical assumptions (e.g. the number of Galactic components and level of the noise). Hence, a reliable comparison is difficult. What is missing is a statistical analysis of the properties of the proposed methodologies. Here, we consider the *Internal Linear Combination* method (ILC) which, among the separation techniques, requires the smallest number of a priori assumptions. This feature is of particular interest in the context of CMB polarization measurements at small angular scales where the lack of knowledge of the polarized backgrounds represents a serious limit.

Methods. The statistical characteristics of ILC are examined through an analytical approach and the basic conditions are fixed so as to work satisfactorily.

Results. ILC provides satisfactory results only under rather restrictive conditions. This is a critical fact to take into consideration in planning future ground-based observations (e.g., with ALMA) where, contrary to satellite experiments, there is the possibility of having a certain control over the experimental conditions.

Key words. methods: data analysis – methods: statistical – cosmology: cosmic microwave background

1. Introduction

The experimental progresses in the detection of cosmological and astrophysical emissions require a parallel development of data analysis techniques in order to extract the maximum physical information from the data. An example of interest is represented by signals that are mixture of the emission of distinct physical mechanisms. The study of the underlying physical processes needs the separation of the different components that contribute to the observed signals. Such is the case of the Cosmic Microwave Background (CMB) observations where there is the need to separate the CMB from diffuse foreground originated by our own Galaxy. In this context, an extensive literature is available and many approaches have been proposed (for a review, see [Delabrouille & Cardoso 2007](#)). However, no general consensus about their real performances and properties has been reached. This is due to the approach followed to determine the characteristics of these techniques that is based on numerical simulations that make use of semi-empirical models of the CMB and the Galactic foreground. Such models often have different level of sophistication and/or are based on different physical assumptions (e.g., the number of Galactic components and level of the noise). Hence, a reliable comparison is difficult. A trustworthy assessment of the real capabilities of such methodologies would

require a rigorous analysis of their theoretical statistical characteristics independently of the specific context where they are applied. However, to our knowledge, in the literature nothing has never been presented in this sense (however, see [Saha et al. 2007](#), for a discussion concerning the estimation of the CMB power spectrum). This is even more true for CMB polarization measurements where the available a priori information is quite limited and the use of *blind* separation techniques an unavoidable choice ([Stivoli et al. 2006](#)).

A situation such as this one is clearly unsatisfactory. This also because in the near future some innovative ground-based experiments are planned for polarization observations at very high spatial resolution, such as with the Atacama Large submillimetre/millimetre array (ALMA). One important advantage of these experiments is that, contrary to the satellite observations, they will allow a certain control of the experimental conditions. Hence, a full exploitation of the capabilities of this facility (and other instruments) requires a careful planning of the observations. The recording of good quality data will allow an effective application of the chosen separation methodology. For this reason, in this work we start exploring the capabilities of algorithms aimed at a careful subtraction of the foreground sources when the amount of available a priori information is limited (an expected situation for polarization observations). In particular,

we consider one of the most widely used approaches for the separation of different emissions, the *internal linear combination* method (ILC), which was adopted, for instance, in the reduction of the data from the Wilkinson Microwave Anisotropy Probe (WMAP) satellite for CMB observations (Bennett et al. 2003). Among the separation techniques, ILC requires the smallest number of a priori assumptions. For the reason presented above, we do not perform an application to any astrophysical dataset here, but rather we study the general properties of this technique in order to fix the conditions under which it can be expected to produce reliable results.

In the following, the available data are assumed in the form of N_o maps, taken at different frequencies, containing N_p pixels each. More precisely, if $S^{(i)}(p)$ provides the value of the p th pixel for a map obtained at channel “ i ”, our starting model is:

$$S^{(i)}(p) = \mathfrak{E}_c^{(i)}(p) + S_f^{(i)}(p) + \mathcal{N}^{(i)}(p) \quad (1)$$

where $\mathfrak{E}_c^{(i)}(p)$, $S_f^{(i)}(p)$ and $\mathcal{N}^{(i)}(p)$ are the contributions due to the CMB, the diffuse Galactic foreground and the experimental noise, respectively. Although not necessary for later arguments, it is assumed that all of these contributions are representable by means of stationary random fields. At least locally, in many experimental situations this is an acceptable approximation. If not, it is often made anyway since it permits a statistical treatment of the problem of interest and the results can be used as a benchmark in the analysis of more complex scenarios. In the present context, this assumption holds on small patches of the sky. Finally, without loss of generality, for ease of notation the random fields are supposed as the realization of zero-mean spatial processes. In the present work the contribution of non-diffuse components (e.g., due to SZ cluster, point-sources, ...) are not considered and they are assumed to have been removed through other methodologies.

The paper is organized as follows: in Sect. 2 the relevant analytical formulas are introduced, in Sect. 3 the statistical framework of ILC in noiseless observations is presented, in Sect. 4 the case of noisy data is considered and the conclusions are drawn in Sect. 5.

2. Formalization of the problem

The idea behind ILC is simple. The main assumption is that model (1) can be written as

$$S^{(i)}(p) = \mathfrak{E}_c(p) + S_f^{(i)}(p) + \mathcal{N}^{(i)}(p), \quad (2)$$

i.e. the template of the CMB component is independent of the observing channel. A way to exploit this fact is to average N_o images $\{S^{(i)}(p)\}_{i=1}^{N_o}$ giving a specific weight w_i to each of them so as to minimize the impact of the foreground and noise (Bennett et al. 2003). This means to look for a solution of type

$$\widehat{\mathfrak{E}}_c(p) = \sum_{i=1}^{N_o} w_i S^{(i)}(p). \quad (3)$$

If the constraint $\sum_{i=1}^{N_o} w_i = 1$ is imposed, Eq. (3) becomes

$$\widehat{\mathfrak{E}}_c(p) = \mathfrak{E}_c(p) + \sum_{i=1}^{N_o} w_i [S_f^{(i)}(p) + \mathcal{N}^{(i)}(p)]. \quad (4)$$

Now, from this equation it is clear that, for a given pixel “ p ”, the only variable terms are in the summation. Hence, under the assumption of independence of $\mathfrak{E}_c(p)$ from $S_f^{(i)}(p)$ and $\mathcal{N}^{(i)}(p)$, the weights $\{w_i\}$ have to minimize the variance of $\widehat{\mathfrak{E}}_c(p)$, i.e.

$$\{w_i\} = \arg \min_{\{w_i\}} \text{VAR} \left[\mathfrak{E}_c(p) + \text{VAR} \left[\sum_{i=1}^{N_o} w_i (S_f^{(i)}(p) + \mathcal{N}^{(i)}(p)) \right] \right], \quad (5)$$

where $\text{VAR}[s(p)]$ is the *expected variance* of $s(p)$. If $\mathbf{S}^{(i)}$ denotes a row vector such as $\mathbf{S}^{(i)} = [S^{(i)}(1), S^{(i)}(2), \dots, S^{(i)}(N_p)]$ and the $N_o \times N_p$ matrix \mathbf{S} is defined as

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}^{(1)} \\ \mathbf{S}^{(2)} \\ \vdots \\ \mathbf{S}^{(N_o)} \end{pmatrix}, \quad (6)$$

then Eq. (2) becomes

$$\mathbf{S} = \mathfrak{E}_c + \mathbf{S}_f + \mathcal{N}. \quad (7)$$

In this case, the weights are given by (Eriksen et al. 2004)

$$\mathbf{w} = \frac{\mathbf{C}_S^{-1} \mathbf{1}}{\mathbf{1}^T \mathbf{C}_S^{-1} \mathbf{1}}, \quad (8)$$

where \mathbf{C}_S is the $N_o \times N_o$ cross-covariance matrix of the random processes that generate \mathbf{S} , i.e.

$$\mathbf{C}_S = \text{E}[\mathbf{S}\mathbf{S}^T], \quad (9)$$

and $\mathbf{1} = (1, 1, \dots, 1)^T$ is a column vector of all ones. Here, $\text{E}[\cdot]$ denotes the *expectation operator*. Hence, the ILC estimator takes the form

$$\widehat{\mathfrak{E}}_c = \mathbf{w}^T \mathbf{S}, \quad (10)$$

$$= \alpha \mathbf{1}^T \mathbf{C}_S^{-1} \mathbf{S}, \quad (11)$$

with $\mathbf{1}^T \mathbf{w} = 1$ and the scalar quantity α given by

$$\alpha = [\mathbf{1}^T \mathbf{C}_S^{-1} \mathbf{1}]^{-1}. \quad (12)$$

In practical applications, matrix \mathbf{C}_S is unknown and has to be estimated from the data. Typically, this is done by means of the estimator

$$\widehat{\mathbf{C}}_S = \frac{1}{N_p} \mathbf{S}\mathbf{S}^T. \quad (13)$$

In this case, the ILC estimator is given by Eqs. (10)–(12) with \mathbf{C}_S and \mathbf{w} replaced, respectively, by $\widehat{\mathbf{C}}_S$ and

$$\widehat{\mathbf{w}} = \frac{\widehat{\mathbf{C}}_S^{-1} \mathbf{1}}{\mathbf{1}^T \widehat{\mathbf{C}}_S^{-1} \mathbf{1}}. \quad (14)$$

If the observed maps are not zero-mean, they have to be centered before the computation of $\widehat{\mathbf{C}}_S$. After that, the resulting weights can be applied directly to the original (i.e., non-centered) \mathbf{S} . The computation of $\widehat{\mathbf{C}}_S$ is the only point where the fact of working with non-zero mean maps has to be taken into account.

Although in the literature it might appear that the estimator (11) has *optimal* properties, this is not true. In Eq. (2) the term $S_f^{(i)}(p) + \mathcal{N}^{(i)}(p)$ is considered as a single noise component

¹ In the present work, p indexes pixels in the classic spatial domain. However, the same formalism applies if other domains are considered, for example, the Fourier one.

(e.g., see [Eriksen et al. 2004](#); [Hinshaw et al. 2007](#)). In this way the problem is apparently simplified since it is reduced to the separation of two components only. No a priori information on this “global” noise is required. However, this approach can lead to wrong conclusions. For example, since all the components in the mixtures \mathbf{S} are assumed to be zero-mean, from Eq. (4) one could conclude that

$$E[\widehat{\mathfrak{E}}_c | \mathfrak{E}_c] = \mathfrak{E}_c + \mathbf{w}^T E[\mathbf{S}_f] + \mathbf{w}^T E[\mathcal{N}] = \mathfrak{E}_c, \quad (15)$$

i.e. the ILC estimator is unbiased². This is not correct: the claim that $\widehat{\mathfrak{E}}_c$ is unbiased requires ones to prove that

$$E[\widehat{\mathfrak{E}}_c | \mathfrak{E}_c, \mathbf{S}_f] = \mathfrak{E}_c + \mathbf{w}^T \mathbf{S}_f + \mathbf{w}^T E[\mathcal{N}] = \mathfrak{E}_c. \quad (16)$$

The reason is that \mathbf{S}_f is a fixed realization of a random process. There is no way to obtain another one. Even if observed many times (under the same experimental conditions) the foreground components (for instance the Galaxy) will always appear the same. Only the noise component \mathcal{N} will change. This has important consequences. In order to discuss this issue, it is useful to start with the case of noiseless observations that can be thought to reproduce a situation of very high signal-to-noise ratio. In this case, model (2) becomes

$$\mathbf{S}^{(i)} = \mathfrak{E}_c + \mathbf{S}_f^{(i)}. \quad (17)$$

3. Statistical characteristics of ILC: noiseless observations

A common assumption in CMB observations is that $\mathbf{S}_f^{(i)}$ is given by the linear mixture of the contribution of N_c physical processes $\{\mathfrak{E}_j\}_{j=1}^{N_c}$ (e.g., free-free, dust re-radiation, ...)

$$\mathbf{S}_f^{(i)} = \sum_{j=1}^{N_c} a_{ij} \mathfrak{E}_j, \quad (18)$$

with a_{ij} constant coefficients. In practice, it is assumed that for the j th physical process a template \mathfrak{E}_j exists independent of the specific channel “ i ”. Although rather strong, it is not unrealistic to assume that this condition is satisfied when small enough patches of the sky are considered. Inserting Eq. (18) into Eq. (17) one obtains

$$\mathbf{S} = \mathbf{A} \mathfrak{E}, \quad (19)$$

with

$$\mathfrak{E} = \begin{pmatrix} \mathfrak{E}_c \\ \mathfrak{E}_1 \\ \mathfrak{E}_2 \\ \vdots \\ \mathfrak{E}_{N_c} \end{pmatrix}, \quad (20)$$

and

$$\mathbf{A} = \begin{pmatrix} 1 & a_{11} & a_{12} & \dots & a_{1N_c} \\ 1 & a_{21} & a_{22} & \dots & a_{2N_c} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_{N_o1} & a_{N_o2} & \dots & a_{N_oN_c} \end{pmatrix}. \quad (21)$$

Here, matrix \mathbf{A} , assumed to be of full rank, is shown partitioned in a way that will be useful for later calculations.

In the following, three cases are considered that correspond to possible experimental situations.

3.1. Case $N_o = N_c + 1$

If $N_o = N_c + 1$, i.e. when number of observations is equal to the number of the components (CMB included), then \mathbf{A} is a square $N_o \times N_o$ matrix. In this case,

$$\mathbf{C}_S = \mathbf{A} \mathbf{C}_\mathfrak{E} \mathbf{A}^T \quad (22)$$

with $\mathbf{C}_\mathfrak{E}$ the $N_o \times N_o$ cross-covariance matrix of the random processes that generate the templates \mathfrak{E} , i.e.

$$\mathbf{C}_\mathfrak{E} = E[\mathfrak{E} \mathfrak{E}^T]. \quad (23)$$

If this equation is inserted in Eq. (11), one obtains

$$\widehat{\mathfrak{E}}_c = \alpha \mathbf{1}^T \mathbf{A}^{-T} \mathbf{C}_\mathfrak{E}^{-1} \mathfrak{E}, \quad (24)$$

with the scalar α given by

$$\alpha = [\mathbf{1}^T \mathbf{A}^{-T} \mathbf{C}_\mathfrak{E}^{-1} \mathbf{A}^{-1} \mathbf{1}]^{-1}, \quad (25)$$

and $\mathbf{A}^{-T} \equiv (\mathbf{A}^{-1})^T$. Now, since it is trivially verified that

$$\mathbf{1}^T = \mathbf{e}_1^T \mathbf{A}^T \quad (26)$$

with

$$\mathbf{e}_1 \equiv (1, 0, \dots, 0)^T, \quad (27)$$

it is

$$\mathbf{1}^T \mathbf{A}^{-T} = \mathbf{e}_1^T. \quad (28)$$

Hence, $\alpha = (\mathbf{C}_\mathfrak{E}^{-1})_{11} = (E[\mathfrak{E}_c \mathfrak{E}_c^T])^{-1} = \sigma_{cc}^{-1}$ is the inverse of the expected variance of the CMB template. As a consequence, if the random process that generates the template \mathfrak{E}_c is uncorrelated with those that generate the templates $\{\mathfrak{E}_j\}$ (in general, this is a reasonable assumption), i.e. if

$$\mathbf{C}_\mathfrak{E} = \begin{pmatrix} \sigma_{cc} & 0 & 0 & \dots & 0 \\ 0 & \sigma_{11} & \sigma_{12} & \dots & \sigma_{1N_o} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \sigma_{N_o1} & \sigma_{N_o2} & \dots & \sigma_{N_oN_o} \end{pmatrix}, \quad (29)$$

from Eq. (24) and the fact that $\mathbf{C}_\mathfrak{E}^{-1}$ has the form

$$\mathbf{C}_\mathfrak{E}^{-1} = \begin{pmatrix} \sigma_{cc}^{-1} & 0 & 0 & \dots & 0 \\ 0 & & & & \\ \vdots & & \mathbf{\Omega}^{-1} & & \\ 0 & & & & \end{pmatrix}, \quad (30)$$

with $\mathbf{\Omega}$ the bottom-right block of the matrix in the rhs of Eq. (29), one obtains that

$$\widehat{\mathfrak{E}}_c = \mathfrak{E}_c. \quad (31)$$

Here, no use is made of the operator $E[.]$ since we are dealing with fixed realizations of the random processes that generate the CMB as well the Galactic components. Therefore, if matrix \mathbf{C}_S is known, the ILC solution is exact.

This condition changes if matrix \mathbf{C}_S has to be estimated through Eq. (13) and a general treatment of this problem is quite difficult. For this reason, we consider the case of observations that span a sky area much wider than the correlation lengths of the maps $\{\mathbf{S}^{(i)}\}$. Under this condition, $\widehat{\mathbf{C}}_\mathfrak{E} = \mathfrak{E} \mathfrak{E}^T / N_p$ can be written in the form

$$\widehat{\mathbf{C}}_\mathfrak{E} = \mathbf{C}_\mathfrak{E} + \Delta \mathbf{C}_\mathfrak{E} \quad (32)$$

² The expression $E[a|b]$ indicates *conditional expectation* of a given b .

with

$$\Delta \mathbf{C}_{\mathfrak{E}} = \begin{pmatrix} \delta_{cc} & \delta_{c1} & \delta_{c2} & \dots & \delta_{cN_o} \\ \delta_{c1} & \delta_{11} & \delta_{12} & \dots & \delta_{1N_o} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_{cN_o} & \delta_{1N_o} & \delta_{2N_o} & \dots & \delta_{N_o N_o} \end{pmatrix} \quad (33)$$

a perturbing matrix with small zero-mean entries. Because of this, it can be expanded in a Taylor series around $\mathbf{C}_{\mathfrak{E}}^{-1}$ up to the linear term obtaining

$$\widehat{\mathbf{C}}_{\mathfrak{E}}^{-1} \approx \mathbf{C}_{\mathfrak{E}}^{-1} - \mathbf{C}_{\mathfrak{E}}^{-1} \Delta \mathbf{C}_{\mathfrak{E}} \mathbf{C}_{\mathfrak{E}}^{-1}, \quad (34)$$

or

$$\widehat{\mathbf{C}}_{\mathfrak{E}}^{-1} \approx \left(\begin{array}{c|c} \sigma_{cc}^{-1} - \sigma_{cc}^{-2} \delta_{cc} & -\sigma_{cc}^{-1} \delta^T \Omega^{-1} \\ \hline -\sigma_{cc}^{-1} \Omega^{-1} \delta & \Omega^{-1} - \Omega^{-1} \Theta \Omega^{-1} \end{array} \right), \quad (35)$$

with $\delta = (\delta_{c1}, \delta_{c2}, \dots, \delta_{cN_o})^T$ and Θ the bottom-right block of the matrix on the rhs of Eq. (33)³. From this result and from Eqs. (24), (27) and (28), one obtains that

$$\widehat{\mathfrak{E}}_c = (1, \delta_{\mathfrak{E}}) \mathfrak{E}, \quad (36)$$

or

$$\Delta \mathfrak{E}_c = \widehat{\mathfrak{E}}_c - \mathfrak{E}_c = (0, \delta_{\mathfrak{E}}) \mathfrak{E}, \quad (37)$$

with $\delta_{\mathfrak{E}}$ a row vector given by

$$\delta_{\mathfrak{E}} = -\frac{\sigma_{cc}}{\sigma_{cc} - \delta_{cc}} \delta^T \Omega^{-1} \approx -\delta^T \Omega^{-1}. \quad (38)$$

Hence, not unexpectedly, the ILC solution differs from the true one by an amount that depends on the sample correlation between \mathfrak{E}_c (the CMB) and the Galactic templates.

3.2. Case $N_o < N_c + 1$

If $N_o < N_c + 1$, the number of the components (CMB included) is larger than the number of channels. Since in this case \mathbf{A} is a rectangular $N_o \times (N_c + 1)$ matrix, the inverse \mathbf{A}^{-1} is not defined. Hence, the ILC solution $\widehat{\mathfrak{E}}_c$ as given by Eq. (11) cannot be written in the form (24). The only possibility to still have $\widehat{\mathfrak{E}}_c = \mathfrak{E}_c$ is that

$$\alpha \mathbf{1}^T (\mathbf{A} \mathbf{C}_{\mathfrak{E}}^{-1} \mathbf{A}^T)^{-1} \mathbf{A} = \mathbf{e}_1^T. \quad (39)$$

In the present case, there is no reason to expect that this condition will be satisfied. As a consequence, the ILC solution differs from the true one by an amount given by

$$\Delta \mathfrak{E}_c = [\alpha \mathbf{1}^T (\mathbf{A} \mathbf{C}_{\mathfrak{E}} \mathbf{A}^T)^{-1} \mathbf{A} - \mathbf{e}_1^T] \mathfrak{E}. \quad (40)$$

This result implies that, similarly to other techniques (e.g., Generalized Least Squares), it is risky to use the ILC technique in situations where the number of observing channels is smaller than the number of components. Unfortunately, the importance of $\Delta \mathfrak{E}_c$ depends on the specific characteristics of both \mathbf{A} and $\mathbf{C}_{\mathfrak{E}}$

³ The entries of $\Delta \mathbf{C}_{\mathfrak{E}}$ are not independent. If \mathbf{C} is a symmetric, positive definite matrix, then the same holds for its inverse \mathbf{C}^{-1} . However, in general, this is not true for a matrix $\mathbf{C} = \mathbf{C} + \Delta \mathbf{C}$ obtained perturbing \mathbf{C} with an arbitrary matrix. This means that the entries of $\Delta \mathbf{C}$ have to satisfy certain conditions. If \mathbf{C} is written in the form $[(\mathbf{Q} + \Delta \mathbf{Q})(\mathbf{Q} + \Delta \mathbf{Q})^T]$, with $\mathbf{Q} = \mathbf{C}^{1/2}$ and $\Delta \mathbf{Q}$ a matrix of zero-mean random quantities, and then expanded in Taylor series, one can find the same result as in Eq. (34) with $\Delta \mathbf{C} = (\Delta \mathbf{Q}) \mathbf{Q}^T + \mathbf{Q} (\Delta \mathbf{Q})^T$.

and, hence, a general treatment is not possible. However, when the Galactic component is the dominant one, it can even happen that $\Delta \mathfrak{E}_c > \mathfrak{E}_c$. For example, in the hypothesis that $\mathbf{C}_{\mathfrak{E}} = \sigma^2 \mathbf{I}$ and

$$\mathbf{A} = \begin{pmatrix} 1.0 & 2.0 & 3.0 \\ 1.0 & 2.5 & 3.5 \end{pmatrix}, \quad (41)$$

Eq. (40) provides

$$\Delta \mathfrak{E}_c = (-0.33, -0.33, 0.33) \mathfrak{E}. \quad (42)$$

This example does not illustrate a situation excessively unfavorable for the separation. In fact, close to the Galactic plane the CMB component is expected to be largely dominated by the Galactic emission.

3.3. Case $N_o > N_c + 1$

If $N_o > N_c + 1$, the number of components (CMB included) is smaller than the number of channels. As in the previous section, \mathbf{A} is a $N_o \times (N_c + 1)$ rectangular matrix but now we face a more difficult situation. The ILC solution $\widehat{\mathfrak{E}}_c$ cannot be written in the form (11) since matrix \mathbf{C}_S is singular. The only way out is to resort to the *pseudo-inverse* \mathbf{C}_S^\dagger

$$\mathbf{C}_S^\dagger = \mathbf{V} \mathbf{D}^\dagger \mathbf{V}^T. \quad (43)$$

Here, \mathbf{V} is the orthogonal matrix obtained from the *singular value decomposition* of \mathbf{C}_S ,

$$\mathbf{C}_S = \mathbf{V} \mathbf{D} \mathbf{V}^T, \quad (44)$$

\mathbf{D}^\dagger is a diagonal matrix whose entries are given by $d_{ii}^\dagger = d_{ii}^{-1}$ if $d_{ii} \neq 0$, 0 otherwise, and $\{d_{ii}\}$ are the elements of the diagonal matrix \mathbf{D} . In this case, the ILC solution is given by

$$\widehat{\mathfrak{E}}_c = \alpha \mathbf{1}^T \mathbf{C}_S^\dagger \mathbf{S}, \quad (45)$$

where

$$\alpha = [\mathbf{1}^T \mathbf{C}_S^\dagger \mathbf{1}]^{-1}. \quad (46)$$

If \mathbf{A} is a full column-rank matrix, then Eq. (45) can be rewritten in the form

$$\widehat{\mathfrak{E}}_c = \alpha \mathbf{1}^T (\mathbf{A}^\dagger)^T \mathbf{C}_{\mathfrak{E}}^{-1} \mathfrak{E}, \quad (47)$$

where

$$\alpha = [\mathbf{1}^T (\mathbf{A}^\dagger)^T \mathbf{C}_{\mathfrak{E}}^{-1} \mathbf{A}^\dagger \mathbf{1}]^{-1}, \quad (48)$$

and

$$\mathbf{A}^\dagger = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T, \quad (49)$$

and

$$(\mathbf{A}^\dagger)^T = \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1}. \quad (50)$$

Now, since it is trivially verified that $\mathbf{1}^T = \mathbf{e}_1^T \mathbf{A}^T$, one obtains that

$$\mathbf{1}^T (\mathbf{A}^\dagger)^T = \mathbf{e}_1^T, \quad (51)$$

and then, from Eq. (45), $\widehat{\mathfrak{E}}_c = \mathfrak{E}_c$. In other words, if matrix \mathbf{C}_S is known, in this case the ILC method also provides an exact separation. When $\widehat{\mathbf{C}}_S$ has to be used, results similar to those presented in Sect. 3.1 are obtained.

4. Statistical characteristics of ILC: noisy observations

In situations of maps contaminated by measurement noise \mathcal{N} , assumed to be the realization of zero-mean stationary stochastic processes, model (19) becomes

$$\mathbf{S} = \mathbf{A}\boldsymbol{\xi} + \mathcal{N}. \quad (52)$$

In turn, under the condition of \mathcal{N} uncorrelated with \mathbf{S} , \mathbf{C}_S becomes

$$\mathbf{C}_S = \mathbf{A}\mathbf{C}_\xi\mathbf{A}^T + \boldsymbol{\Omega}_N, \quad (53)$$

with

$$\boldsymbol{\Omega}_N = \mathbf{E}[\mathcal{N}\mathcal{N}^T]. \quad (54)$$

If noise is small and $N_o = N_c + 1$, inserting Eq. (53) in Eq. (11) and expanding the resulting expression in a Taylor series up to the linear term, it is possible to see that

$$\widehat{\boldsymbol{\xi}}_c = \alpha^*[\mathbf{1}^T - \gamma]\mathbf{A}^{-T}\mathbf{C}_\xi^{-1}\boldsymbol{\xi}, \quad (55)$$

with $\gamma = \sigma_{cc}^{-1}\mathbf{e}_1^T\mathbf{A}^{-1}\boldsymbol{\Omega}_N$ and $\alpha^* = 1/[\sigma_{cc}^{-1} - \sigma_{cc}^{-2}\mathbf{e}_1^T\mathbf{A}^{-1}\boldsymbol{\Omega}_N\mathbf{A}^{-T}\mathbf{e}_1]$. A similar result holds when $N_o > N_c + 1$ with $(\mathbf{A}^\dagger)^T$ substituting \mathbf{A}^{-T} . From Eq. (55) it is clear that $\mathbf{E}[\widehat{\boldsymbol{\xi}}_c|\mathbf{A}\boldsymbol{\xi}] \neq \boldsymbol{\xi}_c$, i.e. the ILC estimator is biased. Since both γ and α^* depend on the specific characteristics of \mathbf{A} and $\boldsymbol{\Omega}_N$, this does not allow a general treatment of the question. However, especially in the case of a high level of noise contamination, the bias affecting $\widehat{\boldsymbol{\xi}}_c$ is expected to be severe.

It is possible to obtain an unbiased ILC estimator computing the weights \mathbf{w} by means of Eq. (11) with the cross-covariance matrix \mathbf{C}_S substituted by

$$\mathbf{C}_S^* = \mathbf{C}_S - \boldsymbol{\Omega}_N. \quad (56)$$

This operation, however, has a cost: the weights \mathbf{w} no longer minimize the variance of $\{\widehat{\boldsymbol{\xi}}_c(p)\}$. As a consequence, the influence of noise can be dramatically amplified. The reason for this result lies in the condition number, $\kappa(\mathbf{C}_S^*)$, of \mathbf{C}_S^* that tends to be larger than $\kappa(\mathbf{C}_S)$. Because of this, the entries of $(\mathbf{C}_S^*)^{-1}$ can take values in a range much wider than those of \mathbf{C}_S^{-1} and the same holds for the corresponding weights \mathbf{w} . According to a theorem of Weyl (e.g., see [Bhatia 1997](#), Theorem III.2.2), it is

$$\lambda_1(\mathbf{C}_S) - \lambda_1(\boldsymbol{\Omega}_N) \leq \lambda_1(\mathbf{C}_S^*) \leq \lambda_1(\mathbf{C}_S) - \lambda_{N_o}(\boldsymbol{\Omega}_N), \quad (57)$$

$$\lambda_{N_o}(\mathbf{C}_S) - \lambda_1(\boldsymbol{\Omega}_N) \leq \lambda_{N_o}(\mathbf{C}_S^*) \leq \lambda_{N_o}(\mathbf{C}_S) - \lambda_{N_o}(\boldsymbol{\Omega}_N), \quad (58)$$

where $\lambda_i(\mathbf{H})$ denotes the i th eigenvalues of a $N \times N$ symmetric, positive definite matrix \mathbf{H} with $\lambda_1 \geq \lambda_2 \dots \geq \lambda_N > 0$. These inequalities indicate a high probability that $\lambda_{N_o}(\mathbf{C}_S^*)/\lambda_{N_o}(\mathbf{C}_S) < \lambda_1(\mathbf{C}_S^*)/\lambda_1(\mathbf{C}_S)$. In this way,

$$\kappa(\mathbf{C}_S^*) = \frac{\lambda_1(\mathbf{C}_S^*)}{\lambda_{N_o}(\mathbf{C}_S^*)} > \frac{\lambda_1(\mathbf{C}_S)}{\lambda_{N_o}(\mathbf{C}_S)} = \kappa(\mathbf{C}_S). \quad (59)$$

For example, this is strictly true if $\boldsymbol{\Omega}_N$ is a diagonal matrix with non-zero entries equal to a constant value ϱ , since

$$\kappa(\mathbf{C}_S^*) = \frac{\lambda_1(\mathbf{C}_S) - \varrho}{\lambda_{N_o}(\mathbf{C}_S) - \varrho}. \quad (60)$$

From these consideration, in the case of a low SNR and large $\kappa(\mathbf{C}_S^*)$, one has to expect significant amplification of the noise

for the unbiased ILC estimator. However, matrix \mathbf{C}_S^* is very ill-conditioned only in situations where the images $\{\mathcal{S}^{(i)}\}$ are very similar. For example, in CMB applications this could happen if the maps were obtained at very close frequencies, and this should be considered when planning the experiment. As a consequence, in practical applications it is advisable to check whether the resulting $\kappa(\mathbf{C}_S^*)$ permits the use of the unbiased ILC estimator. The benefit is that noise filtering is typically an easier operation than bias removal. Even if \mathbf{C}_S^* is ill-conditioned, as it happens when $N_o > N_c + 1$, its κ can be reduced by means of the SVD technique presented in Sect. 3.3.

These conclusions hold only in situations of isotropic noise (i.e. noise with identical characteristics everywhere). In the contrary case, poor results are to be expected since noise behaves as an additional channel-dependent component.

5. Conclusions

The arguments presented in the previous sections show that for a reliable use of the ILC estimator some conditions have to be satisfied. In particular:

1. The observations have to cover a sky area much wider than the spatial scale (correlation length) of the observed maps in such a way as to allow an accurate estimate of the cross-covariance matrix \mathbf{C}_S .
2. Model (19)–(21) has to hold. In particular, ILC cannot be expected to produce satisfactory results if some of the Galactic templates depend on the observing channels. This point can be realized if a Galactic channel-dependent template $\boldsymbol{\xi}_j$ is thought of as the sum of N_j channel-independent templates. In this case, model (19)–(21) still holds but with an effective number of Galactic components that now is $N_c^* = \sum_{j=1}^{N_c} N_j$. Since the number of observing channel is typically rather limited, in practical applications it has to be expected that $N_o < N_c^* + 1$ and then the point below applies.
3. The condition $N_o \geq N_c^* + 1$ has to hold, i.e. the number of observing channels has to be equal to or larger than the number of the physical components (CMB included) that contribute to form the observed maps. In the contrary case, the solution can suffer severe distortion.
4. In the observed maps the level of the noise has to be rather low otherwise a severe bias can be introduced in the ILC estimator. This can be removed but at the price of an amplification of the noise level in the final result. However, especially in situations of high SNR, the use of the unbiased ILC estimator can offer some advantages.

Thus, using the ILC estimator is not trivial. In particular, the first two points are in conflict with each other. In the case of wide maps, model (19)–(21) is not applicable since on large spatial scales the Galactic templates are expected to depend on the observing channel. In this respect, a simple tests can be of help that is based on the analysis of the eigenvalues of $\widehat{\mathbf{C}}_S$. If $N_o > N_c^* + 1$, matrix $\widehat{\mathbf{C}}_S$ has to be (almost) singular with a number of (almost) zero elements in the diagonal matrix $\widehat{\mathbf{D}}$ of Eq. (44) equal to $N_o - N_c^* - 1$. Unfortunately, this test can be influenced by the noise; because of the statistical fluctuations, the entries of $\widehat{\mathbf{D}}$ that should be close to zero can take larger values.

The obvious conclusion is that, in order to obtain reliable results, the use of ILC requires a careful planning of the observations. The area of the sky to observe as well as the acceptable level of the noise are factors that have to be fixed in

advance. An “a posteriori” correction of the distortions introduced in the ILC solution by the violation of the above conditions is unreliable as all of these distortions critically depend on the true solution that one tries to estimate. Therefore, they cannot be obtained from the data only. The alternative represented by the numerical simulations that make use of semi-empirical templates is also unreliable. In particular, there is the possibility of introducing spurious features in the final results. This holds also when these results are used as “priors” in more sophisticated separation techniques.

As the use of ILC in the Fourier domain provides the same results as those obtainable in the spatial one, it could seem that this approach does not offer particular benefit. This can be false if the maps have different spatial resolutions (i.e. the observing channels have different point spread functions) and/or whether a frequency-dependent separation is desired (Tegmark et al. 2003). Although in this way it is possible to improve some properties of the separated maps (e.g., the spatial resolution), this has a cost in the amplification of the noise level.

Acknowledgements. We thank Dr. Carlo Baccigalupi for careful reading of this manuscript.

References

- Baccigalupi, C., Perrotta, F., de Zotti, G., et al. 2004, *MNRAS*, 354, 55
 Bhatia, R. 1997, *Matrix Analysis* (Berlin: Springer)
 Bennett, C. L., Hill, R. S., Hinshaw, G., et al. 2003, *ApJS*, 148, 97
 Cardoso, J. F. 1999, *Neural Computation*, 11, 157
 Delabrouille, J., & Cardoso, J. F. 2007 [[arXiv:astro-ph/0702198v2](https://arxiv.org/abs/astro-ph/0702198v2)]
 Eriksen, H. K., Banday, A. J., Górski, K. M., & Lilje, P. B. 2004, *ApJ*, 612, 633
 Hinshaw, G., Nolta, M. R., Bennett, C. L., et al. 2007, *ApJS*, 170, 288
 Hyvärinen, A., Karhunen, J., & Oja, E. 2001, *Independent Component Analysis* (New York: John Wiley & Sons)
 Maino, D., Farusi, A., Baccigalupi, C., et al. 2002, *MNRAS*, 334, 53
 Saha, R., et al. 2007 [[arXiv:0706.3567v1](https://arxiv.org/abs/0706.3567v1)]
 Stivoli, F., Baccigalupi, C., Maino, D., & Stompor, R. 2006, *MNRAS*, 372, 615
 Tegmark, M., de Oliveira-Costa, A., & Hamilton, A. J. 2003, *Phys. Rev. D.*, 68, 123523