

Regularization of inverse problems with adaptive discrepancy terms; application to multispectral data

Sandrine Anthoine

I3S, Université de Nice Sophia-Antipolis, CNRS;
2000 route des Lucioles, 06903 Sophia-Antipolis Cedex, France

ABSTRACT

In this paper, a general framework for the inversion of a linear operator in the case where one seeks several components from several observations is presented. The estimation is done by minimizing a functional balancing discrepancy terms by regularization terms. The regularization terms are adapted norms that enforce the desired properties of each component.

The main focus of this paper is the definition of the discrepancy terms. Classically, these are quadratic. We present novel discrepancy terms adapt to the observations. They rely on adaptive projections that emphasize important information in the observations. Iterative algorithms to minimize the functionals with adaptive discrepancy terms are derived and their convergence and stability is studied.

The methods obtained are compared for the problem of reconstruction of astrophysical maps from multifrequency observations of the Cosmic Microwave Background. We show the added flexibility provided by the adaptive discrepancy terms.

Keywords: inverse problems, iterative algorithm, adaptive discrepancy terms, wavelets, multispectral astrophysical data.

1. INTRODUCTION

In a general inverse problem, the goal is to estimate an object F from an observation G where $G=T(F)$ (for example looking for the original from a blurred picture). Here we assume that T , the linear operator modeling the observation process, is known. Even so, the problem is often ill-posed and therefore needs to be "regularized". This amounts to finding an estimate \tilde{F} which has the two following properties:

1. \tilde{F} should generate observations close to the data ($T(\tilde{F}) \approx G$).
2. \tilde{F} should have properties we expect from a priori knowledge. (In the example of the blurred picture, we expect \tilde{F} to have sharp features.)

A number of classical methods for solving inverse problems try to balance the fitness to the data ($T(\tilde{F}) \approx G$), measured by a discrepancy term J_{disc} , with the regularity of the solution (i.e. the properties of \tilde{F}), measured a regularization term J_{reg} , by minimizing a cost functional of the type:

$$J(F) = J_{disc}(T(F), G) + \alpha J_{reg}(F) \quad (1)$$

The main focus of this paper is the definition of the discrepancy term. Generally, this term is chosen to be the L^2 norm of the residual $G - T(F)$ i.e. $J_{disc}(T(F), G) = \|G - T(F)\|_{L^2}^2 = \int |G(x) - T(F(x))|^2 dx$ or a similar quadratic norm. Such a term treats uniformly all the elements in G (for example all the pixels in an image). However, it is clear that some part of G carry more information than others and should therefore be treated differently. For example, in a blurred picture, one is usually more interested in recovering well the face of a person than the uniform part of the sky. A simple quadratic norm is not able to identify such important features and thus may be improved. J.-L. Starck and co-authors in¹ present an iterative algorithm that focuses on these important features in blurred astrophysical images by introducing projections on the "multiresolution support". These are projections on a subspace defined by the wavelet transform of the observations. They are adaptive and allow to consider only important features of the data and discard the noise in the case of deconvolution of astrophysical data presented in¹. Following this idea, we propose the use of general projections to

Email: anthoine@i3s.unice.fr

define adaptive discrepancy measures. The idea is that the image space of the projection defines important features in the observation - these should be well predicted by the estimate of F - while the kernel of the projection defines information that is less important or even not relevant (for example noise in the observation). Using the mathematical framework introduced in², we study the mathematical properties of the resulting algorithms. We show that convergence is guaranteed and that stability holds in a certain sense. However we point out that the use of projections may result in a loss of information that prevents to recover some parts of the data. We show that this can be remedied by introducing the notion of “relaxed projection”, which consists in only down-weighting the importance of the non-feature space instead of cancelling it.

After the present introduction, this paper is organized as follows. In the second section we introduce the notations as well as the variational framework we use, i.e. the class of functionals we seek to minimize as well as an iterative algorithm to solve the minimization. The third section concerns the generalization of discrepancy terms via the use of projections and its mathematical study. The potential loss of information induced by the projections is remedied in section four by the introduction of relaxed projections. Finally, section five is devoted to the application of the resulting algorithms to the estimation of astrophysical maps from multifrequency observations.

2. A CLASS OF VARIATIONAL FUNCTIONAL TO REGULARIZE INVERSE PROBLEMS

2.1 Notations

2.1.1 Inverse problem with several objects and observations

The most general problem we will consider is the case where we seek M objects or components f_1, \dots, f_M from L observations g_1, \dots, g_L . In the case of the estimation of astrophysical maps from multifrequency observations, each object f_i is the map of an astrophysical phenomena (ex: the map of galaxy clusters) and each g_l is an observation of the sky at wavelength ν_l .

We make the following assumptions:

- Each object belongs to a Hilbert space $\mathcal{H}_m^i: \forall m = 1..M, f_m \in \mathcal{H}_m^i$.
- Each observation belongs to a Hilbert space $\mathcal{H}_l^o: \forall l = 1..L, g_l \in \mathcal{H}_l^o$.
- We know the linear bounded operators $T_{m,l}: \mathcal{H}_m^i \rightarrow \mathcal{H}_l^o$

such that the model for the observations is linear with additive noise:

$$\forall l = 1..L, \quad g_l = \sum_{m=1}^M T_{m,l} f_m + n_l \quad (2)$$

where n_l are noise terms.

To estimate the objects f_1, \dots, f_M from g_1, \dots, g_L , we will minimize functionals composed of a sum of discrepancy terms (one per observation) and regularization terms (one per component) such as:

$$J(f_1, f_2, \dots, f_M) = \sum_{l=1}^L \rho_l \left\| \left(\sum_{m=1}^M T_{m,l} f_m - g_l \right) \right\|_{\mathcal{H}_l^o}^2 + \sum_{m=1}^M \gamma_m \|f_m\|_{X_m}; \quad (3)$$

where the γ_m and ρ_l are strictly positive scalars and the “norms” $\|\cdot\|_{X_m}$ are of the form:

$$\|f\|_{X_m} = \sum_{\lambda \in \Lambda} w_\lambda^m |\langle f, \varphi_\lambda^m \rangle|^{p_m} \quad (4)$$

where for all m , $\varphi^m = \{\varphi_\lambda^m\}_{\lambda \in \Lambda}$ is a generating family of \mathcal{H}_m^i , $w_\lambda^m > 0$ and $1 \leq p_m \leq 2$.

The discrepancy terms (first sum in Eq.(3)) are classical quadratic terms. The particular form of the $\|\cdot\|_{X_m}$ is chosen so that one can adapt the regularization terms (second sum in Eq.(3)) to the properties of each object. Indeed, for each m , one chooses the decomposition system $\varphi^m = \{\varphi_\lambda^m\}_{\lambda \in \Lambda}$ on which to measure the smoothness of the m th object, as well as the type of l_p measure and weights needed. This gives a large panel of available smoothness measures that can fit various kind of data. For example, the Cosmic Microwave Background signal, which is the relic radiation of our Universe, is well modeled by a Gaussian process with known spectral power P . The Gaussianity leads to a quadratic measure, while the power spectrum can be enforced in Fourier space. Therefore, an adapted measure is $\sum_k P(k)^{-1} |\langle f, \exp(-2\pi jk) \rangle|^2$. As for galaxy clusters, these being rare, small and intense objects, the wavelet transform of such a map is sparse (only a few coefficients of large amplitude). Therefore, an adapted term is the l_1 norm of its wavelet coefficients: $\sum_{j,k} |\langle f, \psi_{j,k} \rangle|$.

2.1.2 Simplifying notations: case of one object and observation

One can simplify greatly the notations by "vectorizing" the previous problem as follows:

find one object $F = (f_1, f_2, \dots, f_M)^T$ from one observation $G = (g_1, g_2, \dots, g_L)^T$, knowing $T = \{T_{m,l}\}_{m,l}$, the linear operator such that:

$$G = TF + N \quad (5)$$

The functional in Eq.(3) is then:

$$J(F) = \|TF - G\|_{\mathcal{H}^o}^2 + \gamma \|F\|; \quad (6)$$

where the norm in Hilbert space \mathcal{H}^o is the weighted Euclidean norm: $\|G\|_{\mathcal{H}^o}^2 = \sum_{l=1}^L \rho_l \|G_l\|_{\mathcal{H}^o}^2$ and $\|F\|$ is the mixed norm $\|F\| = \sum_{m=1}^M \gamma_m \|f_m\|_{X_m} = \sum_{m=1}^M \gamma_m \sum_{\lambda \in \Lambda} w_\lambda^m |\langle f, \varphi_\lambda^m \rangle|^{p_m}$.

Written this way, Eq.(6) is very close to Eq.(3) with $M=L=1$, which reads:

$$J(f) = \|Tf - g\|_{\mathcal{H}^o}^2 + \gamma \|f\| = \|Tf - g\|_{\mathcal{H}^o}^2 + \gamma \sum_{\lambda \in \Lambda} w_\lambda |\langle f, \varphi_\lambda \rangle|^p \quad (7)$$

It is true that the weighted norm induced on \mathcal{H}^o makes it a standard Hilbert space, hence the discrepancy terms do match perfectly. But the regularization terms do not match: for $M=1$, we get in Eq. (7) a simple l_p sum (with a single exponent p), which is not true for $M>1$.

However, the minimization of Eq.(6) can be done by slightly modifying the iterative algorithm that we use for Eq.(7) and moreover, the proofs of convergence and stability carry to this more complicated case (see³ for details). Since in this paper we are concerned with modifying the discrepancy term, we will only present the theory with a simple regularization term as in Eq.(7) to alleviate the notations, keeping in mind the mixed regularization terms for the application.

Note that we will use the following notations:

- the sequence of weight is: $\mathbf{w} = \{w_\lambda\}_{\lambda \in \Lambda}$.
- the functional in Eq.(7) is $\mathbf{J}_{\gamma, \mathbf{w}, p}(f) = \|Tf - g\|_{\mathcal{H}^o}^2 + \gamma \sum_{\lambda \in \Lambda} w_\lambda |\langle f, \varphi_\lambda \rangle|^p$.
- the scalar product is: $f_\lambda = \langle f, \varphi_\lambda \rangle$.
- we call $\|\cdot\|_{\mathbf{w}, p}$ -norm the quantity: $\sum_{\lambda \in \Lambda} w_\lambda |\langle \cdot, \varphi_\lambda \rangle|^p$.

2.2 A class of functionals and the study of their minimization

In this section, we summarize the findings in², which concern the minimization of Eq.(7), i.e. of the functional $\mathbf{J}_{\gamma, \mathbf{w}, p}$.

2.2.1 Iterative algorithm

The authors propose the following iterative algorithm to obtain a minimizer:

ALGORITHM 2.1.

$$\begin{cases} f^0 & \text{arbitrary} \\ f^n & = \mathbf{S}_{\gamma, \mathbf{w}, p}(f^{n-1} + T^*(g - Tf^{n-1})), \quad n \geq 1 \end{cases}$$

At each iteration, one computes the Landweber iterate $f^{n-1} + T^*(g - Tf^{n-1})$ and modifies it with the $\mathbf{S}_{\gamma, \mathbf{w}, p}$ function. The $\mathbf{S}_{\gamma, \mathbf{w}, p}$ treats independently each coefficient of the argument h on the decomposition system $\varphi = \{\varphi_\lambda\}_{\lambda \in \Lambda}$:

$$\mathbf{S}_{\mathbf{w}, p}(h) = \sum_{\lambda} S_{w_\lambda, p}(h_\lambda) \varphi_\lambda, \quad (8)$$

with the functions $S_{w, p}$ from \mathbb{R} to itself given by

$$S_{w, p}(x) \stackrel{\text{def}}{=} \left(x + \frac{wp}{2} \text{sign}(x) |x|^{p-1} \right)^{-1}, \quad \text{for } 1 \leq p \leq 2, \quad (9)$$

where $(\cdot)^{-1}$ denotes the inverse so that $\forall x, S_{w, p}(x + \frac{wp}{2} \text{sign}(x) |x|^{p-1}) = x$.

In particular, for $p = 1$, $S_{w, 1}$ is the soft-thresholding operator:

$$S_{w, 1}(x) = \text{sign}(x)(|x| - w/2)^+$$

whereas for $p = 2$, one simply gets:

$$S_{w, 2}(x) = \frac{x}{1+w}.$$

2.2.2 Convergence and stability

The two following theorems summarize the findings presented in². The first theorem states that the iterative algorithm 2.1 converges strongly in the norm associated in the Hilbert space \mathcal{H}^i for any initial guess f^0 .

THEOREM 2.2 (CONVERGENCE). *Let T be a bounded linear operator from \mathcal{H}^i to \mathcal{H}^o , with $\|T\| < 1$. Take $p \in [1, 2]$, and let $\mathbf{S}_{\mathbf{w},p}$ be the shrinkage operator defined by (8), where the sequence $\mathbf{w} = \{w_\lambda\}_{\lambda \in \Lambda}$ is such that there exists a constant $c > 0$ such that $\forall \lambda \in \Lambda : w_\lambda \geq c$. Then the sequence of iterates*

$$f^n = \mathbf{S}_{\gamma, \mathbf{w}, p}(f^{n-1} + T^*(g - T f^{n-1})), \quad n = 1, 2, \dots,$$

with f^0 arbitrarily chosen in \mathcal{H}^i , converges strongly to a minimizer of the functional $\mathbf{J}_{\gamma, \mathbf{w}, p}$.

If the minimizer f^* of $\mathbf{J}_{\gamma, \mathbf{w}, p}$ is unique, (which is guaranteed e.g. by $p > 1$ or $\ker(T) = \{0\}$), then every sequence of iterates f^n converges strongly to f^* , regardless of the choice of f^0 .

The second theorem is concerned with the stability of the method. It gives sufficient conditions to ensure that the estimate recovered from a perturbed observation, $g = T f_0 + e$, will approximate the object f_0 as the amplitude of the perturbation $\|e\|_{\mathcal{H}^o}$ goes to 0.

THEOREM 2.3 (STABILITY). *Assume that T is a bounded operator from \mathcal{H}^i to \mathcal{H}^o with $\|T\| < 1$, that $\gamma > 0$, $1 \leq p \leq 2$ and that the entries in the sequence $\mathbf{w} = \{w_\lambda\}_{\lambda \in \Lambda}$ are bounded below by $c > 0$.*

Assume that either $p > 1$ or $\ker(T) = \{0\}$. For any $g \in \mathcal{H}^o$ and any $\gamma > 0$, define $f_{\gamma, \mathbf{w}, p; g}^$ to be the minimizer of $\mathbf{J}_{\gamma, \mathbf{w}, p}$ with observation g . If $\gamma = \gamma(\epsilon)$ satisfies*

$$\lim_{\epsilon \rightarrow 0} \gamma(\epsilon) = 0 \quad \text{and} \quad \lim_{\epsilon \rightarrow 0} \frac{\epsilon^2}{\gamma(\epsilon)} = 0, \quad (10)$$

then we have, for any $f_o \in \mathcal{H}^i$,

$$\lim_{\epsilon \rightarrow 0} \left[\sup_{\|g - T f_o\|_{\mathcal{H}^o} \leq \epsilon} \|f_{\gamma(\epsilon), \mathbf{w}, p; g}^* - f^\dagger\|_{\mathcal{H}^i} \right] = 0, \quad (11)$$

where f^\dagger is the unique element of minimum $\|\cdot\|_{\mathbf{w}, p}$ -norm in the set $\mathcal{S}_{f_o} = \{f; T f = T f_o\}$.

Note that in particular when T is invertible, $f^\dagger = f$ which means that Algorithm 2.1 provides a stable inversion.

So far, we have a convergent and regularizing iterative algorithm that converges to a minimizer of the functional $\mathbf{J}_{\gamma, \mathbf{w}, p}$. Such a minimizer is an estimate of the object f that compromises between generating an observation close to the data g in a quadratic sense and having the smallest $\|\cdot\|_{\mathbf{w}, p}$ -norm. Note the the design of the $\|\cdot\|_{\mathbf{w}, p}$ -norm is such that it will preserve or enhance desirable properties of f . The quadratic discrepancy term in $\mathbf{J}_{\gamma, \mathbf{w}, p}$ is devoid of such considerations and therefore does not enhance more important features that should be matched in the observations. In the rest of this paper, we will present adaptive discrepancy terms that aim at fixing this point.

3. ADAPTIVE DISCREPANCY TERMS (I): USING PROJECTIONS

3.1 An Algorithm using Adaptive Projections

3.1.1 Original idea

In¹, the authors are concerned with the deconvolution of an astrophysical image. The observations of interest were blurred and noisy pictures of galaxies. For these, denoising by wavelet-shrinkage was already known to improve the quality of noisy observations. The wavelet shrinkage procedure on g is nothing more than applying to g an adaptive projection: the projection on the ‘‘multiresolution support’’ of g , i.e. on the subspace defined by the largest wavelet coefficients of g . The fact that wavelet shrinkage improves the observation shows that the ‘‘multiresolution support’’ of g naturally defines a subspace that describes the important features of g .

The authors of¹ proposed to use this multiresolution support not only on g itself but also in the context of deblurring by using it to evaluate how well an estimate f fits the data g . They proposed an iterative algorithm very close to Algorithm 2.1, for $p = w_\lambda = 1$ except that the residual $(g - T f^{n-1})$ is projected on the multiresolution support of g : $(g - T f^{n-1})$ becomes $M_g(g - T f^{n-1})$ where M_g is the projection.

We propose to extend this idea to any kind of adaptive projections and study the mathematical properties of the resulting algorithm.

3.1.2 Iterative algorithm with adaptive projection

We first define the notion of adaptive projection: an adaptive projection defined by the data g is the orthogonal projection on a subspace defined by the fact that the coefficients of g on an orthonormal basis are greater than predefined thresholds. Mathematically:

DEFINITION 3.1. Given an orthonormal basis $\{\beta_\lambda\}_{\lambda \in \Lambda}$ of \mathcal{H}^o , an element g in \mathcal{H}^o and a sequence of nonnegative thresholds $\tau = \{\tau_\lambda\}_{\lambda \in \Lambda}$, the adaptive projection $M_{g,\tau}$ is the map from \mathcal{H}^o into itself defined by:

$$\forall h \in \mathcal{H}^o, \quad M_{g,\tau}(h) = \sum_{\lambda \text{ s.t. } |g_\lambda| > \tau_\lambda} h_\lambda \beta_\lambda$$

(where, as usual, h_λ denotes the scalar product $\langle h, \beta_\lambda \rangle$)

We propose the following algorithm:

ALGORITHM 3.2.

$$\begin{cases} f^0 & \text{arbitrary} \\ f^n & = \mathbf{S}_{\gamma, \mathbf{w}, p}(f^{n-1} + T^* M_{g,\tau}(g - T f^{n-1})), \quad n \geq 1 \end{cases}$$

Note that if T is a convolution, $\{\beta_\lambda\}_{\lambda \in \Lambda}$ is a wavelet basis, $p = 1$ and $\forall \lambda \in \Lambda, w_\lambda = 1$, this is what was proposed in¹. From what we saw before, it is straightforward to infer that Algorithm 3.2 should converge to a minimizer of

$$\mathbf{J}_{\gamma, \mathbf{w}, p, \tau}(f) = \|M_{g,\tau}(Tf - g)\|_{\mathcal{H}^o}^2 + \gamma \|f\|_{\mathbf{w}, p} \quad (12)$$

which is a functional with an adaptive discrepancy term.

3.2 Mathematical Properties

3.2.1 A convergent iterative algorithm

The strong convergence of Algorithm 3.2 to a minimizer of Eq. (12) is guaranteed by Theorem 2.2 (under the same conditions as in Theorem 2.2): apply this theorem to $g' = M_{g,\tau} g$ and $T' = M_{g,\tau} T$ to get the solution (this works because $M_{g,\tau}$ is a self-adjoint projection).

3.2.2 Diagonal case: a new kind of thresholding

To gain insight on this algorithm, we first study the case of a diagonal operator T . We assume that

$$T(h) = \sum_{\lambda \in \Lambda} t_\lambda h_\lambda \varphi_\lambda$$

where the t_λ are scalars. As a reminder, for Algorithm 2.1, the minimizer is

$$\operatorname{argmin}(\mathbf{J}_{\gamma, \mathbf{w}, p}) = \mathbf{S}_{\gamma, \mathbf{w}/t^2, p}(T^{-1}g) = \sum_{\lambda \in \Lambda / t_\lambda \neq 0} S_{\gamma w_\lambda / t_\lambda^2, p}(g_\lambda / t_\lambda) \varphi_\lambda.$$

When $p = 1$, this reduces to the soft-thresholded version of $T^{-1}g$ on the basis $\varphi = \{\varphi_\lambda\}_{\lambda \in \Lambda}$ with the thresholds $\gamma w_\lambda / t_\lambda^2$.

When the adaptive discrepancy term is introduced, we get:

$$\begin{aligned} \mathbf{J}_{\gamma, \mathbf{w}, p, \tau}(f) &= \|M_{g,\tau}(Tf - g)\|_{\mathcal{H}^o}^2 + \gamma \|f\|_{\mathbf{w}, p} \\ &= \sum_{\lambda \text{ s.t. } |g_\lambda| > \tau_\lambda} |(Tf - g)_\lambda|^2 + \gamma \sum_{\lambda \in \Lambda} w_\lambda |f_\lambda|^p \\ &= \sum_{\lambda \text{ s.t. } |g_\lambda| > \tau_\lambda} \left(|t_\lambda f_\lambda - g_\lambda|^2 + \gamma w_\lambda |f_\lambda|^p \right) + \gamma \sum_{\lambda \text{ s.t. } |g_\lambda| \leq \tau_\lambda} w_\lambda |f_\lambda|^p \end{aligned} \quad (13)$$

The equations for each f_λ are now decoupled so that the minimizer f^* is defined by:

$$\begin{cases} f_\lambda^* = S_{\gamma w_\lambda / t_\lambda^2, p}(g_\lambda / t_\lambda) & \text{if } |g_\lambda| > \tau_\lambda \quad \text{and} \quad t_\lambda \neq 0 \\ f_\lambda^* = 0 & \text{if } |g_\lambda| \leq \tau_\lambda \quad \text{or} \quad t_\lambda = 0 \end{cases} \quad (14)$$

Introducing the hard-thresholding operator with threshold m :

$$H_\tau(x) = \begin{cases} x & \text{if } |x| > m \\ 0 & \text{otherwise,} \end{cases} \quad (15)$$

one can rewrite the preceding equation:

$$\begin{cases} f_\lambda^* = S_{\gamma w_\lambda / t_\lambda^2, p}(H_{\tau_\lambda / t_\lambda}(g_\lambda)) & \text{if } t_\lambda \neq 0 \\ f_\lambda^* = 0 & \text{if } t_\lambda = 0. \end{cases} \quad (16)$$

Thus we obtain the previous shrinkage operator $S_{\gamma w_\lambda / t_\lambda^2, p}$ composed with a hard-thresholding operator $H_{\tau/t}$ that we call “adaptive thresholding operator”. The hard-thresholding operation is known to be a way to enhance the solution after application of the pseudo inverse. On the other hand the shrinkage operator $S_{\gamma w_\lambda / t_\lambda^2, p}$ regularizes the same solution with respect to a smoothness defined by the $\|\cdot\|_{w,p}$ -norm. We find here that the introduction of the discrepancy term with adaptive projections is simply an intermediate solution between both of these regularizations.

When $p = 1$ and $t_\lambda = 1$, we obtain a compromise between hard and soft-thresholding if $\tau_\lambda > \gamma w_\lambda$. To illustrate this, we graph in Fig.1 the hard-thresholding function with threshold τ (left), the soft-thresholding function with threshold γ (right) and the function obtained in Eq.(16) (middle) in the case $\tau > \gamma$ (here $w = 1$). We give a further illustration of the diagonal case in Fig. 3, section 5.

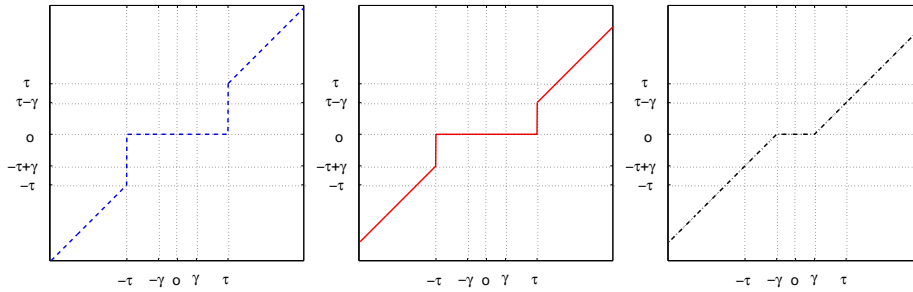


Figure 1. Left: hard-thresholding operator H_τ ; middle: adaptive thresholding operator; right: soft-thresholding operator $S_{\gamma,1}$.

3.2.3 Stability is an issue

The study on diagonal operators suggests that introducing adaptive projections gives flexibility by defining a new shrinkage operator. In this section, we see that this flexibility comes to a price: the resulting algorithm is not stable in the sense of Theorem 2.3. There is stability in the sense that if the parameter (τ, w, \dots) are chosen properly as the noise level decreases - i.e. when the observation g gets closer to the true observation Tf_o - then the solutions converge to a well-defined limit. However this limit is not necessarily f_o , even if T is invertible.

In a nutshell, what happens is that stability requires that the thresholds $\tau = \{\tau_\lambda\}_{\lambda \in \Lambda}$ are large enough compared to $\|g - Tf_o\|$. This implies that the subspace defined by the indexes λ such that $\{Tf_o\}_\lambda = 0$ will necessarily be in the kernel of the adaptive projections $M_{g,\tau}$ as soon as g is close enough to Tf_o . Therefore the information in this subspace will be lost. The result is then that as the observation becomes ideal (i.e. close to Tf_o) the solution of Algorithm 3.2 will approach the element of minimal $\|\cdot\|_{w,p}$ -norm in the set \mathcal{M}_{T,f_o} of elements of \mathcal{H}^i that have the same image under T as f_o except maybe on the coordinates λ such that $(Tf_o)_\lambda = 0$.

Note that even if T is one-to-one, this set is not necessarily reduced to f_o :

EXAMPLE 1. If T is the identity, $f_1 = (1, 0) \in \mathbb{R}^2$, then $\mathcal{M}_{f_1} = \{(1, x), x \in \mathbb{R}\}$ on the canonical basis.

In this case however the minimizer of the $\|\cdot\|_{w,p}$ -norm is f_1 itself whatever the choices of the parameters $\gamma, \tau, w = \{w_\lambda\}_{\lambda \in \Lambda, \dots}$ are. Algorithm 3.2 will therefore provide the desired result. This is not the case in the following example, where T is also an invertible operator in \mathbb{R}^2 :

EXAMPLE 2. Consider $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, the bounded and linear operator defined by:

$$T : \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \mapsto \frac{1}{4} \begin{pmatrix} 2f_1 + f_2 \\ f_1 - f_2 \end{pmatrix} \quad \text{and} \quad f_a = \begin{pmatrix} a \\ a \end{pmatrix} \quad \text{for some } a \neq 0.$$

- T has a bounded inverse: $T^{-1} : \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \mapsto \frac{4}{3} \begin{pmatrix} f_1 + f_2 \\ f_1 - 2f_2 \end{pmatrix}$ and $\|T\| = \frac{1}{2} < 1$.
- $Tf_a = \begin{pmatrix} \frac{3a}{4} \\ 0 \end{pmatrix}$ and $\mathcal{M}_{f_a} = \{f : (Tf)_1 = (Tf_a)_1\} = \{f : 2f_1 + f_2 = 3a\}$.

The element in \mathcal{M}_{f_a} with minimal l^1 norm is: $f_a^\dagger = \begin{pmatrix} \frac{3a}{2} \\ 0 \end{pmatrix}$, and not f_a itself. Thus the minimizers of Eq. (12) do not converge to f_a as the observations converge to Tf_a . In other words, information on the second coordinate in image plane has been lost that prevents the algorithm to invert T even with arbitrary accurate data.

We now formalize this result. We first define \mathcal{M}_{T,f_o} and the set $\mathcal{H}_{T,\mathbf{w},p}^i$ of elements for which \mathcal{M}_{T,f_o} has a unique minimizer of the $\|\cdot\|_{\mathbf{w},p}$ -norm.

DEFINITION 3.3 (\mathcal{M}_{T,f_o}). Given two Hilbert spaces \mathcal{H}^i and \mathcal{H}^o , an operator $T : \mathcal{H}^i \rightarrow \mathcal{H}^o$, an orthonormal basis $\{\beta_\lambda\}_{\lambda \in \Lambda}$ of \mathcal{H}^o and an element f_o of \mathcal{H}^i . The set \mathcal{M}_{T,f_o} is the subset of elements of \mathcal{H}^i that verify:

$$f \in \mathcal{M}_{T,f_o} \iff M_{Tf_o,0}(Tf) = Tf_o \iff \left[\{Tf_o\}_\lambda \neq 0 \Rightarrow \{Tf\}_\lambda = \{Tf_o\}_\lambda \right]$$

DEFINITION 3.4 ($\mathcal{H}_{T,\mathbf{w},p}^i$). Given a Hilbert space \mathcal{H}^i , $\mathcal{H}_{T,\mathbf{w},p}^i$ is the subset of elements of \mathcal{H}^i that verify: f_o is in $\mathcal{H}_{T,\mathbf{w},p}^i$ if and only if the set $\mathcal{M}_{T,f_o} = \{f : M_{Tf_o,0}Tf = Tf_o\}$ has a unique element of minimum $\|\cdot\|_{\mathbf{w},p}$ -norm.

When $p > 1$, then $\mathcal{H}_{T,\mathbf{w},p}^i = \mathcal{H}^i$, regardless of T . This is not true if $p = 1$, even if $\ker T = \{0\}$. It turns out that Algorithm 3.2 is regularizing for elements f in $\mathcal{H}_{T,\mathbf{w},p}^i$, and that the minimizer obtained in the limit $\|Tf_o - g\|_{\mathcal{H}^o}$ goes to zero is exactly the minimizer of the $\|\cdot\|_{\mathbf{w},p}$ -norm in \mathcal{M}_{T,f_o} . This is the object of the following theorem:

THEOREM 3.5. Assume that T is a bounded operator from \mathcal{H}^i to \mathcal{H}^o with $\|T\| < 1$, that $\gamma > 0$, $p \in [1, 2]$ and that the entries in the sequence $\mathbf{w} = \{w_\lambda\}_{\lambda \in \Lambda}$ are bounded below uniformly by a strictly positive number c .

For any $g \in \mathcal{H}^o$ and any $\gamma > 0$ and any nonnegative sequence $\tau = \{\tau_\lambda\}_{\lambda \in \Lambda}$, define $f_{\gamma,\mathbf{w},p,\tau;\gamma}^*$ to be a minimizer of $\mathbf{J}_{\gamma,\mathbf{w},p,\tau}(f)$ with observation g . If $\gamma = \gamma(\epsilon)$ and $\tau = \tau(\epsilon)$ satisfy:

1. $\lim_{\epsilon \rightarrow 0} \gamma(\epsilon) = 0$ and $\lim_{\epsilon \rightarrow 0} \frac{\epsilon^2}{\gamma(\epsilon)} = 0$
2. $\forall \lambda \in \Lambda, \lim_{\epsilon \rightarrow 0} \tau_\lambda(\epsilon) = 0$ and $\exists \delta > 0, \text{ s.t. } [\epsilon < \delta \Rightarrow \forall \lambda \in \Lambda, \tau_\lambda(\epsilon) > \epsilon]$

then we have, for any $f_o \in \mathcal{H}_{T,\mathbf{w},p}^i$:

$$\lim_{\epsilon \rightarrow 0} \left[\sup_{\|g - Tf_o\|_{\mathcal{H}^o} \leq \epsilon} \|f_{\gamma(\epsilon),\mathbf{w},p,\tau(\epsilon)}^* - f_o^\dagger\|_{\mathcal{H}^i} \right] = 0,$$

where f_o^\dagger is the unique element of minimum $\|\cdot\|_{\mathbf{w},p}$ -norm in the set \mathcal{M}_{T,f_o} .

The detailed proof of this theorem is given in³, p.18-24 and is not reproduced here. It is based on two ingredients:

- The two lemmas provided in Appendix. A. show that condition 2 in Theorem 3.5 is needed to obtain the weak convergence of the adaptive projection operators $M_{g,\tau}$ when $\|g - Tf_o\| \rightarrow 0$.
- Using this weak convergence, one can then adapt the proof of Theorem 2.3 provided in².

4. ADAPTIVE DISCREPANCY TERMS (II): RELAXED PROJECTIONS

In the previous section, we showed that introducing adaptive projections in the discrepancy term allows to take into account features that are more important in the data but results in a loss of information that may be harmful to the estimation of the object sought. The reason is that the projections used cancel some information. To fix this instability problem still keeping the spirit of the previous method, one can imagine to only dampen the non-feature space defined by the adaptive projections instead of cancelling it. As we see in the next section, the resulting “relaxed projections” still emphasize the same features but without losing any information; therefore the stability as defined in Theorem 2.3 is restored.

4.1 Relaxed Adaptive Projections

The “relaxed projection” $M_{g,\tau,\mu}$ with dampening parameter μ and corresponding to the orthogonal adaptive projection $M_{g,\tau}$ is

$$M_{g,\tau,\mu} = M_{g,\tau} + \mu(\text{Id} - M_{g,\tau}) \quad (17)$$

or more formally:

DEFINITION 4.1. Given an orthonormal basis of \mathcal{H}^o , $\beta = \{\beta_\lambda\}_{\lambda \in \Lambda}$, an element g in \mathcal{H}^o , a sequence of nonnegative thresholds $\tau = \{\tau_\lambda\}_{\lambda \in \Lambda}$ and a scalar $\mu > 0$, $M_{g,\tau,\mu}$ is the map from \mathcal{H}^o into itself defined by:

$$\forall h \in \mathcal{H}^o, \quad M_{g,\tau,\mu}(h) = \sum_{\lambda \text{ s.t. } |g_\lambda| > \tau_\lambda} h_\lambda \beta_\lambda + \mu \sum_{\lambda \text{ s.t. } |g_\lambda| \leq \tau_\lambda} h_\lambda \beta_\lambda$$

This operator is introduced in the discrepancy term so that we now seek to minimize the functional

$$\mathbf{J}_{\gamma,\mathbf{w},p,\tau,\mu}(f) = \|M_{g,\tau,\mu}(Tf - g)\|_{\mathcal{H}^o}^2 + \gamma \|f\|_{\mathbf{w},p}, \quad (18)$$

via the following iterative algorithm:

ALGORITHM 4.2.

$$\begin{cases} f^0 & \text{arbitrary} \\ f^n & = \mathbf{S}_{\gamma,\mathbf{w},p}(f^{n-1} + T^* M_{g,\tau,\mu}^2(g - Tf^{n-1})), \quad n \geq 1 \end{cases}$$

Note that in this case, one needs to square the relaxed projection operator in the iterative algorithm. This is because unlike $M_{g,\tau}$, $M_{g,\tau,\mu}$ is not a self-adjoint projection. This equation can be easily checked by replacing T by $M_{g,\tau,\mu}T$ and g by $M_{g,\tau,\mu}g$ in the original functional $\mathbf{J}_{\gamma,\mathbf{w},p}$ of Eq.(7) and in Algorithm 2.1. In practice, we use the fact that $M_{g,\tau,\mu}^2 = M_{g,\tau,\mu^2}$; so the operator is still easy to compute.

The previous change of variable used in Theorem 2.2 also proves the **strong convergence** of Algorithm 4.2 to a minimizer of Eq. (18) (under the same conditions as in Theorem 2.2).

4.2 Stability is recovered

The introduction of the dampening factor ensures that all the information in the data will be taken into account and we recover the stability in the usual sense: if the data become ideal ($g \rightarrow Tf_o$) and the parameters γ , $\tau = \{\tau_\lambda\}_{\lambda \in \Lambda}$ and μ are chosen accordingly, then the solution converges to f_o when f_o is the unique antecedent of Tf_o .

The conditions on the parameters are given in the following theorem:

THEOREM 4.3. Assume that T is a bounded operator from \mathcal{H}^i to \mathcal{H}^o with $\|T\| < 1$ and that the entries in the sequence $\mathbf{w} = \{w_\lambda\}_{\lambda \in \Lambda}$ are bounded below uniformly by a strictly positive number c .

For any $g \in \mathcal{H}^o$ and any $\gamma > 0$, $0 < \mu \leq 1$ and nonnegative sequence $\tau = \{\tau_\lambda\}_{\lambda \in \Lambda}$, define $f_{\gamma,\mathbf{w},p,\tau,\mu;g}^*$ to be a minimizer of $\mathbf{J}_{\gamma,\mathbf{w},p,\tau,\mu}(f)$ with observation g . If $\gamma = \gamma(\epsilon)$, $\tau = \tau(\epsilon)$ and $\mu = \mu(\epsilon)$ satisfy:

1. $\lim_{\epsilon \rightarrow 0} \gamma(\epsilon) = 0$ and $\lim_{\epsilon \rightarrow 0} \frac{\epsilon^2}{\gamma(\epsilon)} = 0$
2. $\forall \lambda \in \Lambda, \lim_{\epsilon \rightarrow 0} \tau_\lambda(\epsilon) = 0$ and $\forall \lambda \in \Lambda, \exists \delta(\lambda) > 0, \text{ s.t. } [\epsilon < \delta(\lambda) \Rightarrow \tau_\lambda(\epsilon) > \epsilon]$
3. $\lim_{\epsilon \rightarrow 0} \mu(\epsilon) = \mu_o$, with $0 < \mu_o \leq 1$

then for any f_o such that there is a unique minimizer of the $\|\cdot\|_{\mathbf{w},p}$ -norm in the set $\mathcal{S}f_o = \{f : Tf = Tf_o\}$:

$$\lim_{\epsilon \rightarrow 0} \left[\sup_{\|g - Tf_o\|_{\mathcal{H}^o} \leq \epsilon} \|f_{\gamma(\epsilon),\mathbf{w},p,\tau(\epsilon),\mu(\epsilon);g}^* - f_o^\dagger\|_{\mathcal{H}^i} \right] = 0,$$

where f_o^\dagger is the unique element of minimum $\|\cdot\|_{\mathbf{w},p}$ -norm in the set $\mathcal{S}f_o$.

The proof of this theorem is detailed in³, p.28-31 and is similar to that of Theorem 3.5. The weak convergence of the adaptive operators is ensured by conditions 2 and 3 of Theorem 4.3 and the corresponding lemma is provided in Appendix B.

It is clear that in practice, by choosing μ small, the properties of g enhanced by both Algorithm 3.2 and 4.2 are similar. The second algorithm is however more stable as it is guaranteed to make a correct guess when the data is sufficiently close to the image of an object f .

5. APPLICATION

5.1 Multispectral Data

In this section we apply the algorithms described previously to the problem of reconstructing maps of astrophysical phenomena from multispectral observations. We consider simulated multispectral observations of the Cosmic Microwave Background (CMB) radiation with the observation conditions relative to the Atacama Cosmology Telescope (ACT). In this case, we observe the same portion of sky at different wavelengths ν_l . The observations are blurred mixtures of the physical phenomena we seek f_1, \dots, f_M that can be modeled by:

$$\forall l = 1..L, \quad g(\nu_l) = g_l = b_l * \sum_{m=1}^M a_{m,l} f_m + n_l. \quad (19)$$

The blurring b_l changes with the wavelength ν_l and is Gaussian. The mixture coefficients $a_{m,l}$ are called frequency dependencies and give the contribution of phenomena m to observation l . The noise terms n_l have a known variance σ_l that also depend on the wavelength ν_l . Note that here, the operator $T_{m,l}$ from Eq.(2) is a mixture followed by a convolution $T_{m,l}(\cdot) = b_l * \sum_{m=1}^M a_{m,l}(\cdot)_m$. For ACT, the observation wavelength are low: $\nu = 145, 217$ or 265 GHz. (Details about the noise and blur level can be found in³, p.88.)

Here, we seek to reconstruct two components:

- the CMB ($= f_1$): this is an electromagnetic radiation that fills the whole of the Universe (see Figure 2, left panel). Its existence and properties are considered one of the major confirmations of the Big Bang theory.
- the galaxy clusters, noted SZ ($= f_2$): the clusters can be seen through their Sunyaev-Zeldovich effect (SZ effect in short) which is due to high energy electrons in the galaxy clusters that interact with Cosmic Microwave Background photons.

In fact, we focus on the detection and estimation of the galaxy clusters in observations such as can be done with ACT.

A complete model of the observations would have to include other astrophysical phenomena such as infrared point sources or our Galaxy dust. We will not consider them here, since their contribution at low wavelengths, such as the ones considered here, are negligible.

Figure 2 illustrates the simulated data we use. The two left panels show the astrophysical map we seek to reconstruct from the observations shown on the two right panels. (The units of the maps is the micro-Kelvin).

General parameters of the functional algorithms

In this multispectral case, the reconstruction methods proposed earlier have one regularization term for each component and one regularization term per observation (see Eq.(3)).

As can be seen from the observations, the contribution of the galaxy clusters (SZ) is negligible compared to this of the CMB. We rely on the fact that these maps have very different spatial properties to disentangle them. These properties are reflected by the regularization terms. The CMB component is regularized by a weighted l_2 -norm in Fourier space, the weights being proportional to its spectral power. The SZ component is regularized by an l_1 -norm on its wavelets coefficients. The wavelet transform used for regularization is the dual tree complex wavelet transform.^{4,5}

We compare the results obtained with the classical discrepancy terms of Eq.(3) to these obtained with various adaptive projections, relaxed (Eq.(18)) and not (Eq.(12)). In any case, the adaptive/relaxed projection is done on an orthonormal wavelet transform (Symmlet, 2 vanishing moments) and the threshold parameter τ are set to the noise standard deviation.

The general balancing parameters ρ_l are set to 1. The γ_m are learned from a database of simulations.

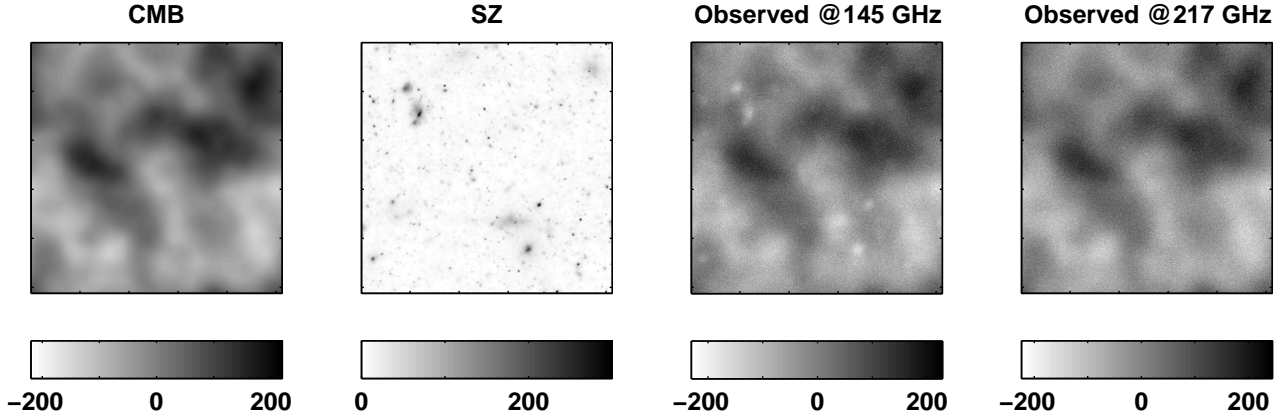


Figure 2. Multispectral data (units: μK); left to right: CMB map, galaxy clusters map, observation at 145GHz, observation at 217GHz

5.2 Denoising galaxy cluster maps

To illustrate the effect of introducing adaptive projections, we took a one-dimensional slice of the galaxy cluster map in Fig. 2 and added noise to it (top panels of Fig. 3) with a standard deviation of σ .

We show in the four bottom panels of this figure the results of the denoising using

- hard-thresholding (Fig. 3, middle left panel) with threshold $\tau = \sigma$;
- soft-thresholding (Fig. 3, bottom right panel) with threshold $\gamma = \sigma$; This is obtained with the initial iterative algorithm (Algo. 2.1).
- the adaptive thresholding (Fig. 3, middle right and bottom left panels) seen in subsection 3.2.2. This is obtained with the adaptive algorithm (Algo. 3.2) Note that there is no stability issue in this example.

One can see that increasing the introduction of the soft-thresholding on top of the hard-thresholding smooths the solution. The pure soft-thresholding however suffers from that fact that it dampens peaks compared to the pure hard-thresholding. In the case of galaxy clusters, these peak of intensity correspond to the central part of the cluster and indicate its age. Therefore, the dampening obtained by soft-thresholding is detrimental. On the other hand, the lack of smoothness of the hard-thresholded solution will induces false positive in the detection of clusters. The introduction of the adaptive thresholding via the use of projections in the discrepancy term allows to tune both effects. It gives an interesting compromise keeping a bit of the advantages of the pure hard or soft-thresholded solutions (see Fig. 3, middle right and bottom left panels).

5.3 Reconstruction of CMB and galaxy clusters maps from multispectral observations

The simultaneous reconstruction of both the CMB and galaxy cluster maps from the multispectral observations as seen in subsection 5.1 has been performed with the different iterative algorithms proposed in section 2, 3 and 4. All parameters were described in 5.1 except for the relaxed projection dampening parameter μ (see Eq.(17)) which is fixed here to $\mu = 0.1$ when using Algorithm 4.2.

Fig. 4 displays the results obtained for

- the initial algorithm (Algo. 2.1) with classical l_2 discrepancy terms. The results are labelled “ $\mu = 1$ ”.
- the relaxed projection algorithm (Algo. 4.2) with stable adaptive discrepancy terms. The results are labelled “ $\mu = 0.1$ ”.

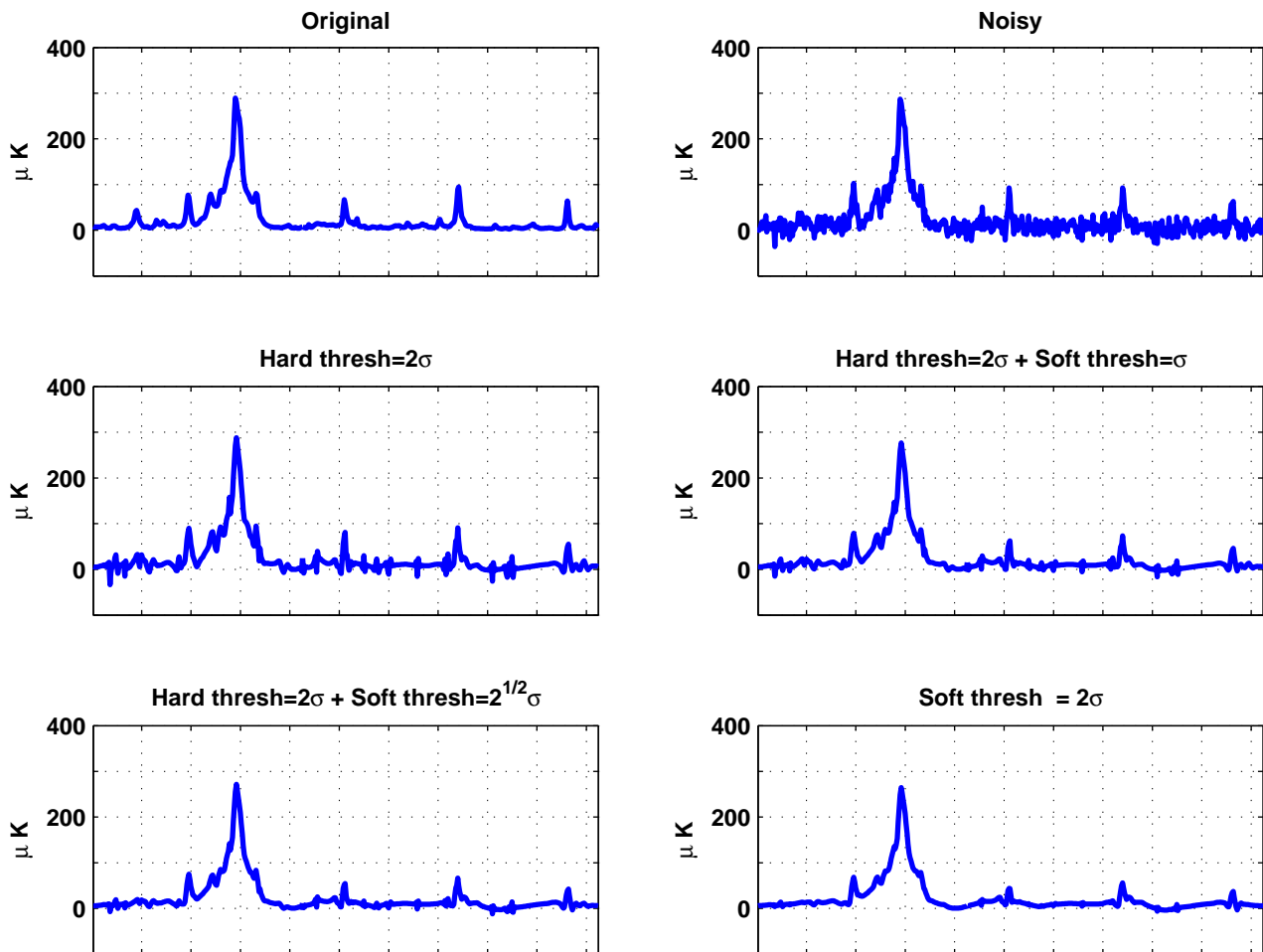


Figure 3. Denoising galaxy clusters map. Top left: 1D profile of a cluster map. Top right: noisy 1D profile of a cluster map (noise variance σ^2). Middle left: noisy data hard-thresholded ($\tau = 2\sigma$). Middle right: noisy data soft/hard-thresholded ($\tau = 2\sigma$, $\gamma = \sigma$). Bottom left: noisy data soft/hard-thresholded ($\tau = 2\sigma$, $\gamma = \sqrt{2}\sigma$). Bottom right: noisy data soft-thresholded ($\gamma = 2\sigma$).

The observed maps and the CMB and galaxy clusters maps that we seek to recover are shown on the left panels of Fig. 2. The reconstructed CMB maps are in the two left panels of Fig. 4. The reconstructed galaxy maps are in the two right panels of Fig. 4.

The following analysis is illustrated by the results shown in 4 but is valid in a more general study with 24 similar simulations.

5.3.1 Analysis of CMB reconstruction

All the reconstructed CMB maps are accurate to the microKelvin precision. The Root Mean Square Error of the different reconstructions to the original (true) CMB map is not affected by the introduction of the adaptive discrepancy term.

The precision obtained for this component is highly satisfactory and allows to proceed to further treatment for astrophysical purposes.

5.3.2 Analysis of the galaxy clusters reconstruction

All the reconstructed galaxy cluster maps have a low accuracy (worst case 100 microKelvin). The Root Mean Square Error of the different reconstructions to the original (true) clusters map is not affected by the introduction of the adaptive discrepancy term. Hence as far as global measures are concerned, all the presented algorithms perform in the same manner

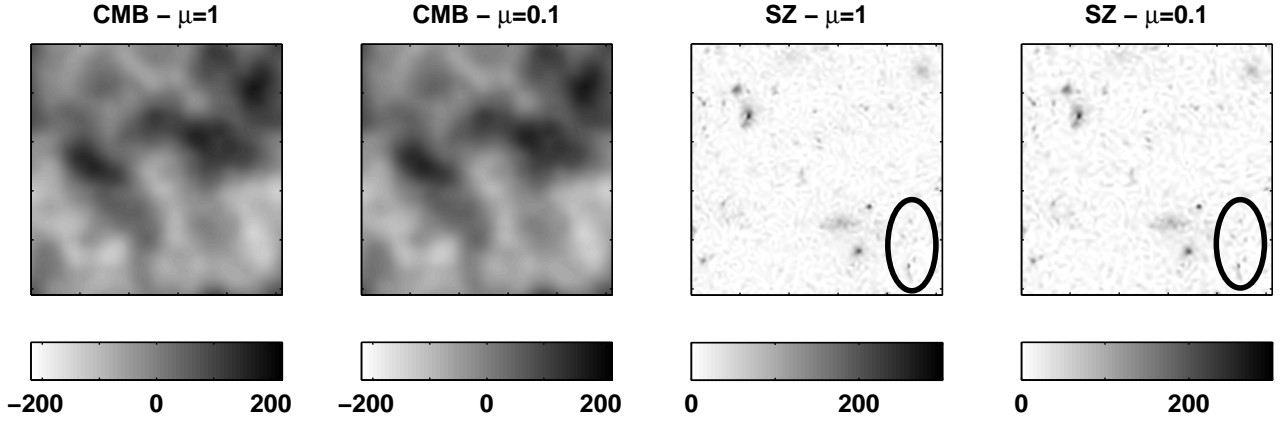


Figure 4. Reconstructed maps; without projections: first and third images ($\mu = 1$); with adaptive projections: second and fourth images ($\mu = 0.1$); far/middle left: CMB; far/middle right: galaxy clusters

for galaxy clusters. These poor results are expected by the fact that the contribution of the galaxy clusters to the observation is well below the CMB contribution and the noise level.

However, as explained in,³ global measures are not satisfactory to evaluate the quality of a reconstructed cluster map. Indeed, the goal is to locate the presence of clusters and quantify some of their statistical characteristics like size, intensity or age... Detailed study of the reliability of these quantities has been done for Algorithm 2.1³ and show that it actually gives good results in this perspective. Here, we do not reproduce the all study for Algorithm 3.2 and 4.2 but simply compare them to Algorithm 2.1.

As can be inferred from Fig. 4, the results are very similar. The introduction of adaptive discrepancy terms yield a slight improvement in the estimation of the central intensity of a cluster (see the three clusters in the upper part of the circle in Fig. 4). This improvement is not statistically significant however it illustrates how adaptive discrepancy terms provide a novel way of tuning the algorithm to the data.

APPENDIX A. ELEMENTS OF THE PROOF OF THE STABILITY OF (I)

To prove Theorem 3.5, we need to examine the behavior of the projections $M_{g(\epsilon), \tau(\epsilon)}$ as ϵ goes to zero. This is done in the next two lemmas. The first lemma (Lemma A.1) gives necessary and sufficient conditions on the sequence $\tau = \{\tau_\lambda\}_{\lambda \in \Lambda}$ to that these projections converge in a weak sense as ϵ goes to zero. We will be interested in the case where the weak limit operator is $M_{Tf_0, 0}$. The second lemma (Lemma A.1) refines these conditions, so that in addition, the sequence $M_{g(\epsilon), \tau(\epsilon)}$ converges strongly to $M_{Tf_0, 0}$ on the set: $T(\mathcal{M}_{f_0})$.

LEMMA A.1. For $f \in \mathcal{H}^i$, let $\{g(\epsilon, f)\}_{\epsilon > 0}$ be an arbitrary family of elements in \mathcal{H}^o that satisfy $\|g(\epsilon, f) - Tf\|_{\mathcal{H}^o} < \epsilon$, $\forall \epsilon > 0$.

1. $\forall h \in \mathcal{H}^o$, $M_{g(\epsilon, f), \tau(\epsilon)}h$ converges weakly as ϵ goes to 0 **if and only if** $\forall \lambda : \exists \delta(\lambda)$ such that either (a) or (b) holds, with
 - (a) $\forall \epsilon \in (0, \delta(\lambda))$, $|[g(\epsilon, f)]_\lambda| > \tau_\lambda$,
 - (b) $\forall \epsilon \in (0, \delta(\lambda))$, $|[g(\epsilon, f)]_\lambda| \leq \tau_\lambda$.
2. $M_{g(\epsilon, f), \tau(\epsilon)}$ converges weakly, independently of the choice of f and of the family $g(\epsilon, f)$, as ϵ goes to 0 **if and only if** $\forall \lambda : \text{both (a) and (b) hold, with}$
 - (a) $\exists \delta(\lambda)$ such that $\forall \epsilon \in (0, \delta(\lambda))$, $\tau_\lambda(\epsilon) > \epsilon$
 - (b) $\lim_{\epsilon \rightarrow 0} \tau_\lambda(\epsilon) = 0$

In that case, the weak-limit operator is necessarily $M_{Tf,0}$.

3. When conditions 2.(a) and 2.(b) above hold, if $h(\epsilon)$ converges weakly to h , then $M_{g(\epsilon,f),\tau(\epsilon)}h(\epsilon)$ converges weakly to $M_{Tf,0}h$ as ϵ goes to 0.

Proof. [Proof of Lemma A.1] Let us examine the behavior of $M_{g(\epsilon,f),\tau(\epsilon)}$ coordinate by coordinate. Since $[M_{g(\epsilon,f),\tau(\epsilon)}h]_\lambda$ equals either h_λ or 0, depending on whether or not $|[g(\epsilon,f)]_\lambda| > \tau_\lambda(\epsilon)$, it follows that $M_{g(\epsilon,f),\tau(\epsilon)}(h)$ will converge weakly as ϵ goes to 0 if and only if for all coordinates λ , one of the following holds:

Either there exists some $\delta(\lambda) > 0$ such that $|[g(\epsilon,f)]_\lambda| > \tau_\lambda(\epsilon)$ for $\epsilon < \delta(\lambda)$. In this case, $[M_{g(\epsilon,f),\tau(\epsilon)}h]_\lambda = h_\lambda$ for $\epsilon < \delta(\lambda)$.

Or there exists some $\delta(\lambda) > 0$ such that $|[g(\epsilon,f)]_\lambda| \leq \tau_\lambda(\epsilon)$ for $\epsilon < \delta(\lambda)$. In this case, $[M_{g(\epsilon,f),\tau(\epsilon)}h]_\lambda = 0$ for $\epsilon < \delta(\lambda)$.

This proves the first assertion.

Let us now consider how uniform this behavior is in the choice of the family $g(\epsilon, f)$. Since $|[g(\epsilon, f) - Tf]_\lambda| \leq \|g(\epsilon, f) - Tf\|_{\mathcal{H}^0} \leq \epsilon$, the set of values that can be assumed by $|g(\epsilon, f)_\lambda|$ is exactly $[|Tf|_\lambda - \epsilon, |Tf|_\lambda + \epsilon]$ (take $g = Tf + r\beta_\lambda$, $r \in [-\epsilon, \epsilon]$ to reach all the values in this set). Therefore, for a fixed f , the weak convergence of the operators $M_{g(\epsilon,f),\tau(\epsilon)}$, regardless of which sequence $g(\epsilon, f)$ is chosen, is equivalent to putting constraints on the sequence $\{\tau(\epsilon)_\lambda\}_{\lambda \in \Lambda}$ that depend of the coordinates $(Tf)_\lambda$. These constraints depends on whether $(Tf)_\lambda \neq 0$ or $(Tf)_\lambda = 0$:

- If $Tf_\lambda \neq 0$ then $\{|g(\epsilon, f)_\lambda|\} = [|Tf|_\lambda - \epsilon, |Tf|_\lambda + \epsilon]$. Therefore, one needs either: $[\epsilon < \delta(\lambda) \Rightarrow \tau_\lambda(\epsilon) > |Tf|_\lambda + \epsilon]$ or $[\epsilon < \delta(\lambda) \Rightarrow \tau_\lambda(\epsilon) \leq |Tf|_\lambda - \epsilon]$. In the first case, β_λ will always be in the kernel of $M_{g(\epsilon,f),\tau(\epsilon)}$ once $\epsilon < \delta(\lambda)$. In the second case β_λ will always in the range of $M_{g(\epsilon,f),\tau(\epsilon)}$ once $\epsilon < \delta(\lambda)$.
- If $Tf_\lambda = 0$ then $\{|g(\epsilon, f)_\lambda|\} = [0, \epsilon]$. Therefore one needs $[\epsilon < \delta(\lambda) \Rightarrow \tau_\lambda(\epsilon) > \epsilon]$. In this case, β_λ will always be in the kernel of $M_{g(\epsilon,f),\tau(\epsilon)}$ once $\epsilon < \delta(\lambda)$.

Note that we do not know beforehand the value of Tf . To be useful, we must derive requirements on the parameters $\tau_\lambda(\epsilon)$ that do not depend on f . The minimum requirements on $\tau(\epsilon)$ ensuring the operators $M_{g(\epsilon,f),\tau(\epsilon)}$ converge weakly as ϵ goes to 0 are:

- $\forall \lambda, \lim_{\epsilon \rightarrow 0} \tau_\lambda(\epsilon) = 0$: this ensures that if $Tf_\lambda \neq 0$, we will have $\tau_\lambda(\epsilon) < |Tf|_\lambda - \epsilon$ for sufficiently small ϵ .
- $\forall \lambda, \exists \delta(\lambda)$ such that $\epsilon < \delta(\lambda) \Rightarrow \tau_\lambda(\epsilon) < \epsilon$: this ensures that if $Tf_\lambda = 0$, we will have $\tau_\lambda(\epsilon) < |Tf|_\lambda + \epsilon = \epsilon$ for sufficiently small ϵ .

If these conditions are satisfied, the $M_{g(\epsilon,f),\tau(\epsilon)}$ converge weakly as ϵ goes to 0 and one can determine the weak limit:

- for λ s.t. $Tf_\lambda \neq 0$: $\lim_{\epsilon \rightarrow 0} \tau_\lambda(\epsilon) = 0$ hence there exists $\delta(\lambda, f)$ such that $\epsilon < \delta(\lambda, f)$ implies $\tau_\lambda(\epsilon) < |Tf|_\lambda - \epsilon$. It follows that: $|g(\epsilon, f)_\lambda| > \tau_\lambda(\epsilon)$ so that $M_{g(\epsilon,f),\tau(\epsilon)}(\beta_\lambda) = \beta_\lambda$ for any $g(\epsilon, f)$ and any $\epsilon < \delta(\lambda, f)$
- for λ s.t. $Tf_\lambda = 0$: $\epsilon < \delta(\lambda)$ implies $\tau_\lambda(\epsilon) > \epsilon$. It follows that if $\epsilon < \delta(\lambda)$, then $|g(\epsilon, f)_\lambda| > \tau_\lambda(\epsilon)$ so that $M_{g(\epsilon,f),\tau(\epsilon)}(\beta_\lambda) = 0$ for any $g(\epsilon, f)$ and any $\epsilon < \delta(\lambda)$.

This proves that the weak limit of $M_{g(\epsilon,f),\tau(\epsilon)}$ for any fixed f is $M_{Tf,0}$ and finishes the proof of the second part of Lemma A.1.

Finally, assuming $h(\epsilon)$ converges weakly to h , we have $\forall \lambda$:

$$\left| [M_{g(\epsilon,f),\tau(\epsilon)}h(\epsilon) - M_{Tf,0}h]_\lambda \right| \quad (20)$$

$$= \left| [M_{g(\epsilon,f),\tau(\epsilon)}(h(\epsilon) - h) + (M_{g(\epsilon,f),\tau(\epsilon)} - M_{Tf,0})h]_\lambda \right| \quad (21)$$

$$= \left| [M_{g(\epsilon,f),\tau(\epsilon)}(h(\epsilon) - h)]_\lambda \right| + \left| [M_{g(\epsilon,f),\tau(\epsilon)}h - M_{Tf,0}h]_\lambda \right| \quad (22)$$

The second term vanishes as ϵ goes to 0 because $M_{g(\epsilon,f),\tau(\epsilon)}$ converges weakly to $M_{Tf,0}$ when the conditions 2.(a) and 2.(b) hold. Moreover, we have seen in the proof of the second part of the lemma that for any λ :

- either there exists a $\delta(\lambda)$ such that $M_{g(\epsilon, f), \tau(\epsilon)}(\beta_\lambda) = 0$ for any $\epsilon < \delta(\lambda)$.

In that case, $\left| \left[M_{g(\epsilon, f), \tau(\epsilon)}(h(\epsilon) - h) \right]_\lambda \right| = 0$, for $\epsilon < \delta(\lambda)$.

- or there exists a $\delta(\lambda)$ such that $M_{g(\epsilon, f), \tau(\epsilon)}(\beta_\lambda) = \beta_\lambda$ for any $\epsilon < \delta(\lambda)$.

In that case, $\left| \left[M_{g(\epsilon, f), \tau(\epsilon)}(h(\epsilon) - h) \right]_\lambda \right| = \left| [h(\epsilon) - h]_\lambda \right|$, for $\epsilon < \delta(\lambda)$; and the weak convergence of $h(\epsilon)$ to h allows to conclude that $\left| \left[M_{g(\epsilon, f), \tau(\epsilon)}(h(\epsilon) - h) \right]_\lambda \right| \rightarrow 0$

This proves that $M_{g(\epsilon, f), \tau(\epsilon)}h(\epsilon)$ converges weakly to $M_{Tf, 0}h$ and finishes the proof of Lemma A.1. \square

We shall now see how to ensure strong convergence of the $M_{g(\epsilon, f), \tau(\epsilon)}(h)$ when h is in \mathcal{M}_f .

LEMMA A.2. *If there exists a value of δ independent of λ such that $\forall \epsilon < \delta$ and $\forall \lambda, \tau_\lambda(\epsilon) > \epsilon$, then the two following properties hold:*

1. *For any choice of f and of the family $g(\epsilon, f)$:*

$$\forall \epsilon < \delta, M_{g(\epsilon, f), \tau(\epsilon)} = M_{Tf, 0} M_{g(\epsilon, f), \tau(\epsilon)} = M_{g(\epsilon, f), \tau(\epsilon)} M_{Tf, 0} = \sum_{\substack{\lambda \text{ s.t. } Tf_\lambda \neq 0 \\ \text{and } |g_\lambda| \geq \tau_\lambda}} \langle \cdot, \beta_\lambda \rangle \beta_\lambda.$$

2. *In particular, for any choice of $f \in \mathcal{H}_{T, \mathbf{w}, p}^i$ and of the family $g(\epsilon, f)$, (i.e. whenever \mathcal{M}_f has a unique minimizer f^\dagger of the $\|\cdot\|_{\mathbf{w}, p}$ -norm):*

$$\forall \epsilon < \delta, M_{g(\epsilon, f), \tau(\epsilon)}(Tf^\dagger) = M_{g(\epsilon, f), \tau(\epsilon)}(Tf).$$

Proof. [Proof of Lemma A.2:] The first part of Lemma A.2 results from properties of orthogonal projections. If P_1 and P_2 are two orthogonal projections, then:

$$\begin{aligned} P_1 P_2 &= P_2 P_1 \\ \ker(P_2) \subset \ker(P_1) &\Leftrightarrow P_1 P_2 = P_1. \end{aligned}$$

Hence, we already proved $M_{g(\epsilon, f), \tau(\epsilon)} M_{Tf, 0} = M_{Tf, 0} M_{g(\epsilon, f), \tau(\epsilon)}$ and

$$M_{g(\epsilon, f), \tau(\epsilon)} M_{Tf, 0} = M_{g(\epsilon, f), \tau(\epsilon)} \Leftrightarrow [(Tf)_\lambda = 0 \Rightarrow |g(\epsilon, f)_\lambda| \leq \tau_\lambda(\epsilon)].$$

When f and ϵ are fixed, the right hand side holds for any $g(\epsilon, f)$ if and only if $[(Tf)_\lambda = 0 \Rightarrow \epsilon < \tau_\lambda(\epsilon)]$ which proves the first part of Lemma A.2.

For f in $\mathcal{H}_{T, \mathbf{w}, p}^i$, f^\dagger is well defined and verifies $M_{Tf, 0} Tf^\dagger = Tf$. Applying $M_{g(\epsilon, \tau(\epsilon))}$ to this equality and using the previous result finishes the proof of Lemma A.2. \square

APPENDIX B. ELEMENTS OF THE PROOF OF THE STABILITY OF (II)

LEMMA B.1. *Suppose that $\tau = \tau(\epsilon)$ and $\mu = \mu(\epsilon)$ verify conditions 2 and 3 of Theorem 4.3. Then the two following properties hold:*

1. *For any h in \mathcal{H}^o , $M_{g(\epsilon, f), \tau(\epsilon), \mu(\epsilon)}^2 h$ converges weakly to $M_{Tf, 0, \mu^o}^2 h$ as ϵ goes to 0.*
2. *If $h(\epsilon)$ converges weakly to h as ϵ goes to 0, then $M_{g(\epsilon, f), \tau(\epsilon), \mu(\epsilon)}^2 h(\epsilon)$ converges weakly to $M_{Tf, 0, \mu^o}^2 h$ as ϵ goes to 0.*

Proof. [Proof of Lemma B.1:] In the proof of Lemma A.1, we have seen that under conditions imposed on $\tau(\epsilon)$ (conditions 3 and 4 of Theorem 4.3), the following happens:

- for λ s.t. $Tf_\lambda \neq 0$: $\lim_{\epsilon \rightarrow 0} \tau_\lambda(\epsilon) = 0$ hence there exists $\delta(\lambda, f)$ such that $\epsilon < \delta(\lambda, f)$ implies $\tau_\lambda(\epsilon) < |Tf_\lambda| - \epsilon$. It follows that: $|g(\epsilon, f)_\lambda| > \tau_\lambda(\epsilon)$.

- for λ s.t. $Tf_\lambda = 0$: $\epsilon < \delta(\lambda)$ implies $\tau_\lambda(\epsilon) > \epsilon$. It follows that if $\epsilon < \delta(\lambda)$, then $|g(\epsilon, f)_\lambda| > \tau_\lambda(\epsilon)$.

So that in the first case: $M_{g(\epsilon, f), \tau(\epsilon), \mu(\epsilon)}^2(\beta_\lambda) = \beta_\lambda$ for any $g(\epsilon, f)$ and any $\epsilon < \delta(\lambda, f)$; and in the second case: $M_{g(\epsilon, f), \tau(\epsilon), \mu(\epsilon)}^2(\beta_\lambda) = \mu(\epsilon)^2 \beta_\lambda$ for any $g(\epsilon, f)$ and any $\epsilon < \delta(\lambda)$. Since $\mu(\epsilon)$ converges to some μ_o by assumption (condition 5 of Theorem 4.3), it follows that $M_{g(\epsilon, f), \tau(\epsilon), \mu(\epsilon)}^2 h$ converges to $M_{Tf_o, 0, \mu_o}^2 h$ as (ϵ) goes to 0. This proves the first part of Lemma B.1.

To prove the second part of Lemma B.1, we use again the splitting trick we used in A.1.(3):

$$\left| [M_{g(\epsilon, f), \tau(\epsilon), \mu(\epsilon)}^2 h(\epsilon) - M_{Tf_o, 0, \mu_o}^2 h]_\lambda \right| \quad (23)$$

$$= \left| [M_{g(\epsilon, f), \tau(\epsilon), \mu(\epsilon)}^2 (h(\epsilon) - h) + (M_{g(\epsilon, f), \tau(\epsilon), \mu(\epsilon)}^2 - M_{Tf_o, 0, \mu_o}^2) h]_\lambda \right| \quad (24)$$

$$= \left| [M_{g(\epsilon, f), \tau(\epsilon), \mu(\epsilon)}^2 (h(\epsilon) - h)]_\lambda \right| + \left| [(M_{g(\epsilon, f), \tau(\epsilon), \mu(\epsilon)}^2 - M_{Tf_o, 0, \mu_o}^2) h]_\lambda \right| \quad (25)$$

And the same argument as we used in Lemma A.1.(3) allows to conclude. \square

ACKNOWLEDGMENTS

Most of this work was done in Princeton University, during the Ph.D of S.A supported by the BWF training program. The author would like to thank her collaborator Elena Pierpaoli for providing the astrophysical problem and data at the basis of this work, and her Ph.D. advisor Ingrid Daubechies.

REFERENCES

1. J.-L. Starck, F. Murtagh, and A. Bijaoui, "Multiresolution support applied to image filtering and restoration," *Graphical Models and Image Processing* **57**, pp. 20–431, 1995.
2. I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.* **57**, pp. 1413–1541, 2004.
3. S. Anthoine, *Different Wavelet-based Approaches for the Separation of Noisy and Blurred Mixtures of Components. Application to Astrophysical Data*. Ph.d. dissertation, Princeton University, 2005.
4. N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Applied and Computational Harmonic Analysis* **10**(3), pp. 234–253, 2001.
5. I. W. Selesnick, "The design of approximate hilbert transform pairs of wavelet bases," *IEEE Transactions on Signal Processing* **50**(5), pp. 1144–1152, 2002.