DOI : 10.1088/1742-6596/124/1/012008

On the convergence of iterative shrinkage algorithms with adaptive discrepancy terms

Sandrine Anthoine

I3S lab., Université de Nice Sophia-Antipolis - CNRS; 2000 route des lucioles, 06903 Sophia-Antipolis cedex, France

E-mail: anthoine@i3s.unice.fr

Abstract. In this paper, the inversion of a linear operator is tackled by a procedure called iterative shrinkage. Iterative shrinkage is a procedure that minimizes a functional balancing quadratic discrepancy terms with l_p regularization terms. In this work, we propose to replace the classical quadratic discrepancy terms with adaptive ones. These adaptive terms rely on adapted projections on a suitable basis. Two versions of these adaptive terms are proposed (one with a straightforward use of the projections and the other with relaxed projections) together with iterative algorithms minimizing the obtained functional. We prove the convergence and stability of corresponding algorithms. Moreover we prove that for a straightforward use of these adaptive projections, although the process is consistent, valuable information may be lost, which is not the case with the "relaxed" projections. We illustrate both algorithms on multispectral astronomical data.

1. Introduction

In this paper, we consider the general inverse problem of finding an object f from its (possibly noisy) observation g. We assume that the observation process can be modeled by a known bounded linear operator T. Hence one can model the observation g as $g = T(f) + n$ where n is a noise term. Even knowing T, this problem is ill-posed and therefore needs to be "regularized". The compromise that is sought is to have an estimate $\hat{f}$ of f that both

(i) fits the data: $T(\hat{f}) \sim g$

(ii) has "desirable" properties (for example, sharpness if f is a picture).

The regularization of inverse problems may be tackled with different models (stochastic or not) on the data f or noise n and different criteria of optimality to define what the "best" guess is. In this work we focus on functional analysis techniques to solve this problem. The optimality of the guess is defined by a cost functional J which measures both the fitness to the data and the properties of f itself. The estimate $\hat{f}$ we seek is a minimizer of this cost functional J, which is typically the sum of

(i) a discrepancy term that measures the fitness to the data: $J_{disc}(T(f), g)$

(ii) a regularization term that measures the "desirable" properties of f: $J_{reg}(f)$,

with a regularization parameter γ to balance the two terms. We have then:

$$J(f) = J_{disc}(T(f), g) + \gamma J_{reg}(f) \tag{1}$$

$T : \mathcal{H}^i \to \mathcal{H}^o$ is a bounded linear operator between two Hilbert spaces. Classically, $\mathcal{H}^i = \mathcal{H}^o$ and is an L^2 space and the discrepancy term is simply the quadratic norm of the residual

$$J_{disc}(T(f), g) = \left\|Tf - g\right\|^2_{\mathcal{H}^o}.$$

The regularization term may include a priori knowledge on the properties that are desirable on f. This formulation with cost functional is specially practical when the properties on f are well modeled by its belonging to a particular functional space, since then one can simply choose J_{reg} as the norm in this space. For images for example, the norm in the space of bounded variation functions has been successfully used ([1]....). In this work, we use weighted l_p-norms on a basis of $\mathcal{H}^i$ as the regularization term. The regularization terms we consider have then the following form:

$$J_{reg}(f) = \sum_{\lambda\in\Lambda} w_\lambda |\langle f, \varphi_\lambda\rangle|^p \overset{def}{=} \|\!|f|\!\|^p_{\mathbf{w},p} \tag{2}$$

where

- $\boldsymbol{\varphi}=\{\varphi_\lambda\}_{\lambda\in\Lambda}$is a basis of $\mathcal{H}^i$.
- p is an exponent, $1 \leq p \leq 2$.
- $\mathbf{w}=\{\mathbf{w}_\lambda\}_{\lambda\in\Lambda}$are strictly positive weights.

This allows to cover a wide variety of norm and consequently properties for f. For example, choosing $p = 2$ and $\boldsymbol{\varphi}= \{\varphi_\lambda\}_{\lambda\in\Lambda}$as the Fourier basis gives a simple quadratic norm with constraints on the power spectrum of f. When $p = 1$, the norm promotes sparsity in the basis chosen. We will give concrete examples of the regularization norm we choose for a particular application. Many other regularization term may be proposed, but here we focus on the discrepancy term.

The discrepancy term has been less closely studied in the literature and most methods rely on the classical quadratic term described earlier. However, it is clear that some a priori knowledge on the observation could also be useful to regularize the inversion. For example, often methods are known that alleviate the noise in the observation, thus emphasizing more important features in the observation. The classical quadratic discrepancy term does not allow to focus on these features.

This paper presents two ways to modify the classical quadratic discrepancy term in order to incorporate this idea of important feature in the observation. The key is to use adaptive projection operators to modify the discrepancy term. The original idea is from J.-L. Starck and co-authors. In [2], they present an iterative algorithm that focuses on 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 in [2]. Following this idea, we propose the use of general projections to 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). Let us denote by M_g a projection operator that depends on g, we consider in this paper the discrepancy terms of the form:

$$J_{disc}(T(f), g) = \left\|\mathrm{M}_g(Tf - g)\right\|^2_{\mathcal{H}^o}. \tag{3}$$

The cost functional we study therefore reads:

$$J(f) = \left\|\mathrm{M}_g(Tf - g)\right\|^2_{\mathcal{H}^o} + \gamma\|\!|f|\!\|^p_{\mathbf{w},p} = \left\|\mathrm{M}_g(Tf - g)\right\|^2_{\mathcal{H}^o} + \gamma\sum_{\lambda\in\Lambda} w_\lambda |\langle f, \varphi_\lambda\rangle|^p \tag{4}$$

Minimizing functionals with classical or adaptive discrepancy terms and l^p regularization terms (Eq.(4) without or with M_g) is obtained via "iterative shrinkage algorithms". These algorithms have received a lot attention in the last few year. Founding papers derive such algorithms in different framework (e.g. Bayesian in [3] or functional in [4]). Many aspects were and are still studied, in particular, different regularization terms (e.g. [5]), the extension to nonlinear problems [6], the separation of different constituents of the data in layers [7], the application to multidimensional data [8; 9], as well as the convergence speed [10; 11]. To the best of our knowledge, there is however no work on generalizing the discrepancy term and the consequences on the convergence and stability of the corresponding modified shrinkage algorithms, which is the main topic of this paper.

Using the mathematical framework introduced in [4], we study the behavior of the iterative algorithm minimizing functionals obtained with adaptive discrepancy terms such as those in Eq.(4). We prove the strong convergence and stability of the proposed iterative algorithms minimizing such functionals. We also show that the straight projections may lead to undesired effects (some useful information being erased by the projections). Therefore, we define "relaxed" projections operators which will not suffer this drawback while retaining the possibly of adaptation to important feature in the observation.

After the introduction, the paper is organized as follows. Section 2 reviews the principals results in [4], which present a complete study of the convergence, stability and convergence rate of iterative algorithms to minimize the functional of Eq. (4) when $\mathrm{M} = \mathrm{Id}$ (i.e without adaptive discrepancy term). In section 3, we present a first version of functional with discrepancy term. These consist in straight adaptive projections. We prove the convergence and stability of the obtained iterative algorithm and exhibit the drawbacks of having plain projection. Section 4 gives the relaxed version of the discrepancy term, together with its analysis. We briefly explain in section 5 the extension of this framework to inverse problem with multiple observation and/or objects. This extension is used in section 7, which is devoted to an application in astronomy.

2. Iterative shrinkage algorithm of [4]:

In this section, we summarize the findings in [4], where an iterative shrinkage algorithm was proposed to minimize Eq.(5) without adaptive discrepancy term, i.e the functional:

$$\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}(f) = \left\|Tf - g\right\|^2_{\mathcal{H}^o} + \gamma \|\!| f |\!\|^p_{\mathbf{w},p} = \left\|Tf - g\right\|^2_{\mathcal{H}^o} + \gamma \sum_{\lambda\in\Lambda} w_\lambda |\left\langle f, \varphi_\lambda \right\rangle|^p. \tag{5}$$

Here, the set $\boldsymbol{\varphi}=\{\varphi_\lambda\}_{\lambda\in\Lambda}$is a fixed basis of $\mathcal{H}^i$.

The main results in [4] are the proofs of convergence, stability and convergence rate of the proposed algorithm. We will build on the first two results to analyze our functionals with adaptive discrepancy terms and therefore remind the results in this section.

2.1. Iterative algorithm

The authors of [4] propose the following iterative algorithm to obtain a minimizer of Eq.(5):

Algorithm 2.1

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

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

$$\mathbf{S}_{\mathbf{w},\mathrm{p}}(h) = \sum_\lambda S_{w_\lambda,p}(h_\lambda)\varphi_\lambda \quad , \tag{6}$$

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

$$S_{w,p}(x) \stackrel{def}{=} \left(x + \frac{wp}{2}\ \operatorname{sign}(x)\ |x|^{p-1}\right)^{-1}\ ,\ \text{for } 1 \le p \le 2, \tag{7}$$

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

- $p = 1$: $\quad S_{w,1}(x) = \begin{cases} x - w/2 & \text{if} \quad x \quad \ge \quad w/2 \\ 0 & \text{if} \quad |x| \quad < \quad w/2 \\ x + w/2 & \text{if} \quad x \quad \le \quad -w/2 \end{cases}$ (soft-thresholding operator).
- $p = 2$: $\quad S_{w,2}(x) = \dfrac{x}{1+w}$.

2.2. Convergence and stability

The two following theorems summarize the findings presented in [4]. The first theorem states that 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 $\mathbf{S}_{\mathbf{w},\mathrm{p}}$ be the shrinkage operator defined by Eq.(6), where the sequence $\mathbf{w} = \{w_\lambda\}_{\lambda\in\Lambda}$ is such that there exists a constant $c > 0$ such that $\forall\lambda \in \Lambda : w_\lambda \ge c$. Then the sequence of iterates*

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

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

If the minimizer $f^\star$ of $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}$ is unique, (which is guaranteed e.g. by $p > 1$ or $\ker(T) = \{0\}$), then every sequence of iterates f^n converges strongly to $f^\star$, 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, $g = Tf_0 + e$, 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 $\gamma > 0$, $1 \le p \le 2$ 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 $f^\star_{\gamma,\mathbf{w},p;g}$ to be the minimizer of $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}$ with observation g. If $\gamma = \gamma(\epsilon)$ satisfies

$$\lim_{\epsilon\to 0} \gamma(\epsilon) = 0 \qquad \text{and} \qquad \lim_{\epsilon\to 0} \frac{\epsilon^2}{\gamma(\epsilon)} = 0\ , \tag{8}$$

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

$$\lim_{\epsilon\to 0}\left[\sup_{\|g-Tf_o\|_{\mathcal{H}^o}\le\epsilon} \|f^\star_{\gamma(\epsilon),\mathbf{w},p;g} - f_o^\dagger\|_{\mathcal{H}^i}\right] = 0\ , \tag{9}$$

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

Let us now make a two remarks concerning theses results.

Remark 1: Note that when T is invertible, $\mathcal{S}_{f_o}$ is reduced to $\{f_o\}$ and therefore $f_o^\dagger = f_o$. This means that Algorithm 2.1 provides a stable inversion of the set of invertible operators: one is ensured that when the observation becomes ideal, so will be the estimation.

Remark 2: The proposed iterative algorithm converges strongly to a minimizer of the functional $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}$. 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 ($\|Tf - g\|^2_{\mathcal{H}^o}$) and having the smallest $|\!|\!| \cdot |\!|\!|_{\mathbf{w},p}$–norm.

By using different bases, exponents p or weights $\mathbf{w}$, one can construct a $|\!|\!| \cdot |\!|\!|_{\mathbf{w},p}$–norm such that it will preserve or enhance desirable properties of f. For example, for $w_\lambda = p = 1$, the regularizing term is nothing more than the sum of absolute value of the coefficients on the basis $\boldsymbol{\varphi}= \{\varphi_\lambda\}_{\lambda\in\Lambda}$. This is frequently used to promote the sparsity of the decomposition of the estimate on the basis $\boldsymbol{\varphi}= \{\varphi_\lambda\}_{\lambda\in\Lambda}$, corresponding to the a priori knowledge that plausible estimates are compactly represented in this basis (few coefficients of large amplitude allow for a good description of the signal).

On the other hand, the quadratic discrepancy term ($\|Tf-g\|^2_{\mathcal{H}^o}$) in $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}$ is fixed and can not be adapted to the problem at hand. For example, it can not enhance more important features that should be matched in the observations, while discarding less important ones. In the rest of this paper, we will present adaptive discrepancy terms that aim at fixing this point.

3. Adaptive discrepancy terms via projections on the observations' feature space

3.1. Founding idea: the multiresolution support

In [2], the authors are concerned with the deconvolution of an astrophysical image. The observations of interest are blurred and noisy pictures of galaxies. For these, denoising by wavelet hard-thresholding was already known to improve the quality of noisy observations. The wavelet hard-thresholding procedure on g is the reconstruction from the wavelet transform of g where all the coefficients smaller than a predefined threshold have been put to zero. Therefore, wavelet hard-thresholding 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 name multiresolution support was defined in [2]). The fact that wavelet thresholding improves the observation shows that the "multiresolution support" of g naturally defines a subspace that describes the important features of g.

The authors of [2] proposed to use this multiresolution support not only to denoise 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 $(g - Tf^{n-1})$ is projected on the multiresolution support of g: $(g - Tf^{n-1})$ becomes $\mathrm{M}_g(g - Tf^{n-1})$ where M_g is the projection on the multiresolution support of g.

We propose to use this idea to extend the class of functionals and algorithms studied in [4] and study the mathematical properties of the resulting algorithms. Indeed, the idea in [2] is in essence to focus on the important features of the observation g knowing that it is not an perfect observation. If the ideal observation is known to be sparse in some basis (e.g. astronomical data have a sparse wavelet representation) then one can exploit this sparsity to focus on relevant pieces of information in the observation g by projecting on the subspace defined by large coefficients. This leads us to using adaptive projections in the discrepancy term of the functionals.

3.2. 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 non-negative thresholds* $\boldsymbol{\tau}=\{\tau_\lambda\}_{\lambda\in\Lambda}$*, the adaptive projection* $\mathrm{M}_{\mathrm{g},\tau}$ *is the map from* $\mathcal{H}^o$ *into itself defined by:*

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

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

We propose the following algorithm:

Algorithm 3.2

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

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 equivalent to what was proposed in [2]. From what we saw before, we can infer that Algorithm 3.2 should converge to a minimizer of

$$\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau}(f) = \big\|\mathrm{M}_{\mathrm{g},\tau}(Tf - g)\big\|^2_{\mathcal{H}^o} + \gamma |\!|\!| f |\!|\!|^p_{\mathbf{w},p} \tag{10}$$

which is a functional with an adaptive discrepancy term.

Before we go on proving the convergence and stability of Algorithm 3.2, we first gain insight on it by looking at the particular case when T is a diagonal operator on the basis $\{\beta_\lambda\}_{\lambda\in\Lambda}$.

3.3. Diagonal case: a mixture of hard- and soft-thresholding

Let us assume that T is diagonal:

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

where the t_λ are scalars.

$$\begin{aligned} \mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau}(f) &= \big\|\mathrm{M}_{\mathrm{g},\tau}(Tf - g)\big\|^2_{\mathcal{H}^o} + \gamma |\!|\!| f |\!|\!|_{\mathbf{w},p} \\ &= \sum_{\lambda\ s.t.\ |g_\lambda|>\tau_\lambda} |(Tf-g)_\lambda|^2 + \gamma \sum_{\lambda\in\Lambda} w_\lambda |f_\lambda|^p \\ &= \sum_{\lambda\ s.t.\ |g_\lambda|>\tau_\lambda} \Big(|t_\lambda f_\lambda - g_\lambda|^2 + \gamma w_\lambda |f_\lambda|^p\Big) + \gamma \sum_{\lambda\ s.t.\ |g_\lambda|\leq\tau_\lambda} w_\lambda |f_\lambda|^p \\ &= \sum_{\substack{\lambda\ s.t.\\ |g_\lambda|>\tau_\lambda \\ \&\ t_\lambda\neq 0}} t_\lambda^2\Big(|f_\lambda - g_\lambda/t_\lambda|^2 + \gamma w_\lambda/t_\lambda^2 |f_\lambda|^p\Big) + \sum_{\substack{\lambda\ s.t.\\ |g_\lambda|>\tau_\lambda \\ \&\ t_\lambda=0}} \Big(|g_\lambda|^2 + \gamma w_\lambda |f_\lambda|^p\Big) + \sum_{\substack{\lambda\ s.t.\\ |g_\lambda|\leq\tau_\lambda}} \gamma w_\lambda |f_\lambda|^p \end{aligned} \tag{11}$$

The equations for each f_λ are now decoupled so that the minimizer $f^\star$ of $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau}$ is defined by:

$$\begin{cases} f^\star_\lambda = S_{\gamma w_\lambda/t_\lambda^2,p}(g_\lambda/t_\lambda) & \text{if } |g_\lambda| > \tau_\lambda \quad \text{and} \quad t_\lambda \neq 0 \\ f^\star_\lambda = 0 & \text{if } |g_\lambda| \leq \tau_\lambda \quad \text{or} \quad t_\lambda = 0 \end{cases} \tag{12}$$

Introducing the hard-thresholding operator with threshold τ:

$$H_\tau(x) = \begin{cases} x & \text{if } |x| > \tau \\ 0 & \text{otherwise,} \end{cases} \tag{13}$$

and rewriting the preceding equation, the minimizer $f^\star$ of $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau}$ for T diagonal is defined by:

$$\begin{cases} f^\star_\lambda = S_{\gamma w_\lambda/t_\lambda^2,p}(H_{\tau_\lambda/\mathbf{t}_\lambda}(g_\lambda)) & \text{if } t_\lambda \neq 0 \\ f^\star_\lambda = 0 & \text{if } t_\lambda = 0. \end{cases} \tag{14}$$

Note that without the adaptive projection (minimizing $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}$ via Algorithm 2.1), the solution is the same without hard-thresholding: the minimizer $f^\star$ of $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}$ for T diagonal is defined by:

$$\begin{cases} f^\star_\lambda = S_{\gamma w_\lambda,p}(g_\lambda) \\ f^\star_\lambda = 0 \end{cases} \tag{15}$$

Thus we obtain the previous shrinkage operator $S_{\gamma w_\lambda/t_\lambda^2,p}$ composed with a hard-thresholding operator $H_{\tau/\mathbf{t}}$. We call the result an "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 $\||\cdot\||_{\mathbf{w},p}$–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.(14) (middle) in the case $\tau > \gamma$ (here $\mathbf{w} = 1$).

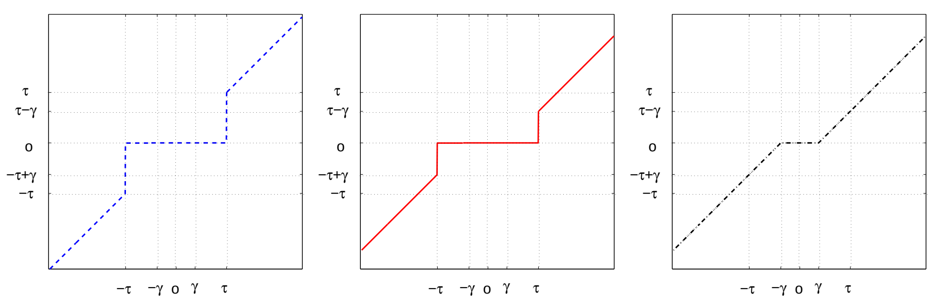

Figure 1. Left: hard-thresholding operator H_τ; middle: adaptive thresholding operator; right: soft-thresholding operator $S_{\gamma,1}$.

3.4. A convergent iterative algorithm

The strong convergence of Algorithm 3.2 to a minimizer of Eq.(10) is guaranteed by Theorem 2.2. Provided that the operator T, the exponent p and the sequence $\mathbf{w}=\{\mathbf{w}_\lambda\}_{\lambda\in\Lambda}$verify the conditions in Theorem 2.2), one can apply this theorem to $g' = \mathrm{M}_{\mathrm{g},\tau}\, g$ and $T' = \mathrm{M}_{\mathrm{g},\tau}\, T$ to get the solution (this works because $\mathrm{M}_{\mathrm{g},\tau}$ is a self-adjoint projection, so that T' verifies the conditions as soon as T does).

3.5. Stability is an issue

3.5.1. Examples

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 stable in the sense of Theorem 2.3, however the limit obtained has undesired properties. There is stability in the sense that if the parameters (τ, $\mathbf{w}$,..) 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 $\boldsymbol{\tau}=\{\tau_\lambda\}_{\lambda\in\Lambda}$ 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 $\mathrm{M}_{\mathrm{g},\tau}$ as soon as g is close enough to Tf_o. Therefore the information in this subspace will 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 $\|\cdot\|_{\mathbf{w},p}$–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 $(Tf_o)_\lambda = 0$.

Let us define formally $\mathcal{M}_{T,f_o}$:

Definition 3.3 **($\mathcal{M}_{T,f_o}$)** *Given two Hilbert spaces $\mathcal{H}^i$ and $\mathcal{H}^o$, an operator $T : \mathcal{H}^i \to \mathcal{H}^o$, 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:*

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

It is clear that even if the operator we seek to invert is one-to-one, for a given object f_o, the set $\mathcal{M}_{T,f_o}$ need not be reduced to $\{f_o\}$ itself:

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

In this case the minimizer of any $\|\cdot\|_{\mathbf{w},p}$–norm in the set $\mathcal{M}_{\mathrm{Id},f_1}$ is f_1 itself, whatever the choices of the parameters p and $\mathbf{w}=\{w_\lambda\}_{\lambda\in\Lambda}$,... are. Algorithm 3.2 will therefore provide the desired result: if g is arbitrarily close to $Tf_1 = f_1$ then the minimizer of $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau}$ for suitable parameters γ and $\boldsymbol{\tau}=\{\tau_\lambda\}_{\lambda\in\Lambda}$is arbitrarily close to f_1.

However, this is not the case in the following example, where T is also an invertible operator in $\mathbb{R}^2$:

Example 2 *Consider $T : \mathbb{R}^2 \to \mathbb{R}^2$, the bounded and linear operator defined by:*

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

- *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* $\mathcal{M}_{T,f_a} = \{f : (Tf)_1 = (Tf_a)_1\} = \{f : 2f_1 + f_2 = 3a\}$.

The element in $\mathcal{M}_{T,f_a}$ 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.(10) (for $p = 1$) 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.

3.5.2. Stability theorem

We now state the stability theorem for Algorithm 3.2. It turns out that Algorithm 3.2 is regularizing for elements f in a particular set: $\mathcal{H}^i_{T,\mathbf{w},p}$, which is defined by:

Definition 3.4 ($\mathcal{H}^i_{T,\mathbf{w},p}$) *Given a Hilbert space $\mathcal{H}^i$, $\mathcal{H}^i_{T,\mathbf{w},p}$ is the subset of elements of $\mathcal{H}^i$ that verify: f_o is in $\mathcal{H}^i_{T,\mathbf{w},p}$ 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 $|\!|\!|.|\!|\!|_{\mathbf{w},p}$-norm.*

When $p > 1$, then $\mathcal{H}^i_{T,\mathbf{w},p} = \mathcal{H}^i$, regardless of T. This is not true if $p = 1$, even if $\ker T = \{0\}$. Algorithm 3.2 is regularizing for elements f in $\mathcal{H}^i_{T,\mathbf{w},p}$, and the minimizer obtained in the limit $\|Tf_o - g\|_{\mathcal{H}^o}$ goes to zero is exactly the minimizer of the $|\!|\!|.|\!|\!|_{\mathbf{w},p}$-norm in the set $\mathcal{M}_{T,f_o}$. 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 non-negative sequence $\boldsymbol{\tau}=\{\tau_\lambda\}_{\lambda\in\Lambda}$, define $f^\star_{\gamma,\mathbf{w},p,\tau;g}$ 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:

$$\text{(i)}\ \lim_{\epsilon\to 0}\gamma(\epsilon) = 0 \quad \text{and} \quad \lim_{\epsilon\to 0}\frac{\epsilon^2}{\gamma(\epsilon)} = 0$$

$$\text{(ii)}\ \forall\lambda\in\Lambda,\ \lim_{\epsilon\to 0}\tau_\lambda(\epsilon) = 0 \quad \text{and} \quad \exists\,\delta > 0,\ \text{s.t:}\ \ [\ \epsilon < \delta \Rightarrow \forall\lambda\in\Lambda,\ \tau_\lambda(\epsilon) > \epsilon\ ]$$

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

$$\lim_{\epsilon\to 0}\left[\sup_{\|g - Tf_o\|_{\mathcal{H}^o}\le\epsilon}\|f^\star_{\gamma(\epsilon),\mathbf{w},p,\tau(\epsilon);\ g} - f^\dagger_o\|_{\mathcal{H}^i}\right] = 0\ ,$$

where $f^\dagger_o$ is the unique element of minimum $|\!|\!|\ |\!|\!|_{\mathbf{w},p}$–norm in the set $\mathcal{M}_{T,f_o}$.

We will prove this stability theorem in a similar manner as Theorem 2.3 is proved in [4]. The proof proceeds as follows: first we prove that the norms $|\!|\!|f^\star_{\gamma(\epsilon),\mathbf{w},p,\tau(\epsilon);\ g}|\!|\!|_{\mathbf{w},p}$ are uniformly bounded. Secondly, we prove that when f_o is in $\mathcal{H}^i_{T,\mathbf{w},p}$, any sequence $\{f^\star_{\gamma(\epsilon_n),\mathbf{w},p,\tau(\epsilon_n);\ g_n}\}_n$ converges weakly to $f^\dagger_o$ when ϵ_n converges to 0. (Here g_n is any element in $\mathcal{H}^o$ verifying $\|g_n - Tf_o\|_{\mathcal{H}^o} \le \epsilon_n$). Finally we prove strong convergence of the $\{f^\star_{\gamma(\epsilon_n),\mathbf{w},p,\tau(\epsilon_n);\ g_n}\}_n$ which proves Theorem 3.5.

The main addition of the proof given here (case with adaptive projection in the discrepancy term) compared to one provided in [4] (case with classical quadratic discrepancy term) is the analysis of the behavior of the adaptive projection operators $\mathrm{M}_{\mathrm{g},\tau}$ when $||g - Tf_o|| \to 0$. More exactly, we prove in Lemmas 3.6 and 3.7 that condition ii) in Theorem 3.5 is needed to obtain the weak convergence of the adaptive projection operators $\mathrm{M}_{\mathrm{g},\tau}$ when $||g - Tf_o|| \to 0$ in section 3.5.3. Using this weak convergence allows to adapt the proof of Theorem 2.3 provided in [4], to give the full proof of Theorem 3.5 in 3.5.4.

3.5.3. Weak convergence of the projection operators

To prove Theorem 3.5, we first examine the behavior of the projections $M_{g(\epsilon),\tau(\epsilon)}$ as ϵ goes to zero in the next two lemmas. The first lemma (Lemma 3.6) gives necessary and sufficient conditions on the sequence $\boldsymbol{\tau}=\{\tau_\lambda\}_{\lambda\in\Lambda}$ 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 $M_{Tf_o,0}$. The second lemma (Lemma 3.7) refines these conditions, so that in addition, the sequence $M_{g(\epsilon),\tau(\epsilon)}$ converges strongly to $M_{Tf_o,0}$ on the set: $T(\mathcal{M}_{T,f_0})$.

Lemma 3.6 *For $f \in \mathcal{H}^i$, let $\{g(\epsilon, f)\}_{\epsilon>0}$ be an arbitrary family of elements in $\mathcal{H}^o$ that satisfy* $\|g(\epsilon, f) - Tf\|_{\mathcal{H}^o} < \epsilon$, $\forall \epsilon > 0$.

(i) $\forall h \in \mathcal{H}^o$, $M_{g(\epsilon,f),\tau(\epsilon)}h$ *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) $\forall \epsilon \in (0, \delta(\lambda))$, $\left|[g(\epsilon, f)]_\lambda\right| > \tau_\lambda$,

(b) $\forall \epsilon \in (0, \delta(\lambda))$, $\left|[g(\epsilon, f)]_\lambda\right| \leq \tau_\lambda$.

(ii) $M_{g(\epsilon,f),\tau(\epsilon)}$ *converges weakly, independently of the choice of f and of the family $g(\epsilon, f)$, as ϵ goes to 0* ***if and only if*** $\forall \lambda$ *: both (a) and (b) hold, with*

(a) $\exists\ \delta(\lambda)$ *such that* $\forall \epsilon \in (0, \delta(\lambda))$, $\tau_\lambda(\epsilon) > \epsilon$

(b) $\lim_{\epsilon \to 0} \tau_\lambda(\epsilon) = 0$

In that case, the weak-limit operator is necessarily $\mathrm{M}_{Tf,0}$.

(iii) When conditions ii.(a) and ii.(b) above hold:
if $h(\epsilon)$ converges weakly to h, then $M_{g(\epsilon,f),\tau(\epsilon)}h(\epsilon)$ *converges weakly to* $\mathrm{M}_{Tf,0}\,h$ *as ϵ goes to* 0.

Proof of Lemma 3.6 Let us examine the behavior of $M_{g(\epsilon,f),\tau(\epsilon)}$ coordinate by coordinate. Since $\left[M_{g(\epsilon,f),\tau(\epsilon)}h\right]_\lambda$ equals either h_λ or 0, depending on whether or not $\left|[g(\epsilon, f)]_\lambda\right| > \tau_\lambda(\epsilon)$, 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 $\left|[g(\epsilon, f)]_\lambda\right| > \tau_\lambda(\epsilon)$ for $\epsilon < \delta(\lambda)$. In this case, $\left[M_{g(\epsilon,f),\tau(\epsilon)}h\right]_\lambda = h_\lambda$ for $\epsilon < \delta(\lambda)$.

Or there exists some $\delta(\lambda) > 0$ such that $\left|[g(\epsilon, f)]_\lambda\right| \leq \tau_\lambda(\epsilon)$ for $\epsilon < \delta(\lambda)$. In this case, $\left[M_{g(\epsilon,f),\tau(\epsilon)}h\right]_\lambda = 0$ 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 $\left|[g(\epsilon, f) - Tf]_\lambda\right| \leq \|g(\epsilon, f) - Tf\|_{\mathcal{H}^o} \leq \epsilon$, the set of values that can be assumed by $|g(\epsilon, f)_\lambda|$ is exactly $\left[\ Tf - \epsilon, Tf + \epsilon\right]$ (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 $\left\{|g(\epsilon, f)_\lambda|\right\} = \left[\ |Tf_\lambda| - \epsilon, |Tf_\lambda| + \epsilon\ \right]$. Therefore, one needs either: $\left[\epsilon < \delta(\lambda) \Rightarrow \tau_\lambda(\epsilon) > |Tf_\lambda| + \epsilon\right]$ or $\left[\epsilon < \delta(\lambda) \Rightarrow \tau_\lambda(\epsilon) \leq |Tf_\lambda| - \epsilon\right]$. 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:

- $\forall \lambda$, $\lim_{\epsilon \to 0} \tau_\lambda(\epsilon) = 0$: 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. $Tf_\lambda \neq 0$: $\lim_{\epsilon\to 0}\tau_\lambda(\epsilon) = 0$ 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 $\mathrm{M}_{Tf,0}$ and finishes the proof of the second part of Lemma 3.6.

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

$$\begin{aligned} & \left|\left[M_{g(\epsilon,f),\tau(\epsilon)}h(\epsilon) - \mathrm{M}_{Tf,0}\, h\,\right]_\lambda\right| & (16)\\ = & \left|\left[M_{g(\epsilon,f),\tau(\epsilon)}(h(\epsilon) - h) + (M_{g(\epsilon,f),\tau(\epsilon)} - \mathrm{M}_{Tf,0})h\,\right]_\lambda\right| & (17)\\ = & \left|\left[M_{g(\epsilon,f),\tau(\epsilon)}(h(\epsilon) - h)\,\right]_\lambda\right| + \left|\left[M_{g(\epsilon,f),\tau(\epsilon)}h - \mathrm{M}_{Tf,0}\, h\,\right]_\lambda\right| & (18)\end{aligned}$$

The second term vanishes as ϵ goes to 0 because $M_{g(\epsilon,f),\tau(\epsilon)}$ converges weakly to $\mathrm{M}_{Tf,0}$ 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, $\left|\left[M_{g(\epsilon,f),\tau(\epsilon)}(h(\epsilon) - h)\,\right]_\lambda\right| = \left|\left[h(\epsilon) - h\,\right]_\lambda\right|$, for $\epsilon < \delta(\lambda)$; and the weak convergence of $h(\epsilon)$ to h allows to conclude that $\left|\left[M_{g(\epsilon,f),\tau(\epsilon)}(h(\epsilon) - h)\,\right]_\lambda\right| \to 0$

This proves that $M_{g(\epsilon,f),\tau(\epsilon)}h(\epsilon)$ converges weakly to $\mathrm{M}_{Tf,0}\, h$ and finishes the proof of Lemma 3.6. ∎

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

Lemma 3.7 *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:*

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

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

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

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

Proof of Lemma 3.7: The first part of Lemma 3.7 results from properties of orthogonal projections. If P_1 and P_2 are two orthogonal projections, then:

$$\begin{aligned} P_1\ P_2 &= P_2\ P_1 \\ \ker(P_2) \subset \ker(P_1) &\Leftrightarrow P_1P_2 = P_1. \end{aligned}$$

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

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

When f and ϵ are fixed, the right hand side holds for any $g(\epsilon, f)$ if and only if $\left[(Tf)_\lambda = 0 \Rightarrow \epsilon < \tau_\lambda(\epsilon)\right]$ which proves the first part of Lemma 3.7.

For f in $\mathcal{H}^i_{T,\mathbf{w},p}$, $f^\dagger$ 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 3.7. ∎

3.5.4. Proof of Theorem 3.5

With the help of these two lemma, we can now proceed to the

Proof of Theorem 3.5: Let us consider f_o in $\mathcal{H}^i_{T,\mathbf{w},p}$, i.e. f_o verifies that $\mathcal{M}_{T,f_o}$ has a unique minimizer $\|\!|\ \|\!|_{\mathbf{w},p}$–norm. We note this minimizer $f_o^\dagger$. We fix the following sequences: $\{\epsilon_n\}_n$ such that $\epsilon_n \underset{n\to\infty}{\longrightarrow} 0$, $\{g_n\}_n$ such that $\forall n,\ \|g_n - Tf_o\|_{\mathcal{H}^o} \leq \epsilon_n$, and $\{\gamma_n\}_n \overset{def}{=} \{\gamma(\epsilon_n)\}_n$ and $\{\tau_n\}_n \overset{def}{=} \{\tau(\epsilon_n)\}_n$ that verify conditions i) to ii) in Theorem 3.5. For every n, we choose a minimizer $f_n^\star \overset{def}{=} f^\star_{\gamma_n,\mathbf{w},p,\tau_n;\,g_n}$ of the functional $J_n(f) \overset{def}{=} J_{\gamma_n,\mathbf{w},p,\tau_n;\,g_n}(f) = \|\,\mathrm{M}_{g_n,\tau_n}(Tf - g_n)\|^2_{\mathcal{H}^o} + \gamma_n \|\!| f \|\!|^p_{\mathbf{w},p}$.

We want to prove that for any such choice of the ϵ_n, g_n, γ_n, τ_n and $f_n^\star$, the sequence $f_n^\star$ converges strongly in $\mathcal{H}^i$ to $f_o^\dagger$, where $f_o^\dagger$ is the unique minimizer of the $\|\!|.\|\!|_{\mathbf{w},p}$-norm in the set $\mathcal{M}_{T,f_o} = \{f : (Tf)_\lambda = (Tf_o)_\lambda,\ \forall \lambda\ s.t.\ (Tf_o)_\lambda \neq 0\}$. We will also note $M_n \overset{def}{=} \mathrm{M}_{g_n,\tau_n}$.

a) The sequences $\{\|\!| f_n^\star \|\!|_{\mathbf{w},p}\}_n$ and $\{\|f_n^\star\|_{\mathcal{H}^i}\}_n$ are uniformly bounded:

By definition of J_n, $\forall n$:

$$\begin{aligned} \|\!| f_n^\star \|\!|^p_{\mathbf{w},p} &\leq \tfrac{1}{\gamma_n} J_n(f_n^\star) \\ \text{so that}\quad \|\!| f_n^\star \|\!|^p_{\mathbf{w},p} &\leq \tfrac{1}{\gamma_n} J_n(f_o^\dagger) \quad \text{since } f_n^\star \text{ minimizes } J_n. \end{aligned}$$

But:

$$\begin{aligned} J_n(f_o^\dagger) &= \|M_n(Tf_o^\dagger - g_n)\|^2_{\mathcal{H}^o} + \gamma_n \|\!| f_o^\dagger \|\!|^p_{\mathbf{w},p} \\ &\leq \|M_n(Tf_o^\dagger - Tf_o)\|^2_{\mathcal{H}^o} + \|M_n(Tf_o - g_n)\|^2_{\mathcal{H}^o} + \gamma_n \|\!| f_o^\dagger \|\!|^p_{\mathbf{w},p} \\ &\leq \|M_n(Tf_o^\dagger - Tf_o)\|^2_{\mathcal{H}^o} + \|\!| M_n \|\!|^2 . \|(Tf_o - g_n)\|^2_{\mathcal{H}^o} + \gamma_n \|\!| f_o^\dagger \|\!|^p_{\mathbf{w},p} \\ &\leq \|M_n(Tf_o^\dagger - Tf_o)\|^2_{\mathcal{H}^o} + \epsilon_n^2 + \gamma_n \|\!| f_o^\dagger \|\!|^p_{\mathbf{w},p} \end{aligned}$$

where we used $\|\!| M_n \|\!|^2 \leq 1$ and $\|Tf_o - g_n\| \leq \epsilon_n$ in the last equation. Hence

$$\forall n,\ \|\!| f_n^\star \|\!|^p_{\mathbf{w},p} \leq \frac{\|M_n(Tf_o^\dagger - Tf_o)\|^2_{\mathcal{H}^o}}{\gamma_n} + \frac{\epsilon_n^2}{\gamma_n} + \|\!| f_o^\dagger \|\!|^p_{\mathbf{w},p}. \tag{19}$$

Since condition ii) of Theorem 3.5 is satisfied, we can use Lemma 3.7.(2). It follows that if n is large enough, $M_n T f_o^\dagger = M_n T f_o$. Moreover, $\frac{\epsilon_n^2}{\gamma_n} \underset{n\to\infty}{\longrightarrow} 0$ by condition i) of Theorem 3.5. This proves that $\{\|\!| f_n^\star \|\!|_{\mathbf{w},p}\}_n$ is uniformly bounded.

Since $\mathbf{w}$ is bounded below by $c > 0$ and $p \leq 2$, the $\|.\|_{\mathcal{H}^i}$-norm is bounded above by $c^{-\frac{1}{p}} \|\!|.\|\!|_{\mathbf{w},p}$:

$$|f_\lambda| = \left(|f_\lambda|^p\right)^{\frac{1}{p}} \leq \left(\tfrac{w_\lambda}{c}|f_\lambda|^p\right)^{\frac{1}{p}} \leq \Big(\sum_{\lambda\in\Lambda} \tfrac{w_\lambda}{c}|f_\lambda|^p\Big)^{\frac{1}{p}} = c^{-\frac{1}{p}} \|\!| f \|\!|_{\mathbf{w},p} \tag{20}$$

so that:

$$\|f\|_{\mathcal{H}^i}^2 = \sum_{\lambda\in\Lambda} |f_\lambda|^2 \leq \sum_{\lambda\in\Lambda} \frac{w_\lambda}{c}|f_\lambda|^p|f_\lambda|^{2-p} \leq \sum_{\lambda\in\Lambda} \frac{w_\lambda}{c}|f_\lambda|^p\left[c^{-\frac{1}{p}}\|f\|_{\mathbf{w},p}\right]^{2-p} \tag{21}$$

$$\|f\|_{\mathcal{H}^i}^2 \leq \frac{1}{c}\|f\|_{\mathbf{w},p}^p\left[c^{-\frac{1}{p}}\|f\|_{\mathbf{w},p}\right]^{2-p} = c^{-\frac{2}{p}}\|f\|_{\mathbf{w},p}^2 \tag{22}$$

Hence, the sequence $\{f_n^\star\}$ is also uniformly bounded in $\mathcal{H}^i$.

b) The sequence $\{f_n^\star\}_n$ converges weakly to $f_o^\dagger$:

Since it is uniformly bounded in $\mathcal{H}^i$, the sequence $\{f_n^\star\}_n$ has at least one weakly convergent subsequence $\{f_k^\star\}_k$. Let us denote its weak limit $\widetilde{f}$. We shall now prove that $\widetilde{f} = f_o^\dagger$.

Since $f_k^\star$ is a minimizer of J_k obtained through the iterative algorithm, 3.2, it verifies the fixed point equation: $f_k^\star = \mathbf{S}_{\gamma_k\mathbf{w},\mathrm{p}}(f_k^\star + T^*M_kg_k - T^*M_kTf_k^\star)$. We note $h_k = f_k^\star + T^*M_kg_k - T^*M_kTf_k^\star$, so that $f_k^\star = \mathbf{S}_{\gamma_k\mathbf{w},\mathrm{p}}(h_k)$. By definition of the weak limit, it follows that:

$$\begin{aligned}
\forall\lambda,\ \widetilde{f}_\lambda &= \lim_{k\to\infty} S_{\gamma_k w_\lambda}\big((h_k)_\lambda\big) \\
&= \lim_{k\to\infty}[(h_k)_\lambda] + \lim_{k\to\infty}[S_{\gamma_k w_\lambda}((h_k)_\lambda) - (h_k)_\lambda] && \text{but } \lim_{k\to\infty}\gamma_k w_\lambda = 0 \\
\text{So, } \forall\lambda,\ \widetilde{f}_\lambda &= \lim_{k\to\infty}[(h_k)_\lambda] && \text{since } \forall x,\ S_v(x) \xrightarrow[v\to 0]{} x \\
&= \lim_{k\to\infty}[(f_k^\star + T^*M_kg_k - T^*M_kTf_k^\star)_\lambda] \\
&= \widetilde{f}_\lambda + \lim_{k\to\infty}[(T^*M_kg_k - T^*M_kTf_k^\star)_\lambda] && \text{since } (f_k^\star)_\lambda \xrightarrow[k\to\infty]{} \widetilde{f}_\lambda.
\end{aligned}$$

As a result: $\forall\lambda,\ \lim_{k\to\infty}[(T^*M_kg_k - T^*M_kTf_k^\star)_\lambda] = 0$.

But since $\|g_k - Tf_o\|_{\mathcal{H}^o} \leq \epsilon_k$, then $\|T^*M_k(g_k - Tf_o)\|_{\mathcal{H}^i} \leq \|T^*\|\|M_k\|\epsilon_k < \epsilon_k$. This proves that for all λ:

$$\lim_{k\to\infty}[(T^*M_kTf_o - T^*M_kTf_k^\star)_\lambda] = 0. \tag{23}$$

Moreover, from Lemma 3.6.(2), we know that $\{M_k(Tf_o)\}_k$ converges weakly to $M_{Tf_o,0}(Tf_o) = Tf_o$. Together with the continuity of T^*, this leads to:

$$T^*M_kTf_k^\star \xrightarrow[k\to\infty]{w} T^*Tf_o. \tag{24}$$

On the other hand, $f_k^\star$ converges weakly to $\widetilde{f}$. Using the continuity of T, we get $Tf_k^\star \xrightarrow[k\to\infty]{w} T\widetilde{f}$. From Lemma 3.6.(3), this also implies $\{M_kTf_k^\star\}_k \xrightarrow[k\to\infty]{w} \mathrm{M}_{Tf_o,0}\,T\widetilde{f}$. and it follows from the continuity of T^* that:

$$T^*M_kTf_k^\star \xrightarrow[k\to\infty]{w} T^*M_{Tf_o,0}T\widetilde{f}. \tag{25}$$

Plugging this last result in Eq. (24), we obtain the equality:

$$T^*M_{Tf_o,0}T\widetilde{f} = T^*Tf_o \tag{26}$$

Since $M_{Tf_o,0}(Tf_o) = Tf_o$, the previous equality reduces to: $T^*M_{Tf_o,0}T(\widetilde{f} - f_o) = 0$. Taking the scalar product with $\widetilde{f} - f_o$, we obtain:

$$\begin{aligned}
&\big\langle \widetilde{f} - f_o,\ T^*M_{Tf_o,0}T(\widetilde{f} - f_o)\big\rangle = 0 \\
&\Leftrightarrow \big\langle M_{Tf_o,0}T(\widetilde{f} - f_o),\ M_{Tf_o,0}T(\widetilde{f} - f_o)\big\rangle = 0 \\
&\Leftrightarrow \|M_{Tf_o,0}T(\widetilde{f} - f_o)\|_{\mathcal{H}^o}^2 = 0 \\
&\Leftrightarrow M_{Tf_o,0}T(\widetilde{f} - f_o) = 0 \\
&\Leftrightarrow M_{Tf_o,0}T\widetilde{f} = Tf_o
\end{aligned}$$

We used for the first equality that $M_{Tf_o,0} = M^*_{Tf_o,0} = M^2_{Tf_o,0}$. This proves that $\widetilde{f}$ belongs to the set $\mathcal{M}_{T,f_o}$.

Let us now prove that $|||\widetilde{f}|||_{\mathbf{w},p} \le |||f_o^\dagger|||_{\mathbf{w},p}$. Because of the weak convergence of the $f_n^\star$ to $\widetilde{f}$, for all λ, the non-negative sequence $\{w_\lambda |f^\star_{n\lambda}|\}_n$ converges to $w_\lambda|\widetilde{f}_\lambda|$. One can then use Fatou's lemma to obtain:

$$|||\widetilde{f}|||^p_{\mathbf{w},p} = \sum_\lambda \lim_{n\to\infty} \{w_\lambda |f^\star_{n\lambda}|\}_n \le \lim_{n\to\infty} \sum_\lambda \{w_\lambda |f^\star_{n\lambda}|\}_n = \lim_{n\to\infty} |||f_n^\star|||^p_{\mathbf{w},p}$$

But we proved earlier that $\limsup_n |||f_n^\star|||^p_{\mathbf{w},p} \le |||f_o^\dagger|||^p_{\mathbf{w},p}$. Therefore, we get:

$$|||\widetilde{f}|||^p_{\mathbf{w},p} \le \lim_{n\to\infty} |||f_n^\star|||^p_{\mathbf{w},p} \le |||f_o^\dagger|||^p_{\mathbf{w},p} \tag{27}$$

By definition, $f_o^\dagger$ is the unique minimizer of the $|||.|||_{\mathbf{w},p}$-norm in $\mathcal{M}_{f_o}$, so this implies that $\widetilde{f} = f_o^\dagger$.

Hence $f_o^\dagger$ is the only possible accumulation point of the sequence $f_n^\star$. Since we proved that the sequence $\{f_n^\star\}_n$ is uniformly bounded in the $\|.\|_{\mathcal{H}^i}$-norm and that it has a unique accumulation point: $f_o^\dagger$, this allows us to conclude that $f_n^\star$ converges weakly to $f_o^\dagger$.

c) The sequence $\{f_n^\star\}_n$ converges strongly to $f_o^\dagger$:

Replacing $\widetilde{f}$ by its value $f_o^\dagger$ in (27), we get: $|||f_o^\dagger|||^p_{\mathbf{w},p} \le \lim_{n\to\infty} |||f_n^\star|||^p_{\mathbf{w},p} \le |||f_o^\dagger|||^p_{\mathbf{w},p}$ which proves that the sequence $\{|||f_n^\star|||^p_{\mathbf{w},p}\}_n$ converges to $|||f_o^\dagger|||^p_{\mathbf{w},p}$. We shall see now that the two results we obtained so far:

$$f_n^\star \xrightarrow[n\to\infty]{w} f_o^\dagger \tag{28}$$

$$|||f_n^\star|||_{\mathbf{w},p} \xrightarrow[n\to\infty]{} |||f_o^\dagger|||_{\mathbf{w},p}\ , \tag{29}$$

imply the strong convergence of the sequence $\{f_n^\star\}_n$ to $f_o^\dagger$. (This argument closely follows [4].)

Let us prove that $\{\|f_n^\star\|_{\mathcal{H}^i}\}_n$ converges to $\|f_o^\dagger\|_{\mathcal{H}^i}$. We have:

$$\Big| \|f_n^\star\|^2_{\mathcal{H}^i} - \|f_o^\dagger\|^2_{\mathcal{H}^i} \Big| = \Big| \sum_\lambda \big(|f^\star_{n\lambda}|^2 - |f^\dagger_{o\lambda}|^2\big) \Big| \le \sum_\lambda \Big| \, |f^\star_{n\lambda}|^2 - |f^\dagger_{o\lambda}|^2 \, \Big| \tag{30}$$

Writing $x^2 = (x^p)^{\frac{2}{p}}$ and using the derivability of $x \to x^{\frac{2}{p}}$, one can bound the last term:

$$\Big| |f^\star_{n\lambda}|^2 - |f^\dagger_{o\lambda}|^2 \Big| \le \frac{2}{p} \max\{\big(|f^\star_{n\lambda}|^p\big)^{\frac{2}{p}-1}, \big(|f^\dagger_{o\lambda}|^p\big)^{\frac{2}{p}-1}\} \Big| |f^\star_{n\lambda}|^p - |f^\dagger_{o\lambda}|^p \Big| \tag{31}$$

$$\le \frac{2}{p} \max\{|f^\star_{n\lambda}|^{2-p}, |f^\dagger_{o\lambda}|^{2-p}\} \Big| |f^\star_{n\lambda}|^p - |f^\dagger_{o\lambda}|^p \Big| \tag{32}$$

$$\le \frac{2}{pc} \max\{|f^\star_{n\lambda}|^{2-p}, |f^\dagger_{o\lambda}|^{2-p}\} \Big| w_\lambda |f^\star_{n\lambda}|^p - w_\lambda |f^\dagger_{o\lambda}|^p \Big| \tag{33}$$

We saw in Eq. (42) that for any $f \in \mathcal{H}^i$ and $\lambda_o \in \Lambda$ $|f_{\lambda_o}| \le c^{\frac{1}{p}} |||f|||_{\mathbf{w},p}$. Plugging this into Eq. (33) and summing over λ, we get:

$$\Big| \|f_n^\star\|^2_{\mathcal{H}^i} - \|f_o^\dagger\|^2_{\mathcal{H}^i} \Big| \le \frac{2}{p} c^{-\frac{2}{p}} \max\{|||f_n^\star|||^{2-p}_{\mathbf{w},p}, |||f_o^\dagger|||^{2-p}_{\mathbf{w},p}\} \sum_{\lambda\in\Lambda} \Big| w_\lambda |f^\star_{n\lambda}|^p - w_\lambda |f^\dagger_{o\lambda}|^p \Big| \tag{34}$$

Since $\{|\!|\!| f_n^\star |\!|\!|_{\mathbf{w},p}^p\}_n$ converges to $|\!|\!| f_o^\dagger |\!|\!|_{\mathbf{w},p}^p$, for n large enough, $\max\{|\!|\!| f_n^\star |\!|\!|_{\mathbf{w},p}^{2-p}, |\!|\!| f_o^\dagger |\!|\!|_{\mathbf{w},p}^{2-p}\}$ is bounded by $2|\!|\!| f_o^\dagger |\!|\!|_{\mathbf{w},p}^{2-p}$. Defining $g_{c,p,f_o} = \frac{4}{p} c^{-\frac{2}{p}} \; |\!|\!| f_o^\dagger |\!|\!|_{\mathbf{w},p}^{2-p}$, we get:

$$\begin{aligned} \left| \|f_n^\star\|_{\mathcal{H}^i}^2 - \|f_o^\dagger\|_{\mathcal{H}^i}^2 \right| &\leq g_{c,p,f_o} \textstyle\sum_{\lambda\in\Lambda} \left| w_\lambda |f_{n\,\lambda}^\star|^p - w_\lambda |f_{o\,\lambda}^\dagger|^p \right| \\ &\leq g_{c,p,f_o} \textstyle\sum_\lambda \Big(w_\lambda |f_{n\,\lambda}^\star|^p + w_\lambda |f_{o\,\lambda}^\dagger|^p - 2 w_\lambda \min\{|f_{n\,\lambda}^\star|, |f_{o\,\lambda}^\dagger|\}^p \Big) \\ &\leq g_{c,p,f_o} \Big(|\!|\!| f_n^\star |\!|\!|_{\mathbf{w},p}^p + |\!|\!| f_o^\dagger |\!|\!|_{\mathbf{w},p}^p - 2 \textstyle\sum_\lambda w_\lambda \min\{|f_{n\,\lambda}^\star|, |f_{o\,\lambda}^\dagger|\}^p \Big) \end{aligned} \quad (35)$$

We already know that $|\!|\!| f_n^\star |\!|\!|_{\mathbf{w},p}^p \xrightarrow[n\to\infty]{} |\!|\!| f_o^\dagger |\!|\!|_{\mathbf{w},p}^p$, we shall see now that the same holds for the last term in the previous inequality. Let us define the sequence $\{u_{n\lambda}\}_n$ for each λ by $u_{n\lambda} = w_\lambda \min\{|f_{n\,\lambda}^\star|, |f_{o\,\lambda}^\dagger|\}^p$. The weak convergence of the $f_n^\star$ to $f_o^\dagger$ implies that for each λ, $u_{n\lambda} \xrightarrow[n\to\infty]{} w_\lambda |f_{o\,\lambda}^\dagger|^p$. Moreover, for all n, $0 \leq u_{n\lambda} \leq w_\lambda |f_{o\,\lambda}^\dagger|^p$ and $\sum_\lambda w_\lambda |f_{o\,\lambda}^\dagger|^p = |\!|\!| f_o^\dagger |\!|\!|_{\mathbf{w},p}^p < \infty$ so that by the dominated convergence theorem, $\lim\limits_{n\to\infty} \sum_\lambda u_{n\lambda} = \sum_\lambda \lim\limits_{n\to\infty} u_{n\lambda}$. Replacing the $u_{n\lambda}$ and their limits by their value, we obtain:

$$\lim_{n\to\infty} \sum_\lambda w_\lambda \min\{|f_{n\,\lambda}^\star|, |f_{o\,\lambda}^\dagger|\}^p = |\!|\!| f_o^\dagger |\!|\!|_{\mathbf{w},p}^p.$$

Hence:

$$\Big(|\!|\!| f_n^\star |\!|\!|_{\mathbf{w},p}^p + |\!|\!| f_o^\dagger |\!|\!|_{\mathbf{w},p}^p - 2 \sum_\lambda w_\lambda \min\{|f_{n\,\lambda}^\star|, |f_{o\,\lambda}^\dagger|\}^p \Big) \xrightarrow[n\to\infty]{} |\!|\!| f_o^\dagger |\!|\!|_{\mathbf{w},p}^p + |\!|\!| f_o^\dagger |\!|\!|_{\mathbf{w},p}^p - 2 |\!|\!| f_o^\dagger |\!|\!|_{\mathbf{w},p}^p = 0$$

so that by taking the limit as n goes to ∞ in Eq.(35), we can conclude that

$$\|f_n^\star\|_{\mathcal{H}^i} \xrightarrow[n\to\infty]{} \|f_o^\dagger\|_{\mathcal{H}^i}.$$

Using the identity $\|f_n^\star - f_o^\dagger\|_{\mathcal{H}^i} = \|f_n^\star\|_{\mathcal{H}^i} + \|f_o^\dagger\|_{\mathcal{H}^i} - 2\langle f_n^\star,\ f_o^\dagger \rangle$, this last result combined with the weak convergence of the $f_n^\star$ to $f_o^\dagger$ proves that the sequence $\{f_n^\star\}_n$ converges strongly in $\mathcal{H}^i$ to $f_o^\dagger$. ∎

Note that we only assumed f_o is in $\mathcal{H}^i_{T,\mathbf{w},p}$ to obtain stability. It could very well be that the functional $J_{\gamma_n,\mathbf{w},p,\tau_n;g_n}$ has several minimizers, in that case, depending on the choice of the starting element for the iterative algorithm 3.2, the element f_n^* might have different values. As a result, the sequence $\{f_n^*\}_n$ is not fixed by the parameters ϵ_n, γ_n, τ_n and g_n. However no matter which of these sequences $\{f_n^*\}_n$ we consider, it will converge strongly to $f_o^\dagger$.

4. Adaptive discrepancy terms via "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 canceling 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" $\mathrm{M}_{\mathrm{g},\tau,\mu}$ with dampening parameter μ and corresponding to the orthogonal adaptive projection $\mathrm{M}_{\mathrm{g},\tau}$ is

$$\mathrm{M}_{\mathrm{g},\tau,\mu} = \mathrm{M}_{\mathrm{g},\tau} + \mu(\mathrm{Id} - \mathrm{M}_{\mathrm{g},\tau}) \tag{36}$$

or more formally:

Definition 4.1 *Given an orthonormal basis if* $\mathcal{H}^o$, $\boldsymbol{\beta}=\{\beta_\lambda\}_{\lambda\in\Lambda}$, *an element* g *in* $\mathcal{H}^o$, *a sequence of non-negative thresholds* $\boldsymbol{\tau}=\{\tau_\lambda\}_{\lambda\in\Lambda}$ *and a scalar* $\mu > 0$, $\mathrm{M}_{\mathrm{g},\tau,\mu}$ *is the map from* $\mathcal{H}^o$ *into itself defined by:*

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

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

$$\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau,\mu}(f) = \left\|\mathrm{M}_{\mathrm{g},\tau,\mu}(Tf-g)\right\|^2_{\mathcal{H}^o} + \gamma |\!|\!| f |\!|\!|^p_{\mathbf{w},p}, \tag{37}$$

via the following iterative algorithm:

Algorithm 4.2

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

Note that in this case, one needs to square the relaxed projection operator in the iterative algorithm. This is because unlike $\mathrm{M}_{\mathrm{g},\tau}$, $\mathrm{M}_{\mathrm{g},\tau,\mu}$ is not a self-adjoint projection. This equation can be easily checked by replacing T by $\mathrm{M}_{\mathrm{g},\tau,\mu}T$ and g by $\mathrm{M}_{\mathrm{g},\tau,\mu}\,g$ in the original functional $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p}}$ of Eq.(4) and in Algorithm 2.1. In practice, we use the fact that $\mathrm{M}_{\mathrm{g},\tau,\mu}{}^2 = \mathrm{M}_{\mathrm{g},\tau,\mu^2}$; 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.(37) (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 ($g \to Tf_o$) and the parameters γ, $\boldsymbol{\tau}=\{\tau_\lambda\}_{\lambda\in\Lambda}$ and μ are chosen accordingly, then the solution converges to f_o when f_o is the unique antecedent of Tf_o.

4.2.1. Stability theorem

The conditions on the parameters to obtain stability in this case 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* $\gamma > 0$, $0 < \mu \le 1$ *and non-negative sequence* $\boldsymbol{\tau}=\{\tau_\lambda\}_{\lambda\in\Lambda}$, *define* $f^\star_{\gamma,\mathbf{w},p,\tau,\mu;\,g}$ *to be a minimizer of* $\mathbf{J}_{\gamma,\mathbf{w},\mathrm{p},\tau,\mu}(f)$ *with observation* g. *If* $\gamma = \gamma(\epsilon)$, $\tau = \tau(\epsilon)$ *and* $\mu = \mu(\epsilon)$ *satisfy:*

(i) $\lim_{\epsilon\to 0} \gamma(\epsilon) = 0$ *and* $\lim_{\epsilon\to 0} \dfrac{\epsilon^2}{\gamma(\epsilon)} = 0$

(ii) $\forall\lambda\in\Lambda,\ \lim_{\epsilon\longrightarrow 0}\tau_\lambda(\epsilon)=0$ *and* $\forall\lambda\in\Lambda,\ \exists\,\delta(\lambda)>0,$ *s.t:* $[\ \epsilon<\delta(\lambda)\Rightarrow\ \tau_\lambda(\epsilon)>\epsilon\ ]$

(iii) $\lim_{\epsilon\longrightarrow 0}\mu(\epsilon)=\mu_o$*, with* $0<\mu_o\leq 1$

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

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

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

The proof of this theorem is very similar to the proof of Theorem 3.5. The weak convergence of the adaptive operators is ensured by conditions ii) and iii) of Theorem 4.3 and the corresponding lemma is Lemma 4.4. This lemma, similarly to Lemmas 3.6 and 3.7, examines the convergence of the operators $\mathrm{M}_{\mathrm{g},\tau,\mu}$ are is proved in section 4.2.2. Section 3.5.4 is then devoted to the proof of Theorem 4.3.

4.2.2. Weak convergence of the projection operators

Lemma 4.4 *Suppose that* $\tau=\tau(\epsilon)$ *and* $\mu=\mu(\epsilon)$ *verify conditions ii) and iii) of Theorem 4.3. Then the two following properties hold:*

(i) For any h *in* $\mathcal{H}^o$*,* $M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}h$ *converges weakly to* $\mathrm{M}^2_{Tf,0,\mu_o}\,h$ *as* ϵ *goes to* 0*.*

(ii) If $h(\epsilon)$ *converges weakly to* h *as* ϵ *goes to* 0*, then* $M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}h(\epsilon)$ *converges weakly to* $\mathrm{M}^2_{Tf,0,\mu_o}\,h$ *as* ϵ *goes to* 0*.*

Proof of Lemma 4.4: In the proof of Lemma 3.6, we have seen that under conditions imposed on $\tau(\epsilon)$ (condition ii) of Theorem 4.3), the following happens:

- for λ s.t. $Tf_\lambda\neq 0$: $\lim_{\epsilon\to 0}\tau_\lambda(\epsilon)=0$ 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: $M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}(\beta_\lambda)=\beta_\lambda$ for any $g(\epsilon,f)$ and any $\epsilon<\delta(\lambda,f)$; and in the second case: $M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}(\beta_\lambda)=\mu(\epsilon)^2\beta_\lambda$ for any $g(\epsilon,f)$ and any $\epsilon<\delta(\lambda)$. Since $\mu(\epsilon)$ converges to some μ_o by assumption (condition iii) of Theorem 4.3), it follows that $M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}h$ converges to $M^2_{Tf_o,0,\mu_o}h$ as (ϵ) goes to 0. This proves the first part of Lemma 4.4.

To prove the second part of Lemma 4.4, we use again the splitting trick we used in 3.6.(3):

$$\left|\left[M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}h(\epsilon)-\mathrm{M}^2_{Tf,0,\mu_o}\,h\ \right]_\lambda\right| \tag{38}$$

$$=\left|\left[M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}(h(\epsilon)-h)+(M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}-\mathrm{M}^2_{Tf,0,\mu_o})h\ \right]_\lambda\right| \tag{39}$$

$$=\left|\left[M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}(h(\epsilon)-h)\ \right]_\lambda\right|+\left|\left[(M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}-\mathrm{M}^2_{Tf,0,\mu_o})h\ \right]_\lambda\right| \tag{40}$$

And the same argument as we used in Lemma 3.6.(3) allows to conclude. ∎

Note that we did not need $0<\mu_o\leq 1$ to prove this lemma.

4.2.3. Proof of Theorem 4.3

Now that the weak convergence of $M^2_{g(\epsilon,f),\tau(\epsilon),\mu(\epsilon)}$ is established, we proceed to the proof of Theorem 4.3.

This proof is very similar to the proof of Theorem 3.5. For the sake of completeness, we give the full details of the first two parts of the proof, indicating by

► ◄

when the argument differs from before. Once we prove that $f_o^\dagger$ is the unique accumulation point of the sequence $\{f_n^\star\}_n$ (weak convergence of the sequence), the proof the strong convergence is strictly identical and we do not repeat it.

Proof of Theorem 4.3: Let us consider f_o in $\mathcal{H}^i$, that verifies that $\mathcal{S}_{T,f_o}$ has a unique minimizer $\|\!| \ \|\!|_{\mathbf{w},p}$–norm. We note this minimizer $f_o^\dagger$. We fix the following sequences: $\{\epsilon_n\}_n$ such that $\epsilon_n \underset{n\to\infty}{\longrightarrow} 0$, $\{g_n\}_n$ such that $\forall n$, $\|g_n - Tf_o\|_{\mathcal{H}^o} \leq \epsilon_n$, and $\{\gamma_n\}_n \overset{def}{=} \{\gamma(\epsilon_n)\}_n$, $\{\mu_n\}_n \overset{def}{=} \{\mu(\epsilon_n)\}_n$ and $\{\tau_n\}_n \overset{def}{=} \{\tau(\epsilon_n)\}_n$ that verify conditions i) to iii) in Theorem 4.3 For every n, we choose a minimizer $f_n^\star \overset{def}{=} f^\star_{\gamma_n,\mathbf{w},p,\tau_n,\mu_n;\,g_n}$ of the functional $J_n(f) \overset{def}{=} J_{\gamma_n,\mathbf{w},p,\tau_n,\mu_n;\,g_n}(f) = \|\,\mathrm{M}_{g_n,\tau_n,\mu_n}(Tf - g_n)\|^2_{\mathcal{H}^o} + \gamma_n \|\!|f|\!\|^p_{\mathbf{w},p}$. We want to prove that for any such choice of the ϵ_n, g_n, γ_n, μ_n, τ_n and $f_n^\star$, the sequence $f_n^\star$ converges strongly in $\mathcal{H}^i$ to $f_o^\dagger$. We will also note $M_n \overset{def}{=} \mathrm{M}_{g_n,\tau_n,\mu_n}$.

a) The sequences $\{\|\!|f_n^\star|\!\|_{\mathbf{w},p}\}_n$ and $\{\|f_n^\star\|_{\mathcal{H}^i}\}_n$ are uniformly bounded:

By definition of J_n, $\forall n$:

$$\begin{aligned} \|\!|f_n^\star|\!\|^p_{\mathbf{w},p} &\leq \tfrac{1}{\gamma_n} J_n(f_n^\star) \\ \text{so that} \quad \|\!|f_n^\star|\!\|^p_{\mathbf{w},p} &\leq \tfrac{1}{\gamma_n} J_n(f_o^\dagger) \quad \text{since } f_n^\star \text{ minimizes } J_n. \end{aligned}$$

► But:

$$\begin{aligned} J_n(f_o^\dagger) &= \|M_n(Tf_o^\dagger - g_n)\|^2_{\mathcal{H}^o} + \gamma_n \|\!|f_o^\dagger|\!\|^p_{\mathbf{w},p} \\ &= \|M_n(Tf_o - g_n)\|^2_{\mathcal{H}^o} + \gamma_n \|\!|f_o^\dagger|\!\|^p_{\mathbf{w},p} \qquad \text{since } Tf_o^\dagger = Tf_o \\ &\leq \|M_n\|^2.\|(Tf_o - g_n)\|^2_{\mathcal{H}^o} + \gamma_n \|\!|f_o^\dagger|\!\|^p_{\mathbf{w},p} \\ &\leq \max\{1, |\mu_n|^2\}.\epsilon_n^2 + \gamma_n \|\!|f_o^\dagger|\!\|^p_{\mathbf{w},p} \qquad \text{since } \|Tf_o - g_n\| \leq \epsilon_n \end{aligned}$$

Hence

$$\forall n,\ \|\!|f_n^\star|\!\|^p_{\mathbf{w},p} \leq \max\{1, |\mu_n|^2\}.\frac{\epsilon_n^2}{\gamma_n} + \|\!|f_o^\dagger|\!\|^p_{\mathbf{w},p}. \tag{41}$$

Since $\frac{\epsilon_n^2}{\gamma_n} \underset{n\to\infty}{\longrightarrow} 0$ and $\mu_n \underset{n\to\infty}{\longrightarrow} \mu_o \in (0,1]$, this proves that $\{\|\!|f_n^\star|\!\|_{\mathbf{w},p}\}_n$ is uniformly bounded. ◄

Moreover, $\mathbf{w}$ is bounded below by $c > 0$ and $p \leq 2$, so the $\|.\|_{\mathcal{H}^i}$-norm is bounded above by $c^{-\frac{1}{p}} \|\!|.|\!\|_{\mathbf{w},p}$:

$$|f_\lambda| = \big(|f_\lambda|^p\big)^{\frac{1}{p}} \leq \big(\tfrac{w_\lambda}{c}|f_\lambda|^p\big)^{\frac{1}{p}} \leq \Big(\sum_{\lambda\in\Lambda} \tfrac{w_\lambda}{c}|f_\lambda|^p\Big)^{\frac{1}{p}} = c^{-\frac{1}{p}} \|\!|f|\!\|_{\mathbf{w},p} \tag{42}$$

so that:

$$\|f\|^2_{\mathcal{H}^i} = \sum_{\lambda\in\Lambda} |f_\lambda|^2 \leq \sum_{\lambda\in\Lambda} \tfrac{w_\lambda}{c}|f_\lambda|^p|f_\lambda|^{2-p} \leq \sum_{\lambda\in\Lambda} \tfrac{w_\lambda}{c}|f_\lambda|^p\big[c^{-\frac{1}{p}}\|\!|f|\!\|_{\mathbf{w},p}\big]^{2-p} \tag{43}$$

$$\|f\|^2_{\mathcal{H}^i} \leq \tfrac{1}{c}\|\!|f|\!\|^p_{\mathbf{w},p}\big[c^{-\frac{1}{p}}\|\!|f|\!\|_{\mathbf{w},p}\big]^{2-p} = c^{-\frac{2}{p}}\|\!|f|\!\|^2_{\mathbf{w},p} \tag{44}$$

Hence, the sequence $\{f_n^\star\}$ is also uniformly bounded in $\mathcal{H}^i$.

b) The sequence $\{f_n^\star\}_n$ converges weakly to $f_o^\dagger$:

Since it is uniformly bounded in $\mathcal{H}^i$, the sequence $\{f_n^\star\}_n$ has at least one weakly convergent subsequence $\{f_k^\star\}_k$. Let us denote its weak limit $\widetilde{f}$. We shall now prove that $\widetilde{f} = f_o^\dagger$.

Since $f_k^\star$ is a minimizer of J_k obtained through the iterative Algorithm 4.2, it verifies the fixed point equation: $f_k^\star = \mathbf{S}_{\gamma_k \mathbf{w},\mathrm{p}}\left(f_k^\star + T^* M_k^2 g_k - T^* M_k^2 T f_k^\star\right)$. We note $h_k = f_k^\star + T^* M_k^2 g_k - T^* M_k^2 T f_k^\star$, so that $f_k^\star = \mathbf{S}_{\gamma_k \mathbf{w},\mathrm{p}}(h_k)$. By definition of the weak limit, it follows that:

$$\begin{array}{rcll}
\forall\lambda,\ \widetilde{f}_\lambda &=& \lim\limits_{k\to\infty} S_{\gamma_k w_\lambda}\big((h_k)_\lambda\big) & \\
&=& \lim\limits_{k\to\infty} [(h_k)_\lambda] + \lim\limits_{k\to\infty} [S_{\gamma_k w_\lambda}((h_k)_\lambda) - (h_k)_\lambda] & \text{but } \lim\limits_{k\to\infty} \gamma_k w_\lambda = 0 \\
\text{So, } \forall\lambda,\ \widetilde{f}_\lambda &=& \lim\limits_{k\to\infty} [(h_k)_\lambda] & \text{since } \forall x,\ S_v(x) \xrightarrow[v\to 0]{} x \\
&=& \lim\limits_{k\to\infty} \left[(f_k^\star + T^* M_k^2 g_k - T^* M_k^2 T f_k^\star)_\lambda\right] & \\
&=& \widetilde{f}_\lambda + \lim\limits_{k\to\infty} \left[(T^* M_k^2 g_k - T^* M_k^2 T f_k^\star)_\lambda\right] & \text{since } (f_k^\star)_\lambda \xrightarrow[k\to\infty]{} \widetilde{f}_\lambda.
\end{array}$$

As a result: $\forall\lambda,\ \lim\limits_{k\to\infty} \left[(T^* M_k^2 g_k - T^* M_k^2 T f_k^\star)_\lambda\right] = 0$.

► Since $\|g_k - Tf_o\| \le \epsilon_k$, then $\|T^* M_k^2 (g_k - Tf_o)\|_{\mathcal{H}^o} \le \|T^*\| \|M_k\|^2 \epsilon_k < \max\{1, |\mu_k|\}^2.\epsilon_k$. Since μ_k converges to $\mu_o \in (0,1]$, and ϵ_k to 0, this proves that for all λ:

$$\lim_{k\to\infty} \left[(T^* M_k^2 T f_o - T^* M_k^2 T f_k^\star)_\lambda\right] = 0. \tag{45}$$

From Lemma 4.4.(1), the sequence $\{M_k^2(Tf_o)\}_k$ converges weakly to $M^2_{Tf_o,0,\mu_o}(Tf_o) = Tf_o$. ◄ Together with the continuity of T^*, this leads to:

$$T^* M_k^2 T f_k^\star \xrightarrow[k\to\infty]{w} T^* T f_o. \tag{46}$$

On the other hand, $f_k^\star$ converges weakly to $\widetilde{f}$. Using the continuity of T, we get $Tf_k^\star \xrightarrow[k\to\infty]{w} T\widetilde{f}$.

► Lemma 4.4.(2) allows then to conclude that $M_k^2 T f_k^\star \xrightarrow[k\to\infty]{w} M^2_{Tf_o,0,\mu_o} T\widetilde{f}$ ◄

and it follows from the continuity of T^* that:

$$T^* M_k^2 T f_k^\star \xrightarrow[k\to\infty]{w} T^* M^2_{Tf_o,0,\mu_o} T\widetilde{f}. \tag{47}$$

Plugging this last result in Eq. (46), we obtain the equality:

$$T^* M^2_{Tf_o,0,\mu_o} T\widetilde{f} = T^* T f_o \tag{48}$$

Note that $M_{Tf_o,0,\mu_o}$ is a self adjoint and that $M^2_{Tf_o,0,\mu_o}(Tf_o) = M_{Tf_o,0,\mu_o}(Tf_o) = Tf_o$. Therefore the previous equality reduces to: $T^* M^2_{Tf_o,0,\mu_o} T(\widetilde{f} - f_o) = 0$. Taking the scalar product with $\widetilde{f} - f_o$, we obtain:

$$\begin{array}{rl}
\blacktriangleright \quad \left\langle \widetilde{f} - f_o,\ T^* M^2_{Tf_o,0,\mu_o} T(\widetilde{f} - f_o)\right\rangle = 0 & \Leftrightarrow\ \left\langle M_{Tf_o,0,\mu_o} T(\widetilde{f} - f_o),\ M_{Tf_o,0,\mu_o} T(\widetilde{f} - f_o)\right\rangle = 0 \\
& \Leftrightarrow\ \|M_{Tf_o,0,\mu} T(\widetilde{f} - f_o)\|^2_{\mathcal{H}^o} = 0 \\
& \Leftrightarrow\ M_{Tf_o,0,\mu_o} T(\widetilde{f} - f_o) = 0 \\
& \Leftrightarrow\ T(\widetilde{f} - f_o) = 0 \quad \text{since } M_{Tf_o,0,\mu_o} \text{ is invertible.} \\
& \Leftrightarrow\ T\widetilde{f} = Tf_o
\end{array}$$

This proves that $\widetilde{f}$ belongs to the set $\mathcal{S}_{T,f_o}$. ◀

Let us now prove that $\|\widetilde{f}\|_{\mathbf{w},p} \leq \|f_o^\dagger\|_{\mathbf{w},p}$. Because of the weak convergence of the $f_n^\star$ to $\widetilde{f}$, for all λ, the non-negative sequence $\{w_\lambda|f_{n\lambda}^\star|\}_n$ converges to $w_\lambda|\widetilde{f}_\lambda|$. One can then use Fatou's lemma to obtain:

$$\|\widetilde{f}\|_{\mathbf{w},p}^p = \sum_\lambda \lim_{n\to\infty} \{w_\lambda|f_{n\lambda}^\star|\}_n \leq \lim_{n\to\infty} \sum_\lambda \{w_\lambda|f_{n\lambda}^\star|\}_n = \lim_{n\to\infty} \|f_n^\star\|_{\mathbf{w},p}^p$$

▶ But we proved earlier that $\|f_n^\star\|_{\mathbf{w},p}^p \leq \max\{1,|\mu_n|\}.\frac{\epsilon_n^2}{\gamma_n} + \|f_o^\dagger\|_{\mathbf{w},p}^p$. Therefore, since the $\lim_{n\to\infty}\mu_n = \mu_o \in (0,1]$ and $\lim_{n\to\infty}\frac{\epsilon_n^2}{\gamma_n} = 0$, we get:

$$\|\widetilde{f}\|_{\mathbf{w},p}^p \leq \lim_{n\to\infty} \|f_n^\star\|_{\mathbf{w},p}^p \leq \|f_o^\dagger\|_{\mathbf{w},p}^p \tag{49}$$

◀

By definition, $f_o^\dagger$ is the unique minimizer of the $\|.\|_{\mathbf{w},p}$-norm in $\mathcal{S}_{T,f_o}$, so this implies that $\widetilde{f} = f_o^\dagger$.

The conclusion of this paragraph is that $f_o^\dagger$ is the only possible accumulation point of the sequence $f_n^\star$. Since we proved that the sequence $\{f_n^\star\}_n$ is uniformly bounded in the $\|.\|_{\mathcal{H}^i}$-norm and that it has a unique accumulation point: $f_o^\dagger$, this allows us to conclude that $f_n^\star$ converges weakly to $f_o^\dagger$.

c) The sequence $\{f_n^\star\}_n$ converges strongly to $f_o^\dagger$:

[This is identical to the proof given for Theorem 3.5] ∎

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, when f is the only antecedent of its own image by T.

5. Extension to inverse problem with several objects and observations

We now consider the more general problem where we seek M objects or components $f_1,..,f_M$ from L observations $g_1,...,g_L$. 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 $T_{m,l} : \mathcal{H}_m^i \to \mathcal{H}_l^o$

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

$$\forall l = 1..L, \quad g_l = \sum_{m=1}^{M} T_{m,l} f_m + n_l \tag{50}$$

where n_l are noise terms.

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

$$J(f_1, f_2, \ldots, f_M) = \sum_{l=1}^{L} \rho_l \left\| \mathrm{M}_{g_l}(\sum_{m=1}^{M} T_{m,l} f_m - g_l) \right\|_{\mathcal{H}_l^o}^2 + \sum_{m=1}^{M} \gamma_m \|f_m\|_{X_m} \; ; \tag{51}$$

where the γ_m and ρ_l are strictly positive scalars and the "norms" $\|\!|.\|\!|_{X_m}$ are $\|\!| \cdot \|\!|_{\mathbf{w},p}$–norm as before:

$$\|\!| f \|\!|_{X_m} = \sum_{\lambda\in\Lambda} w_\lambda^m | \langle f, \varphi_\lambda^m \rangle |^{p_m} \tag{52}$$

where for all m, $\boldsymbol{\varphi}^m = \{\varphi_\lambda^m\}_{\lambda\in\Lambda}$ is a basis of $\mathcal{H}_m^i$, $w_\lambda^m > 0$ and $1 \leq p_m \leq 2$ and for all l, M_{g_l} is an adaptive such as the ones we have seen in section 3 and 4.

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 $\|\!| \cdot \|\!|_{\mathbf{w},p}$–norm 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: $\sum_{j,k} | \langle f, \psi_{j,k} \rangle |$.

We mention here without giving the proofs that such cost functionals (with or without adaptive discrepancy terms) can be minimized by iterative algorithms similar to Algorithm 2.1, 3.2 and 4.2. To do so, one works on the vectorized operator $T = \{T_{m,l}\}_{m,l}$, observation $G = (g_1, g_2, \ldots, g_L)^T$, and object $F = (f_1, f_2, \ldots, f_M)^T$: such that the goal is to minimize:

$$J(F) = \left\| \mathrm{M}_G(TF - G) \right\|_{\mathcal{H}^o}^2 + \gamma \|\!| F \|\!| \ ; \tag{53}$$

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

It is true that the weighted norm induced on $\mathcal{H}^o$ makes it a standard Hilbert space, hence the discrepancy terms do match the ones seen before perfectly. But the regularization terms do not match: we get in Eq.(4) a simple weighted l_p sum (with a single exponent p), which is not true here for $M > 1$.

However, the minimization of Eq.(53) can be done by slightly modifying the iterative algorithm that we use for Eq.(4): it suffices to change the shrinkage operators $\mathbf{S}_{\mathbf{w},\mathrm{p}}$ so that they take into account the different weighted l_p norms for each of the coordinates of F. Moreover, the proofs of convergence and stability carry to this more complicated case (see [8] for details). We use these extended iterative algorithms and functionals in section 7.

6. Extension to complex coefficients and redundant transforms

The algorithms and theorems presented so far apply only to the case where the regularization systems $\boldsymbol{\varphi}^{\boldsymbol{m}} = \{\varphi^m{}_\lambda\}_{\lambda\in\Lambda}$ are orthonormal bases of $\mathcal{H}_1$ and the scalar products $\langle ., \varphi_\lambda^m \rangle$ are real. It will be useful in our application to use redundant and/or complex families instead. To do that, one needs to make two changes, as was pointed out in [4].

Firstly, the definition of the operators $\mathbf{S}_{\mathbf{w},\mathrm{p}}$ has to be extended to complex numbers. This is done by applying $\mathbf{S}_{\mathbf{w},\mathrm{p}}$ only to the modulus of a complex number, keeping the phase fixed:

$$\mathbf{S}_{\mathbf{w},\mathrm{p}}(r.e^{i\theta}) \overset{def}{=} \mathbf{S}_{\mathbf{w},\mathrm{p}}(r).e^{i\theta}, \ r \in \mathbb{R}, \ \theta \in [0, 2\pi]. \tag{54}$$

Secondly, if the family $\boldsymbol{\varphi} = \{\varphi_\lambda\}_{\lambda\in\Lambda}$ is redundant, the set of sequences of scalar products of elements of $\mathcal{H}_1$:

$$\mathcal{C} = \big\{ \{\langle f, \varphi_\lambda \rangle\}_{\lambda\in\Lambda}, \ f \in \mathcal{H}_1 \big\},$$

is a strict subset of the set of square summable sequences $l^2(\mathbb{R})$ (or $l^2(\mathbb{C})$). As a result,

$$f = \sum_\lambda S_{\gamma w_\lambda, p}\Big(\ \{a + T^*(g - Ta) \ \}_\lambda \ \Big) \ \varphi_\lambda \tag{55}$$

does not imply that:

$$\forall\lambda,\ \langle f,\varphi_\lambda\rangle = S_{\gamma w_\lambda,p}\Big(\ \{a+T^*(g-Ta)\ \}_\lambda\Big) \tag{56}$$

In the derivation of algorithms 2.1, 3.2 and 4.2, we used the fact that Eq. (55) and Eq. (56) are equivalent when $\boldsymbol{\varphi}=\{\varphi_\lambda\}_{\lambda\in\Lambda}$ is an orthonormal basis. When $\boldsymbol{\varphi}=\{\varphi_\lambda\}_{\lambda\in\Lambda}$ is redundant, this problem is rectified by projecting the sequence of coefficients obtained at each step of the iteration algorithm onto the set of scalar products $\mathcal{C}$:

$$f^n = P_{\mathcal{C}}\,\mathbf{S}_{\gamma\mathbf{w},\mathrm{p}}\left(f^{n-1}+T^*(g-Tf^{n-1})\right), \quad n\geq 1 \tag{57}$$

where $P_{\mathcal{C}}$ is the projection onto the set $\mathcal{C}$.

7. Application

7.1. 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,..,f_M that can be modeled by:

$$\forall l=1..L,\quad g(\nu_l)=g_l=b_l*\sum_{m=1}^{M}a_{m,l}f_m+n_l. \tag{58}$$

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.(50) 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: ν =145, 217 or 265GHz. (Details about the noise and blur level can be found in [8] , 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).

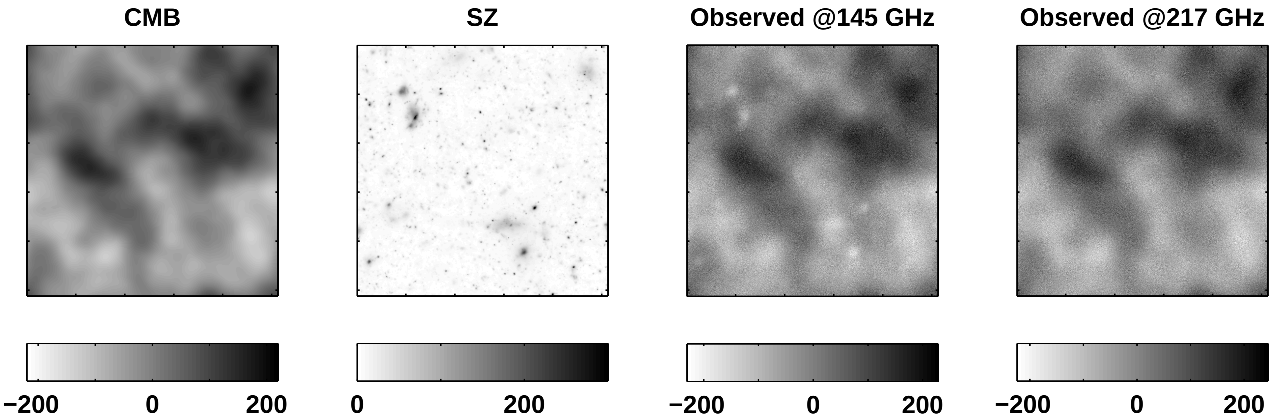

Figure 2. Multispectral data (units:μK); left to right: CMB map, galaxy clusters map, observation at 145GHz, observation at 217GHz

7.2. General parameters of the functional algorithms

In this multispectral case, the reconstruction methods proposed are extended as we have seen in section 5. The functionals we seek to minimize now contain one regularization term for each component and one regularization term per observation (see Eq.(51)).

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 [12; 13].

We compare the results obtained with the classical discrepancy terms (Eq.(51) with $\forall l$: $\mathrm{M}_{g_l} = \mathrm{Id}$, i.e Eq.(4) extended) to these obtained with various adaptive projections, relaxed (Eq.(37) extended) and not (Eq.(10) extended). 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.

Remark. The wavelet transform used for the regularization term of the SZ component is not a basis but a redundant system: the dual tree complex wavelet transform. The redundancy is 4:1 and the frame is tight (energy is conserved between original and transformed space). The choice of a redundant transform was made to remedy the lack of translation and rotation invariance of critically sample wavelet basis. Indeed, shrinkage on a complex dual tree transform yields much less artifacts since coefficients vary smoothly due to their complex envelope where standard wavelet coefficients oscillate much more.

7.3. 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 7.1 has been performed with the different iterative algorithms proposed in section 2, 3 and 4. All parameters were described in 7.1 except for the relaxed projection dampening parameter μ (see Eq.(36)) which is fixed here to $\mu = 0.1$ when using Algorithm 4.2.

Fig. 3 displays the results obtained for

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

The observed maps and the CMB and galaxy clusters maps that we seek to recover are in shown on the left panels of Fig. 2. The reconstructed CMB maps are in the two left panels of Fig. 3. The reconstructed galaxy maps are in the two right panels of Fig. 3.

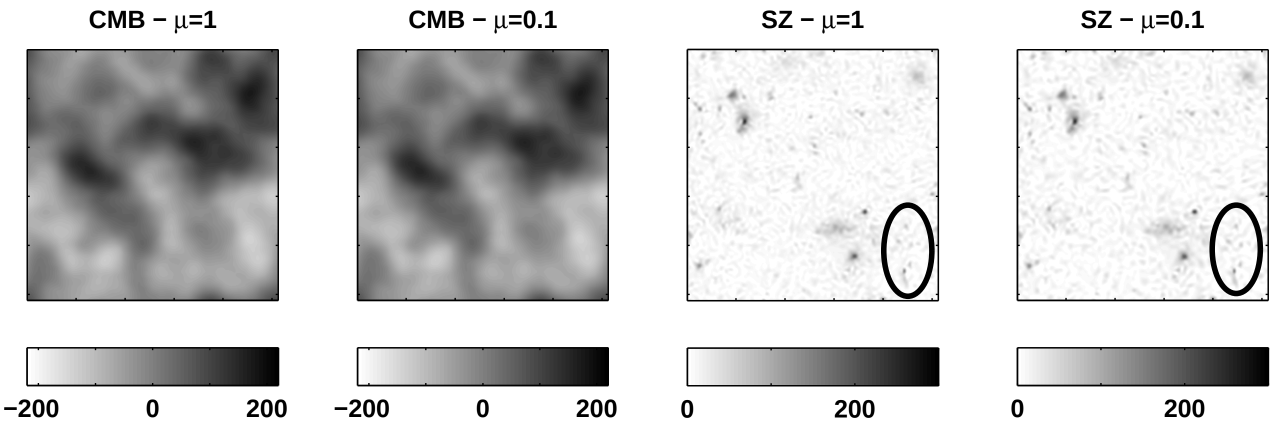

Figure 3. Reconstructed maps; without projections: first and third images ($\mu = 1$); with adaptive projections: second and fourth images ($\mu = 0.1$); far/middle left: CMB; far/middle right: galaxy clusters

The following analysis is illustrated by the results shown in 3 but is valid in a more general study with 24 similar simulations.

7.3.1. 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.

7.3.2. 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 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 [8], 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 be done for Algorithm 2.1 [8] and show that it actually gives good results in this prospective. 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. 3, 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. 3). This improvement is not statistically significant however it illustrates how adaptive discrepancy terms provide a novel way of tuning the algorithm to the data.

8. Conclusion

In this paper, we attacked the problem of inverting a known bounded linear operator by minimizing a cost functional balancing discrepancy term with a regularization term. The regularizing term is a weighted l_p norm in a basis chosen to fit desirable properties of the object sought. Such cost functional with classical quadratic discrepancy terms, along with iterative algorithm that lead to a minimizer, have been thoroughly investigated both from a theoretical and practical point-of-view in the literature. Our contribution in this paper, is to extend such method by modifying the discrepancy term to make it adaptive. Following the seminal idea in [2], we propose two modifications on the discrepancy term using adaptive projections or a relaxed version of these projections that focus on the important features of the observed data. Using the mathematical framework in [4], we prove that the corresponding functionals and iterative algorithms are strongly convergent and stable. We also show that using projection directly may lead to the loss of information in the data, and subsequently unexpected and undesired shape of the solution in the asymptotic case of perfect observations. The functionals using adaptive term with relaxed projections instead are shown not to suffer this drawback.

We have extended the functionals and algorithms to the case where several objects are sought from several observations, making sure that the obtained functional take in account the fact that different objects have different properties. The resulting algorithms (with classical discrepancy term, with projections or with relaxed projections) have been successfully applied to the problem of reconstructing maps of physical phenomena from multifrequency observations of the Cosmic Microwave Background in astronomy. These methods allow to reconstruct maps of the Cosmic Microwave Background and of the galaxy' clusters with enough precision to reliably detect clusters from them. The difference between the algorithms is not significant however small visual differences in the results are in agreement with the differences in the discrepancy terms used.

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 professor Ingrid Daubechies.

References

[1] Rudin L I, Osher S and Fatemi E 1992 *Physica D* **60** 259–268

[2] Starck J L, Murtagh F and Bijaoui A 1995 *Graphical Models and Image Processing* **57**(4) 20–431

[3] Figueiredo M and Nowak R D 1999 *Proc. SPIE Conference on Mathematical Modeling, Bayesian Estimation, and Inverse Problems* vol 3816 pp 97–108

[4] Daubechies I, Defrise M and De Mol C 2004 *Comm. Pure Appl. Math.* **57** 1413–1541 URL `http://www.citebase.org/abstract?id=oai:arXiv.org:math/0307152`

[5] Daubechies I and Teschke G 2005 *Applied and Computational Harmonic Analysis* **19**(1) 1–16

[6] Ramlau R and Teschke G 2005 *Inverse Problems* **21**(5) 1571–1592

[7] Elad M, Starck J L, Querre P and Donoho D 2005 *Applied and Computational Harmonic Analysis* **19** 340–358

[8] Anthoine S 2005 *Different Wavelet-based Approaches for the Separation of Noisy and Blurred Mixtures of Components. Application to Astrophysical Data* Ph.d. dissertation Princeton University URL `http://www.i3s.unice.fr/~anthoine/phd.pdf`

[9] Fornasier M and Rauhut H 2007 *to appear in Applied and Computational Harmonic Analysis*

[10] Vonesh C and Unser M 2007 *Proc. SPIE Optics and Optronics, Wavelet XII* vol 6701 (San Diego, USA)

[11] Daubechies I, Fornasier M and Loris I *to appear in J. Fourier Anal. Appl.* URL `arXiv:0706.4297`

[12] Kingsbury N G 2001 *Journal of Applied and Computational Harmonic Analysis* **10** 234–253

[13] Selesnick I W 2002 *IEEE Trans. Sig. Proc.* **50** 1144–1152