Regularization of multiplicative iterative algorithms with nonnegative constraint

Federico Benvenuto[1] **and Michele Piana**[1,2]

1. DIMA, Università di Genova, Via Dodecaneso 35, 16146, Genova, Italy
2. CNR - SPIN, Via Dodecaneso 33, 16146, Genova, Italy

Abstract. This paper studies the regularization of constrained Maximum Likelihood iterative algorithms applied to incompatible ill-posed linear inverse problems. Specifically we introduce a novel stopping rule which defines a regularization algorithm for the Iterative Space Reconstruction Algorithm in the case of Least-Squares minimization. Further we show that the same rule regularizes the Expectation Maximization algorithm in the case of Kullback-Leibler minimization provided a well-justified modification of the definition of Tikhonov regularization is introduced. The performances of this stopping rule are illustrated in the case of an image reconstruction problem in X-ray solar astronomy.

1. Introduction

In the theory of linear ill-posed inverse problems, discrepancy is a measure of the effectiveness of an element in the solution space to reproduce a given element in the data space through the linear operator modeling the problem. Examples of discrepancy functions that are systematically utilized in this setting are the Least-Squares (LS) discrepancy [23] and the Kullback-Leibler (KL) divergence [19]. The optimization problem associated to the linear inverse problem is the one to minimize the discrepancy function over the entire solution space or, more properly, over a subset of such space. If this subset is the nonnegative orthant, the application of the Karush-Kuhn-Tucker (KKT) conditions [8] transforms the constrained optimization problem into a fixed point problem which can be naturally solved by means of a successive approximation scheme. In the case of LS discrepancy this iterative scheme is the Iterative Space Reconstruction Algorithm (ISRA) [9] while in the case of the KL discrepancy is Expectation-Maximization (EM) [26]. Although both schemes have the convergence property, their limit solutions are not acceptable from a physical viewpoint, since the intrinsic ill-posedness of the original inverse problem induces noise amplification. There are two ways to regularize this numerical instability: first, with an addition of information on the solution realized by adding a penalty term to the discrepancy; second, with the application of a stopping rule on the iterative approximation process preventing the algorithm to reach the limit solution [12].

This paper focuses on this second approach to regularization, with specific attention to the case of incompatible problems formulated in a finite dimension setting. Incompatibility indicates that the noise-free version of the data does not belong to the range of the operator and in this case the inverse problem is also called genuinely ill-posed [28]. We note that a standard approach to stop iterative optimization algorithms is to check the discrepancy value while the iterations run and stop the scheme when a given threshold value is achieved [22, 6]. Our first results show that, for incompatible ill-posed inverse problems, this approach is not regularizing for ISRA and EM. On the other hand, we introduce a new stopping rule, named the KKT principle since it is inspired by the KKT conditions, that is regularizing for ISRA even in the incompatible case. However, we also observe that this same stopping rule fails to regularize EM. Therefore, as the last result of our paper, we introduce a new definition of regularization algorithm that naturally extends the classical Tikhonov definition [28, 12] to the asymptotic case of data with large norm and show that, with such definition, the KKT principle is regularizing for EM.

The plan of the paper is as follows. In Section 2 we setup the optimization problems and introduce the iterative schemes for their solution. Section 3 discusses some existing stopping rules in the case of incompatible problems and introduce a new stopping rule based on the KKT conditions, showing its effectiveness for the regularization of the LS problem. Section 4 provides a new definition of regularization algorithm and shows that, under this definition, the KKT principle regularizes the solution of incompatible problems in the case of KL optimization. Section 5 shows how the KKT stopping rule works when applied to an image reconstruction problem in hard X-ray solar astronomy. Our conclusions will be offered in Section 6.

As a final remark, we observe that most of the arguments treated in this paper has a straightforward statistical interpretation. For example, the optimization of the LS and KL discrepancies can be interpreted as Maximum Likelihood (ML) problems for Gaussian and Poisson statistics, respectively; ISRA and EM provides constrained solutions for these maximum-likelihood problems; and also the new KKT stopping rule can be interpreted in a statistical way. Therefore, although all results in the paper are obtained in a deterministic optimization framework, when appropriate we will point out the statistical aspects of the methods introduced.

2. Optimization problems and iterative schemes

Let us consider two normed vector spaces $\mathcal{X}$ and $\mathcal{Y}$ of dimension M and N, respectively. Let H be an $N \times M$ matrix representing a linear application $\mathcal{X} \to \mathcal{Y}$ whose elements are nonnegative, i.e. for all $j = 1, \ldots, M$ and $i = 1, \ldots, N$

$$H_{ij} \geq 0 \tag{1}$$

and no rows are identically zero, i.e. for all $i = 1, \ldots, N$

$$\sum_{j=1}^{M} H_{ij} > 0 \ . \tag{2}$$

Let $x = (x_1, \ldots, x_M)$ be the vector belonging to the nonnegative convex and closed cone $\mathcal{C}^{\mathcal{X}} = \{x \in \mathcal{X} \mid x_j \geq 0 \ , \ j = 1, \ldots, M\}$ and let Hx the corresponding vector belonging to $H(\mathcal{C}^{\mathcal{X}}) \subset \mathcal{C}^{\mathcal{Y}}$, where $\mathcal{C}^{\mathcal{Y}} = \{y \in \mathcal{Y} \mid y_i \geq 0 \ , \ i = 1, \ldots, N\}$. We consider x as a set of unknown parameters to be estimated, knowing an element $y = (y_1, \ldots, y_N) \in \mathcal{Y}$ which represents the detected signal. The estimation problem is to find x such that

$$y = Hx \quad , \quad x \in \mathcal{C}^{\mathcal{X}} \ . \tag{3}$$

In particular we are interested in the case when the exact solution of equation (3) is nonexistent and the problem is called incompatible, i.e. $y \notin H(\mathcal{C}^{\mathcal{X}})$ [28]. In this case, an estimation of x can be found using the following procedure. Given some discrepancy

$$L : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+ \cup \{0\} \tag{4}$$

which assigns to each pair (x, y) a nonnegative number, we can consider the problem of finding the elements x^* at which the discrepancy takes its minimum within $\mathcal{C}^{\mathcal{X}}$, i.e. such that

$$L(x^*, y) = \min_{x \in \mathcal{C}^{\mathcal{X}}} L(x, y) =: \mu_L(y) > 0 \ , \tag{5}$$

with data y. At least one minimizer $x^* \in \mathcal{C}^{\mathcal{X}}$ exists if L is convex and coercive on $\mathcal{C}^{\mathcal{X}}$. The map $\mu_L : \mathcal{Y} \to \mathbb{R}^+$ is called the incompatibility measure of the problem (3) according to the discrepancy L.

In this work we will consider two discrepancies: the LS function

$$D_{LS}(x,y) = \|Hx - y\|_2^2 \quad , \tag{6}$$

and the KL divergence

$$D_{KL}(x,y) = \sum_{i=1}^{N} y_i \log \frac{y_i}{(Hx)_i} + (Hx)_i - y_i \ , \tag{7}$$

where multiplication and division between vectors are done elements by elements. Common algorithms for the constrained minimization (5) of discrepancies (6) and (7) are gradient-type algorithms. In this work we will focus on multiplicative algorithms that can be readily derived from the KKT conditions of the corresponding minimization problems [8]. In the case of nonnegative constraint, i.e. $x \in \mathcal{C}^{\mathcal{X}}$, the KKT conditions take the form

$$x \ \nabla_x L(x,y) = 0 \quad , \quad x \in \mathcal{C}^{\mathcal{X}} \tag{8}$$

where multiplication and equality between vectors are done element-by-element. In both cases one can transform condition (8) into a fixed point equation and then, applying a successive approximation scheme, one gets a multiplicative iterative algorithm. When L is the LS function (6), equation (8) leads to

$$x^{(k+1)} = x^{(k)} \frac{H^T y}{H^T H x^{(k)}} \tag{9}$$

which, initialized with a constant vector $x^{(0)} = \xi \in \mathcal{C}^{\mathcal{X}}$ provides positive estimates of the solution when $H^T y$ has positive elements and is known as ISRA. It has been introduced as an acceleration of EM [9] and it is convergent to the constrained minimum of D_{LS} [10]. On the other hand, when L is the KL divergence (7), equation (8) leads to

$$x^{(k+1)} = \frac{x^{(k)}}{H^T 1_{\mathcal{Y}}} H^T \frac{y}{Hx^{(k)}} \tag{10}$$

where $1_{\mathcal{Y}}$ indicates the constant vector $(1, \ldots, 1) \in \mathcal{Y}$. The algorithm (10) is usually initialized with a constant vector $x^{(0)} = \xi \in \mathcal{C}^{\mathcal{X}}$, it is known as EM [11] or also as Richardson Lucy algorithm [25, 21] when H represents a convolution operator and it is convergent to the constrained minimum of D_{KL} [26] when $y \in \mathcal{C}^{\mathcal{Y}}$.

3. Stopping rules for incompatible problems

For any fixed $y \in \mathcal{Y}$ let us consider the iterative algorithms (9) and (10). Let us indicate with $\phi_y^{(k)} : \mathcal{X} \to \mathcal{X}$, where $k \in \mathbb{N}$, the map representing their iterative step, i.e.

$$x^{(k+1)} = \phi_y^{(k+1)}(x^{(k)}) \ . \tag{11}$$

For every $\xi \in \mathcal{C}^{\mathcal{X}}$, let us consider the family of operators $\psi_\xi : \mathbb{N} \times \mathcal{Y} \to \mathcal{C}^{\mathcal{X}}$ defined by

$$\psi_\xi^{(k)}(y) = \phi_y^{(k)} \circ \ldots \circ \phi_y^{(1)}(\xi) \tag{12}$$

which determines a map $\mathcal{Y} \to \mathcal{C}^{\mathcal{X}}$ for every k. The convergence property is valid for both algorithms [10, 26] and prescribes that, given a datum $y \in \Omega$ the convergence domain and an initializing vector $\xi \in \mathcal{C}^{\mathcal{X}}$, there exists a unique element $\hat{x} \in \mathcal{C}^{\mathcal{X}}$ such that

$$\lim_{k\to\infty} \|\psi_{\xi}^{(k)}(y) - \hat{x}\|_{\mathcal{X}} = 0 \; . \tag{13}$$

For any fixed initialization $\xi \in \mathcal{C}^{\mathcal{X}}$ we can define the inverse operator

$$\Psi_{\xi} : \Omega \subset \mathcal{Y} \to \mathcal{C}^{\mathcal{X}} \tag{14}$$

such that $\Psi_{\xi}(y) = \hat{x}$. In [2] authors observe that a nonnegative LS solution $\hat{x}$ consists of a set of few significantly nonnegative elements over a major part of zero values. In [4] authors show this property with an example in a deconvolution framework, and note that the non zero components of $\hat{x}$ strongly depend on the noise realization corrupting the data. In general, although $\hat{x}$ is nonnegative, it is not acceptable from a physical viewpoint, since it is corrupted by noise amplification. Therefore some regularization of the operator Ψ_{ξ} is required. It is well known that both algorithms (9) and (10) have the semi-convergence property [5, 12], which means that until a certain number of iterations the algorithms approach the solution of (5) and after that the reconstructions deteriorate. As a result, these methods can be regularized by applying some stopping rule. Now, to investigate the regularization properties of these algorithms, we utilize the definition given in [27] for the regularization of inverse operators and apply it to the case of iterative algorithms in finite dimensional normed spaces. In particular this definition is given for a countable family of regularizing operators and it is specifically referred to the operators Ψ_{ξ} defined in equation (14).

Definition 3.1. *The operator $\Psi_{\xi} : \Omega \subset \mathcal{Y} \to \mathcal{X}$ between two finite dimensional normed vector spaces $\mathcal{Y}$ and $\mathcal{X}$ is called regularizable on $\Omega \subset \mathcal{Y}$ if there exists a family of operator*

$$R_k : \Omega \to \mathcal{X} \tag{15}$$

with $k \in \mathbb{N}$ and a parameter choice rule

$$k : \mathbb{R}^{+} \times \Omega \to \mathbb{N} \tag{16}$$

such that, for all $y \in \Omega$,

$$\lim_{\delta\to 0^{+}} \sup\{\|R_{k(\delta,y^{\delta})}(y^{\delta}) - \Psi_{\xi}(y)\| \mid y^{\delta} \in \Omega \; , \; \|y^{\delta} - y\| \leq \delta\} = 0 \tag{17}$$

holds. For a specific $y \in \Omega$, a pair $(\{R_k\}, k)$ is called a **regularization method** *for Ψ_{ξ} if (17) holds.*

Clearly, the family of operators $\{\psi_{\xi}^{(k)}\}_{k\in\mathbb{N}}$ defined by equation (12), when provided with a rule $k(\delta, y^{\delta})$ such that

$$\lim_{\delta\to 0^{+}} \inf\{k(\delta, y^{\delta}) \mid y^{\delta} \in \Omega \; , \; \|y^{\delta} - y\| \leq \delta\} = \infty \tag{18}$$

holds, satisfies condition (17). Since condition (18) can be readily fulfilled using, for example, the parameter choice rule $k(\delta, y^{\delta}) = \lfloor 1/\delta \rfloor$, where the symbol $\lfloor \cdot \rfloor$ indicates the integer part function, such a pair $(\{\psi_{\xi}^{(k)}\}_{k\in\mathbb{N}}, \lfloor 1/\delta \rfloor)$ defines a regularization method. In general, a family of regularizing operators can be constructed starting from a convergent iterative algorithm by using a stopping rule satisfying condition (18).

We are interested in parameter choice rules for iterative algorithms related with the specific problem and hence depending on the data y^δ. To this aim we now introduce a general definition of stopping rule for iterative algorithms.

Definition 3.2. *Fixed $\tau > 0$, let $f_\tau : \mathbb{N} \times \mathbb{R}^+ \times \Omega \to \mathbb{R}$ a function. We call* **stopping rule** *of an iterative algorithm the parameter choice rule $k : \mathbb{R}^+ \times \Omega \to \mathbb{N}$ defined by*

$$k(\delta, y^\delta) := \inf\{k \in \mathbb{N} \mid f_\tau(k, \delta, y^\delta) \leq 0\} \tag{19}$$

for any pair (δ, y^δ).

This definition implies that for any fixed $\tau > 0$ a stopping rule relates the number of iterations k with the pair (δ, y^δ). Further, different stopping rules are obtained by selecting different choices for the function f_τ. The stopping rules we will consider in this paper are based on continuous f_τ but this restriction is not mandatory for Definition 3.2.

To stop the ISRA iterations, we now introduce the Morozov discrepancy parameter choice rule.

Definition 3.3. *Let D_{LS} be the LS discrepancy (6) and $\{\psi_\xi^{(k)}\}$ the family of operators associated with ISRA. We call Morozov discrepancy stopping rule the function k_M of the form (19) with*

$$f_\tau(k, \delta, y^\delta) := D_{LS}(\psi_\xi^{(k)}(y^\delta), y^\delta) - \tau N \delta^2 \quad . \tag{20}$$

Now we prove that ISRA provided with the Morozov discrepancy principle is not a regularizing algorithm if the problem (3) is incompatible.

Theorem 3.1. *Let us consider the problem (3). Given $y \in \Omega$, let D_{LS} be the discrepancy (6) and $\mu_{LS}(y) > 0$ the corresponding incompatibility measure. For a given $\xi \in \mathcal{C}^{\mathcal{X}}$, let $\Psi_\xi(y)$ the limit solution of ISRA and $\{\psi_\xi^{(k)}\}_{k\in\mathbb{N}}$ the family of operators associated with ISRA. For any fixed $\tau > 0$, if $\delta < \sqrt{\mu_{LS}/\tau N}$ then the Morozov stopping rule k_M is not defined in a neighborhood of $(0, y)$.*

Proof. Given an $\epsilon > 0$, one can always find a δ such that $\mu_m := \inf\{\mu_{LS}(y^\delta) \mid \|y^\delta - y\| < \delta\} > \mu_{LS}(y) - \epsilon$, since the function μ_{LS} is continuous. Hence

$$\mu_{LS}(y) - \epsilon < \mu_m \leq D_{LS}(\psi_\xi^{(k)}(y^\delta), y^\delta) \leq \tau N \delta^2 \ . \tag{21}$$

By taking δ such that $\tau N \delta^2 < \mu_{LS}(y)$, equation (21) is never satisfied and the function k_M is not defined in a neighborhood of $(0, y)$ as ϵ can be arbitrarily chosen. □

In the case of EM, some authors [24] proposed parameter choice rules depending on the data error size and provided theoretical justifications for regularization. Recently, approaches independent from the data error size have been studied in [6, 7, 17, 18]. In [6] authors proposed and motivated with statistical arguments the Poisson discrepancy parameter choice rule in order to estimate the regularization parameter both for optimization problems regularized with a penalty term and for stopping the unpenalized algorithm. We recall this rule in our notation.

Definition 3.4. *Let D_{KL} be the KL divergence (7) and $\{\psi_\xi^{(k)}\}$ the family of operators associated with EM. We call the Poisson discrepancy stopping rule the function k_P of the form (19) with*

$$f_\tau(k, \delta, y^\delta) := D_{KL}(\psi_\xi^{(k)}(y^\delta), y^\delta) - \tau \frac{N}{2} \tag{22}$$

Clearly, the stopping rule (22) does not depend on δ. However, with a little abuse of notation, we used the symbol $f_\tau = f_\tau(k, \delta, y^\delta)$ to indicate it. According to the assertion at page 3 in [27] and to the meaning of Theroem 3.3 in [12], we now prove that the rule (22) is not regularizing in general (i.e. not for any τ). Moreover, we prove that this rule becomes regularizing if and only if an appropriate additional condition on τ is given (see [12] page 157).

Theorem 3.2. *Let us consider the problem (3). Given $y \in \Omega$, let D_{KL} be the discrepancy (7) and $\mu_{KL}(y) > 0$ the incompatibility measure. For a given $\xi \in \mathcal{C}^{\mathcal{X}}$, let $\Psi_\xi(y)$ the limit solution of EM, $\{\psi_\xi^{(k)}\}_{k\in\mathbb{N}}$ the family of operators associated with EM and k_P the Poisson stopping rule. The pair $(\{\psi_\xi^{(k)}\}, k_P)$ defines a regularization method if and only if $\tau = 2\mu_{KL}(y)/N$.*

Proof. Since EM is convergent, condition (17) is fulfilled if (18) holds true. We prove that statement (18) of Definition (3.1) when $k = k_P$ as defined by equations (22) and (19), holds true if and only if $\tau = 2\mu_{KL}(y)/N$.

Given an $\epsilon > 0$, one can always find a δ such that $\mu_m := \inf\{\mu_{KL}(y^\delta) \mid \|y^\delta - y\| < \delta\} > \mu_{KL}(y) - \epsilon$, since the function μ_{KL} is continuous. Hence the set in r.h.s. of k_P definition (19) is described by

$$\mu_{KL}(y) - \epsilon < \mu_m \leq D_{KL}(\psi_\xi^{(k)}(y^\delta), y^\delta) \leq \tau N/2\ . \tag{23}$$

By taking $\tau < 2(\mu_{KL}(y) - \epsilon)/N$ condition 23 is never satisfied and, as ϵ can be arbitrarily chosen, the function k_P is not defined in a neighborhood of $(0, y)$ when $\tau < 2\mu_{KL}(y)/N$.

Now we prove that the function k_P is defined in a neighborhood of $(0, y)$ when $\tau \geq 2\mu_{KL}(y)/N$. Being $\mu_{KL}(cy) = c\mu_{KL}(y)$ for any $c \in \mathbb{R}^+$, for any $\delta > 0$ there exists some y^δ with $\|y^\delta - y\| < \delta$ such that $\mu_{KL}(y^\delta) < \mu_{KL}(y)$. For such a y^δ, when $\tau \geq 2\mu_{KL}(y)/N$, as the EM iteration sequence $\psi_\xi^{(k)}(y^\delta)$ converges to some $\hat{x}^\delta$ and $D_{KL}(\hat{x}^\delta, y^\delta) = \mu_{KL}(y^\delta)$, we have that $D_{KL}(\psi_\xi^{(k)}(y^\delta), y^\delta) < \mu_{KL}(y) \leq \tau N/2$ when k is large enough. This implies that the function k_P is defined at least on the non-empty set $\Delta = \{(\delta, y^\delta) \mid \|y^\delta - y\| < \delta\ ,\ \mu_{KL}(y^\delta) < \mu_{KL}(y)\}$.

Moreover, when $\tau > 2\mu_{KL}(y)/N$, there exists a δ such that $\forall(\delta, y^\delta) \in \Delta$ the equation

$$D_{KL}(\psi_\xi^{(k)}(y^\delta), y^\delta) \leq \tau N/2 \tag{24}$$

is satisfied for some finite value k since the l.h.s. is positive, limited and asymptotically decreasing to $\mu_{KL}(y^\delta) < \tau N/2$ for any y^δ as k tends to infinity. Hence the limit (18) is finite.

Finally, when $\tau = 2\mu_{KL}(y)/N$ the stopping condition becomes $D_{KL}(\psi_\xi^{(k)}(y^\delta), y^\delta) \leq \mu_{KL}(y)$ and, being EM convergent, the limit (18) is ∞. $\square$

From Theorem 3.2 it follows that for incompatible problem (3) it is not ensured that the Poisson discrepancy stopping rule defines a regularization method for the operator Ψ_ξ associated with EM. In fact, for a given data y and matrix H, the fulfillment of the regularizing property depends on the value of the incompatibility measure $\mu_{KL}(y)$, where the value y is unknown.

Now we introduce a stopping rule for ISRA and we will prove that it always applies even to the case of incompatible LS problems.

Definition 3.5. *Let us consider the following functions* $L_{LS}(x, y) = \| x\ H^T\,(Hx - y)\,\|_2^2$ *and* $E_{LS}(x, \delta) = \delta^2 \sum_{j=1}^M x_j^2\ (H_2^T 1_{\mathcal{Y}})_j$*, where* $(H_2)_{ij} = (H_{ij})^2$*. The LS-KKT stopping rule is defined by taking*

$$f_\tau(k, \delta, y^\delta) := L_{LS}(\psi_\xi^{(k)}(y^\delta), y^\delta) - \tau E_{LS}(\psi_\xi^{(k)}(y^\delta), \delta) \tag{25}$$

in equation (19).

Now we prove that ISRA provided with such a stopping rule is a regularization operator for $\Psi_\xi(y)$ even for incompatible problems.

Theorem 3.3. *Let us consider the problem (3) and let* D_{LS} *be the discrepancy (6). For any* $y \in \Omega$ *and for a given* $\xi \in \mathcal{C}^{\mathcal{X}}$*, let* $\Psi_\xi(y)$ *the limit solution of ISRA and* $\{\psi_\xi^{(k)}\}_{k\in\mathbb{N}}$ *the family of operators associated with ISRA. If* k_{LS} *is the LS-KKT stopping rule, then the pair* $(\{\psi_\xi^{(k)}\}_{k\in\mathbb{N}}, k_{LS})$ *is a regularization method for* $\Psi_\xi(y)$ *for any* $y \in \Omega \subset \mathcal{Y}$*.*

Proof. Since ISRA is convergent, condition (17) is fulfilled if (18) holds true. We have to prove statement (18) of Definition (3.1) when $k = k_{LS}$, as defined in equations (25) and (19). Since $\psi_\xi^{(k)}(y^\delta)$ converges to some $\hat{x}^\delta$ and the set $\{x \in \mathcal{C}^{\mathcal{X}} \mid L_{LS}(x, y^\delta) \leq \tau E_{LS}(x, \delta)\}$ is a neighborhood of $\hat{x}^\delta$, the set $\{k \in \mathbb{N} \mid L_{LS}(\psi_\xi^{(k)}(y^\delta), y^\delta) \leq \tau E_{LS}(\psi_\xi^{(k)}(y^\delta), \delta)\}$ is not empty.

Now, we get a lower bound of the set in equation (18) weakening the condition

$$\frac{L_{LS}(\psi_\xi^{(k)}(y^\delta), y^\delta)}{E_{LS}^*(\psi_\xi^{(k)}(y^\delta))} \leq \tau\delta^2 \tag{26}$$

where $E_{LS}^*(\psi_\xi^{(k)}(y^\delta)) = E_{LS}(\psi_\xi^{(k)}(y^\delta), \delta)/\delta^2$ is positive and bounded for all y^δ such that $\|y^\delta - y\| \leq \delta$. By taking the infimum

$$s(k) := \inf_{\|y^\delta - y\| \leq \delta} \frac{L_{LS}(\psi_\xi^{(k)}(y^\delta), y^\delta)}{E_{LS}^*(\psi_\xi^{(k)}(y^\delta))} \quad , \tag{27}$$

we get a map $s : \mathbb{N} \to \mathbb{R}$ defined for every $k \in \mathbb{N}$, which is positive, bounded and asymptotically decreasing to 0, since the algorithm is convergent for every $y \in \Omega$. Hence, we obtain the following lower bound

$$\inf\{k_{LS}(\delta, y^\delta) \mid \forall\ y^\delta,\ \|y^\delta - y\| \leq \delta\} \geq \inf\{k \in \mathbb{N} \mid s(k) \leq \tau\delta^2\} \tag{28}$$

As δ tends to 0, k has to arbitrarily increase to fulfill condition $s(k) \leq \tau\delta^2$ and therefore the thesis is proved. $\square$

Remark 3.1. *The LS-KKT stopping rule can be explained in a statistical framework when the data* Y *is a Gaussian random vector with mean* Hx *and i.i.d. components each one with variance* σ^2*, and the minimizing function is the LS discrepancy. The function* L_{LS} *in Definition 3.5 is the square of the 2-norm of the l.h.s. of KKT conditions (8), and* E_{LS} *is its expected value. Indeed, it is easy to show that* $E_{LS}(x,\sigma) = \mathbb{E}_Y(L_{LS}(x,y))$*. Therefore, the parameter choice rule (25) stops ISRA iterations as soon as* L_{LS} *becomes equal to its expected value* E_{LS}*, up to a scaling parameter* $\tau > 0$*, i.e.*

$$L_{LS}(x,y) \leq \tau E_{LS}\,(x,\sigma) \quad . \tag{29}$$

f_τ *in equation (25) is given by replacing* (x,y) *with* $(\psi_\xi^{(k)}(y^\delta), y^\delta)$ *and the standard deviation* σ *with the error level* δ *in equation (29).*

We observe now that applying the statistical arguments in Remark 3.1 to the case of a Poisson data vector Y with mean Hx and using the KL divergence as a discrepancy function leads to the KL-KKT stopping rule for EM algorithm defined in the following:

Definition 3.6. *Let us consider the following functions* $L_{KL}(x,y) = \|x\left(H^T\left(1 - \frac{y}{Hx}\right)\right)\|_2^2$ *and* $E_{KL}(x) = \sum_{j=1}^M x_j^2\ \left(H_2^T \frac{1}{Hx}\right)_j$*. The KL-KKT stopping rule is defined by taking*

$$f_\tau(k,\delta,y^\delta) := L_{KL}(\psi_\xi^{(k)}(y^\delta), y^\delta) - \tau E_{KL}(\psi_\xi^{(k)}(y^\delta)) \tag{30}$$

in equation (19).

However we now prove that EM with the proposed stopping rule does not define a regularization algorithm in the sense of definition (17).

Theorem 3.4. *Let us consider the problem (3) and let* D_{KL} *be the discrepancy (7). For any* $y \in \Omega$ *and for a given* $\xi \in \mathcal{C}^{\mathcal{X}}$*, let* $\Psi_\xi(y)$ *the limit solution of EM and* $\{\psi_\xi^{(k)}\}_{k\in\mathbb{N}}$ *the family of operators associated with EM. If* k_{KL} *is the KL-KKT stopping rule, then the pair* $(\{\psi_\xi^{(k)}\}_{k\in\mathbb{N}}, k_{KL})$ *is not a regularization method for* $\Psi_\xi(y)$ *for any* $y \in \Omega \subset \mathcal{Y}$*.*

Proof. Since EM is convergent, condition (17) is fulfilled if (18) holds true. We prove that the statement (18) of definition (3.1) when $k = k_{KL}$, as defined in equations (25) and (19), does not hold true. Since $\psi_\xi^{(k)}(y^\delta)$ converges to some $\hat{x}^\delta$ and the set $\{x \in \mathcal{C}^{\mathcal{X}} \mid L_{KL}(x,y^\delta) \leq \tau E_{KL}(x)\}$ is a neighborhood of $\hat{x}^\delta$, the set $\{k \in \mathbb{N} \mid L_{KL}(\psi_\xi^{(k)}(y^\delta), y^\delta) \leq \tau E_{KL}(\psi_\xi^{(k)}(y^\delta))\}$ is not empty. Now, we consider the condition

$$\frac{L_{KL}(\psi_\xi^{(k)}(y^\delta), y^\delta)}{E_{KL}(\psi_\xi^{(k)}(y^\delta))} \leq \tau \tag{31}$$

where $E_{LS}(\psi_\xi^{(k)}(y^\delta))$ is positive and bounded for all y^δ such that $\|y^\delta - y\| \leq \delta$. For any y^δ the l.h.s. is positive, bounded and asymptotically decreasing to 0 since the algorithm is convergent. It means that, fixed $\tau > 0$, exists a value k^* such that

$$\frac{L_{KL}(\psi_\xi^{(k^*)}(y^\delta), y^\delta)}{E_{LS}(\psi_\xi^{(k^*)}(y^\delta))} < \tau \quad \forall\ y^\delta\ ,\ \|y^\delta - y\| \leq \delta\ , \tag{32}$$

and hence the limit (18) is finite. □

The reason why the KL-KKT criterion is not regularizing for EM is due to the fact that such criterion does not depend on the error level δ^2. The regularization property of the KL-KKT stopping rule is re-established by introducing a new definition of regularization algorithm involving the asymptotic behavior of operators R_k with respect to the norm of y.

4. Asymptotic regularization

We remark that the definition of regularization is referred to a specific $y \in \Omega \subset \mathcal{Y}$ and therefore it concerns pointwise convergence of the family $\{R_k\}_{k\in\mathbb{N}}$ to the operator Ψ_ϵ. In the following definition we introduce a new concept of regularization which concerns the convergence of the family $\{R_k\}_{k\in\mathbb{N}}$ when $\|y\| \to +\infty$ and we refer to it as asymptotic regularization.

Definition 4.1. *The operator $\Psi_\xi : \Omega \subset \mathcal{Y} \to \mathcal{X}$ between finite dimensional normed vector spaces $\mathcal{Y}$ and $\mathcal{X}$ is called asymptotically regularizable on a cone $\mathcal{C} \subset \Omega$ if there exists a family of operator*

$$R_k : \mathcal{C} \to \mathcal{X} \tag{33}$$

with $k \in \mathbb{N}$ and a parameter choice rule

$$k : \mathbb{R}^+ \times \mathcal{C} \to \mathbb{N} \tag{34}$$

which satisfies the following condition: $\forall\epsilon > 0$ and $\forall\delta > 0$ exists an M_0 such that

$$y \in \mathcal{C}, y^\delta \in \Omega,\ \|y\| > M_0,\ \|y^\delta - y\| \le \delta \ \Rightarrow\ \|R_{k(\delta,y^\delta)}y^\delta - \Psi_\xi(y)\| < \epsilon \tag{35}$$

holds. A pair $(\{R_k\}, k)$ is called an **asymptotic regularization method** *for Ψ_ξ on $\mathcal{C}$ if (35) holds.*

We note that any family of operators defined by equation (12) and associated to some convergent algorithm satisfies condition (35) if it is provided with a rule $k(\delta, y^\delta)$ satisfying the following condition: $\forall N_0 > 0$ and $\forall\delta > 0$ exists an M_0 such that

$$y \in \mathcal{C}, y^\delta \in \Omega,\ \|y\| > M_0,\ \|y^\delta - y\| \le \delta \ \Rightarrow\ k(\delta, y^\delta) > N_0\ . \tag{36}$$

This condition means that, whatever the distance between the realization of the data y^δ and their exact version y, as the data norm increases, the number of iterations needed to stop the algorithm has to increase. The underlying idea is that a Poisson variable does not admit a noise-free realization unless asymptotically, i.e. when the signal-to-noise ratio tends to infinity and this fact can be mimicked in an analysis framework, by letting the norm of y tend to infinity.

Now, we will prove that EM with the Poisson discrepancy stopping rules does not define an asymptotic regularization method for Ψ_ξ when applied to the incompatible problem (3), while it is if $y \in H(\mathcal{C}^{\mathcal{X}})$. To this aim we need the following general EM property.

Lemma 4.1. EM scaling property *Given $y \in \mathcal{Y}$ and $\xi \in \mathcal{C}^{\mathcal{X}}$, let us consider a scalar $S > 0$. Let $\psi_\xi^{(k)}(y)$ indicate the k-th EM iteration with entry data y as in equation (10). The following relation holds true for every $k \in \mathbb{N}$*

$$\psi_\xi^{(k)}(Sy) = S\psi_\xi^{(k)}(y) \quad . \tag{37}$$

Proof. At the first iteration, with entry data Sy, we have

$$\psi_\xi^{(1)}(Sy) = \frac{\xi}{H^T 1_{\mathcal{Y}}} H^T \frac{Sy}{H\xi} = S\psi_\xi^{(1)}(y) \quad . \tag{38}$$

For induction, by supposing that (37) holds true for k, we have

$$\psi_\xi^{(k+1)}(Sy) = \frac{\psi_\xi^{(k)}(Sy)}{H^T 1_{\mathcal{Y}}} H^T \frac{Sy}{H\psi_\xi^{(k)}(Sy)} = S\psi_\xi^{(k+1)}(y) \quad . \tag{39}$$

and hence the thesis holds true. □

In other words we can say that EM produces scaled reconstructions for scaled input data. Incidentally, we notice that the same property holds also for ISRA. Now we can prove the following:

Theorem 4.1. *Let us consider the problem (3) with $y \in \Omega \subset \mathcal{Y}$, let D_{KL} be the discrepancy (7) and $\mu_{KL}(y) > 0$ the incompatibility measure. For a given $\xi \in \mathcal{C}^{\mathcal{X}}$, let $\Psi_\xi(y)$ the limit solution of EM, $\{\psi_\xi^{(k)}\}_{k\in\mathbb{N}}$ the family of operators associated with EM and k_P the Poisson stopping rule. The pair $(\{\psi_\xi^{(k)}\}, k_P)$ defines an asymptotic regularization method on the cone $H(\mathcal{C}^{\mathcal{X}})$ and it does not outside it.*

Proof. Let $\delta > 0$, y and y^δ be such that $\|y - y^\delta\| < \delta$ with $\|y\| = S > \delta$ and $\|y^\delta\| = S_\delta > \delta$. Let $\bar{y} = y/S$ ans $\bar{y}^\delta = y^\delta/S_\delta$. From the definition of Poisson discrepancy stopping rule (22) and the EM scaling property (Lemma 4.1) we have that k_P is the function assigning to each pair (δ, y^δ) the smaller k such that

$$D_{KL}(\psi_\xi^{(k)}(\bar{y}^\delta), \bar{y}^\delta) \leq \frac{\tau N}{2(S-\delta)} \quad . \tag{40}$$

Moreover, for any ϵ there exists an S such that

$$D_{KL}(\psi_\xi^{(k)}(\bar{y}^\delta), \bar{y}^\delta) \geq \mu_{KL}(\bar{y}^\delta) > \mu_{KL}(\bar{y}) - \epsilon \ , \tag{41}$$

since straightforward computation yields $\|\bar{y} - \bar{y}^\delta\| < 2\delta/S$ and the function μ_{KL} is continuous. When $\mu_{KL}(\bar{y}) > 0$ the conditions (40) and (41) are incompatible as S becomes large enough and hence $k(\delta, y^\delta)$ is not defined. When $\mu_{KL}(\bar{y}) = 0$ the inequalities are compatible and hence, being EM convergent, the regularization property holds when $y \in H(\mathcal{C}^{\mathcal{X}})$. □

The thesis of Theorem 4.1 holds true also using stopping rules different from the Poisson discrepancy. For example the weighted LS discrepancy, a first order approximation of the KL discrepancy, gives rise to a stopping rule [1] for which Theorem 4.1 holds.

Now we will prove that the KL-KKT stopping rule coupled with EM is an asymptotic regularization method. We need the following lemmas.

Lemma 4.2. *For any $y^\delta \in \mathcal{Y}$, let $\psi_\xi^{(k)}(y^\delta)$ be the k-th iteration of EM. For each iteration k the relation*

$$\sum_{i=1}^{N}(H\psi_\xi^{(k)}(y^\delta))_i = \sum_{i=1}^{N} y_i^\delta \tag{42}$$

holds.

Proof. The thesis follows directly from computations. □

Lemma 4.3. *For any $y^\delta \in \mathcal{Y}$, let $\psi_\xi^{(k)}(y^\delta)$ be the k-th iteration of EM. The following inequality holds:*

$$E_{KL}(\psi_\xi^{(k)}(y^\delta)) \leq \sum_{i=1}^{N} y_i^\delta \tag{43}$$

Proof. We begin by noting that

$$\sum_{j=1}^{M}(\psi_\xi^{(k)}(y^\delta))_j^2 \left(H_2^T \frac{1}{H\psi_\xi^{(k)}(y^\delta)}\right)_j = \sum_{i=1}^{N}(H_2(\psi_\xi^{(k)}(y^\delta))^2)_i \left(\frac{1}{H\psi_\xi^{(k)}(y^\delta)}\right)_i . \tag{44}$$

Then

$$\begin{aligned}\sum_{i=1}^{N}\frac{\sum_{j=1}^{M} h_{ij}^2(\psi_\xi^{(k)}(y^\delta))_j^2}{\sum_{j=1}^{M} h_{ij}(\psi_\xi^{(k)}(y^\delta))_j} &\leq \sum_{i=1}^{N}\sqrt{\sum_{j=1}^{M} h_{ij}^2(\psi_\xi^{(k)}(y^\delta))_j^2} \\ &\leq \sum_{i=1}^{N}\sum_{j=1}^{M} h_{ij}(\psi_\xi^{(k)}(y^\delta))_j \leq \sum_{i=1}^{N} y_i^\delta \end{aligned}\tag{45}$$

having used the relation $\sqrt{\sum_{j=1}^{M} h_{ij}^2(\psi_\xi^{(k)}(y^\delta))_j^2} \leq \sum_{j=1}^{M} h_{ij}(\psi_\xi^{(k)}(y^\delta))_j$ and Lemma 4.2. □

Now we can prove the following:

Theorem 4.2. *Let us consider the problem (3) and let D_{KL} be the discrepancy (7). For any $y \in \Omega \subset \mathcal{Y}$ and for a given $\xi \in \mathcal{C}^{\mathcal{X}}$, let $\Psi_\xi(y)$ the limit solution of EM algorithm and $\{\psi_\xi^{(k)}\}_{k\in\mathbb{N}}$ the family of operators associated with EM. If k_{KL} is the KL-KKT stopping rule, then the pair $(\{\psi_\xi^{(k)}\}_{k\in\mathbb{N}}, k_{KL})$ is an asymptotic regularization method for Ψ_ξ.*

Proof. The EM algorithm is convergent, hence condition (35) is fulfilled if (36) holds true. Now we check condition (36). Let $\delta > 0$ and $y \in \Omega$, we have

$$\begin{aligned}&\inf\{k_{KL}(\delta, y^\delta) \mid \forall\ y^\delta,\ \|y^\delta - y\| \leq \delta\} = \\ &\inf\left\{k \in \mathbb{N} \;\middle|\; \frac{L_{KL}(\psi_\xi^{(k)}(y^\delta), y^\delta)}{E_{KL}(\psi_\xi^{(k)}(y^\delta))} \leq \tau,\ \forall\ y^\delta,\ \|y^\delta - y\| \leq \delta\right\},\end{aligned}\tag{46}$$

since $E_{KL}(\psi_\xi^{(k)}(y^\delta))$ is positive. Since $\psi_\xi^{(k)}(y^\delta)$ converges to some $\hat{x}^\delta$ and the set $\{x \in \mathcal{C}^{\mathcal{X}} \mid L_{KL}(x, y^\delta) \leq \tau E_{KL}(x)\}$ is a neighborhood of $\hat{x}^\delta$, the set $\{k \in \mathbb{N} \mid L_{KL}(\psi_\xi^{(k)}(y^\delta), y^\delta) \leq \tau E_{KL}(\psi_\xi^{(k)}(y^\delta))\}$ is not empty. For the EM scaling property (Lemma 4.1)

$$\frac{L_{KL}(\psi_\xi^{(k)}(y^\delta), y^\delta)}{E_{KL}(\psi_\xi^{(k)}(y^\delta))} \leq \tau \quad \Leftrightarrow \quad \frac{L_{KL}(\psi_\xi^{(k)}(\bar{y}^\delta), \bar{y}^\delta)}{E_{KL}(\psi_\xi^{(k)}(\bar{y}^\delta))} \leq \frac{\tau}{S_\delta} \tag{47}$$

where $y^\delta = S_\delta \bar{y}^\delta$ and $\|\bar{y}^\delta\| = 1$. Now, we get a lower bound of the set (46) weakening the condition (47). First, we can consider that, for a given $S > \delta$ and $\|y\| > S$ we have $S_\delta > S - \delta$ for every y^δ such that $\|y - y^\delta\| < \delta$. Hence equation (47) implies

$$\frac{L_{KL}(\psi_\xi^{(k)}(\bar{y}^\delta), \bar{y}^\delta)}{E_{KL}(\psi_\xi^{(k)}(\bar{y}^\delta))} \leq \frac{\tau}{S - \delta} \tag{48}$$

Second, we can take the infimum over $\bar{y}^\delta \in \Sigma^\delta \subset \mathcal{S}^N$ the N-dimensional sphere, i.e.

$$s(k) := \inf_{\bar{y}^\delta \in \mathcal{S}^N} \frac{L_{KL}(\psi_\xi^{(k)}(\bar{y}^\delta), \bar{y}^\delta)}{E_{KL}(\psi_\xi^{(k)}(\bar{y}^\delta))} \quad . \tag{49}$$

The map $s : \mathbb{N} \to \mathbb{R}$ is defined for every $k \in \mathbb{N}$, positive, limited and does not approach to zero for $k \in \mathbb{N}$ since $E_{KL}(\psi_\xi^{(k)}(\bar{y}^\delta)) \leq \sqrt{N}$ is limited (see Lemma 4.3). Moreover, it is asymptotically decreasing to 0, i.e.

$$\lim_{k \to \infty} L_{KL}(\psi_\xi^{(k)}(y), y) = 0 \tag{50}$$

for every $y \in \Omega$. Hence, we proved that

$$\inf\{k_{KL}(\delta, y^\delta) \mid \forall\ y^\delta,\ \|y^\delta - y\| \leq \delta\} \geq \inf\{k \in \mathbb{N} \mid s(k) \leq \frac{\tau}{S - \delta}\}. \tag{51}$$

We can choose S such that k has to arbitrarily increase to fulfill the condition in the r.h.s. and therefore the thesis is proved. $\square$

In Table (1) we present a summary of the results. By comparing the second and the last column, it is clear that asymptotic regularization plays the role of the standard one when EM is used to minimize the KL discrepancy.

LS - ISRA		KL - EM		
Morozov	regularization	Poisson	regularization	asymptotic regularization
$y \in H(\mathcal{C}^{\mathcal{X}})$	yes	$y \in H(\mathcal{C}^{\mathcal{X}})$	no* (Th. 3.2)	yes (Th. 4.1)
$y \notin H(\mathcal{C}^{\mathcal{X}})$	no (Th. 3.1)	$y \notin H(\mathcal{C}^{\mathcal{X}})$	no* (Th. 3.2)	no (Th. 4.1)
KKT		KKT		
$y \in H(\mathcal{C}^{\mathcal{X}})$	yes (Th. 3.3)	$y \in H(\mathcal{C}^{\mathcal{X}})$	no (Th. 3.4)	yes (Th. 4.2)
$y \notin H(\mathcal{C}^{\mathcal{X}})$	yes (Th. 3.3)	$y \notin H(\mathcal{C}^{\mathcal{X}})$	no (Th. 3.4)	yes (Th. 4.2)

Table 1. Summary of (asymptotic) regularization properties of the ISRA and EM algorithms provided with Morozov discrepancy, Poisson discrepancy and KKT stopping rules. The asterisk (*) means except $\tau = \frac{2}{N}\mu(y)$.

5. Numerical experiments

In this section we test the proposed regularization algorithm in the case of image reconstruction from data recorded by a solar hard X-ray satellite. The Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI) [20] mission has been launched by NASA in Febraury 2002 with the aim of investigating emission and energy transport mechanisms during solar flares. RHESSI hardware is made of nine pairs of rotating collimators that time-modulate the incoming photon flux before it is detected by the corresponding Ge detectors. As a consequence, the RHESSI imaging problem consists in locally retrieving the photon flux intensity image starting from a given set of count modulation profiles.

This image reconstruction problem is clearly incompatible, since the model is not exact and the positivity constraint makes impossible the fitting of the data. The count profile formation system is described by a non linear model [14] which is not used to perform inversions owing to its complexity and therefore the imaging problem is usually solved using a linearized model. This linear operator changes along some status parameters of the satellite and it can be retrived using the routines of Solar SoftWare (`http://www.lmsal.com/solarsoft/`) as well as count modulation profiles. The software needed to perform the following analysis can be found at `http://www.dima.unige.it/~benvenuto/reg.tar.gz`.

We studied the behaviour of EM regularized by the KKT principle for the reconstruction of the photon flux map of two real flaring events. The first event is the September 8 2002 flare in the time interval between 01:38:44 and 01:39:35 UT. The data have been collected by detectors 3 through 8, in the energy range between 25 and 30 keV. The second event is the November 3 2003 flare in the time interval between 01:32:42 and 01:42:25 UT. The data have been collected by detectors 3 through 8, in the energy range between 12 and 25 keV. During the first event the total number of counts collected is about 7.45 10^4, the number of data is $N = 3816$. During the second event the total number of counts collected is about 1.38 10^6 and the number of data is $N = 3168$. In both cases the reconstructed field of view is a square of 64 arcseconds side length corresponding to a 64 by 64 pixel image. The noise on the data is mainly Poisson. An estimate of the signal-to-noise ratio can be done by averaging the ratios of the standard deviation over the mean for each datum, i.e.

$$SNR(y) = \frac{1}{N}\sum_{i=1}^{N} \frac{y_i}{\sqrt{y_i}} \ . \tag{52}$$

The estimated signal-to-noise ratio is about 3.95 for the 2002, September 08 event and 19.35 for 2003 November 03.

We applied EM to both data sets, using KL-KKT with $\tau = 1$ in order to stop the iterations. The reconstruction in the left panel of Figure 1 shows, in the case of the September event, one coronal source and two footpoints, which is in accordance with

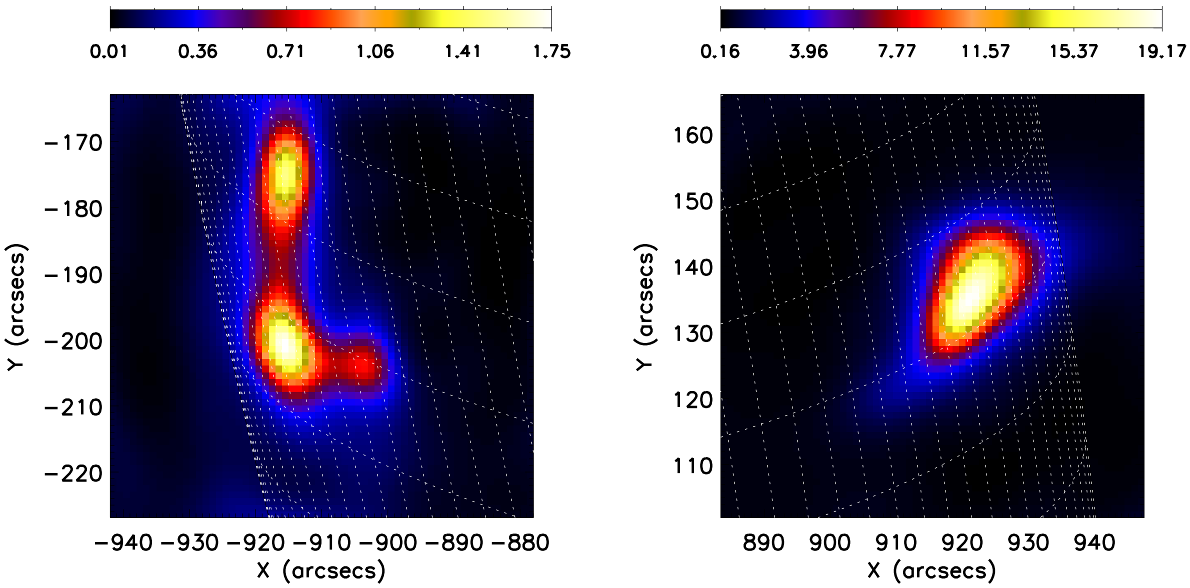

Figure 1. EM reconstructions of the September 8 2002 (left) and the November 3 2003 (right) events stopped with the KL-KKT principle for $\tau = 1$.

typical non-thermal models in solar flare physics [15]. In the case of the November event (Figure 1, right panel) the stopping rule provides a reconstruction comparable with the ones provided by CLEAN [13] of around the same energy and time interval [16]. For both events we also studied the behavior of the KL-KKT principle for smaller values of τ (see Figure 2) and found that the stopping rule always applies and provides higher values of the optimal iteration number. We notice that the performances of this stopping rule are quite similar for the two events, despite the fact that the method is applied to flares characterized by a notably different count amounts. The robustness of the KKT principle is confirmed in Figure 3 and in Figure 4, containing the reconstruction of the two events for four values of τ (in particular, for the November event, using smaller value of τ is able to point out hints of the two footpoints whose presence is confirmed in CLEAN images obtained at higher energy ranges [16]). The behavior of standard stopping rules in the case of this application is significantly less robust. For example, in the case of the November event, the Poisson discrepancy principle never applies for $\tau < 29$ and for variations of τ in the range $[25.5, 25.7]$ it provides optimal iteration numbers that differ of up to 10^3 iterations. We finally notice that the accuracy of this approach to the RHESSI image reconstruction problem is demonstrated also in [3] in the case of realistically simulated synthetic count modulation profiles.

6. Conclusion

In this paper we studied a novel stopping rule for constained ML algorithms that accounts for the constraint in the solution and has a rather straightforward statistical

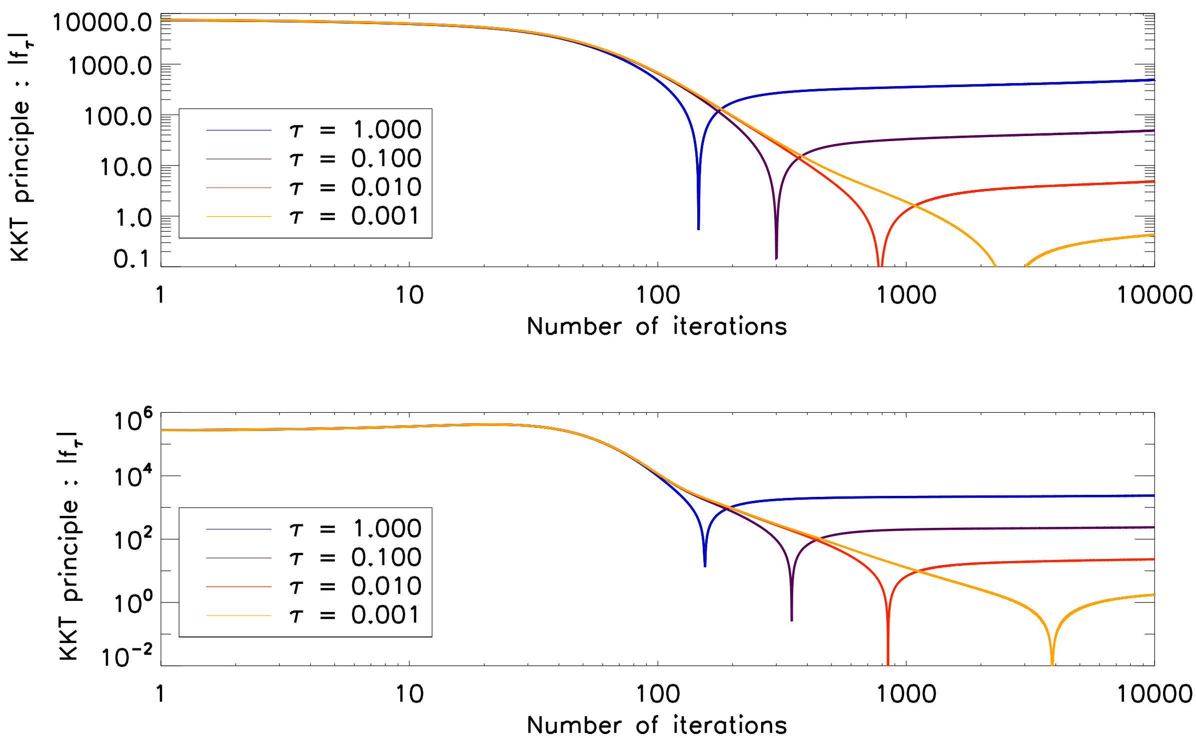

Figure 2. Performance of the KL-KKT principle for September 8 2002 (left) and November 3 2003 (right) events as a function of the number of iterations. On the y-axis the absolute value of f_τ showed in logarithmic scale. The cusps indicate that f_τ passes from positive to negative values and so at that iteration the stopping rule applies. The legend points out the different values of τ for the different curves.

interpretation. We proved that this stopping rule makes ISRA a regularization algorithm and provided a generalization of the Tikhonov definition of regularization for which the stopping rule is regularizing also for EM. We then showed the robustness of this criterion in two applications involving high energy solar data. We used count modulation profiles recorded by the X-ray solar mission RHESSI and we reconstructed the X-ray sources with the new method, obtaining reliable flaring configurations. A comparison with the reference method in these field has been given. Possible developments of this piece of research are: the extension of these new stopping rule to the implementation of the accelerated versions of this image reconstruction approach; the study of its reliability in the case of ML algorithms with penalty terms; and the study of stopping rules explicitly depending on the noise level for EM applied to incompatible problems.

Acknowledgements

This work was supported by the European Community Framework Program 7, “High Energy Solar Physics data in Europe (HESPE)”, Grant Agreement No. 263086.

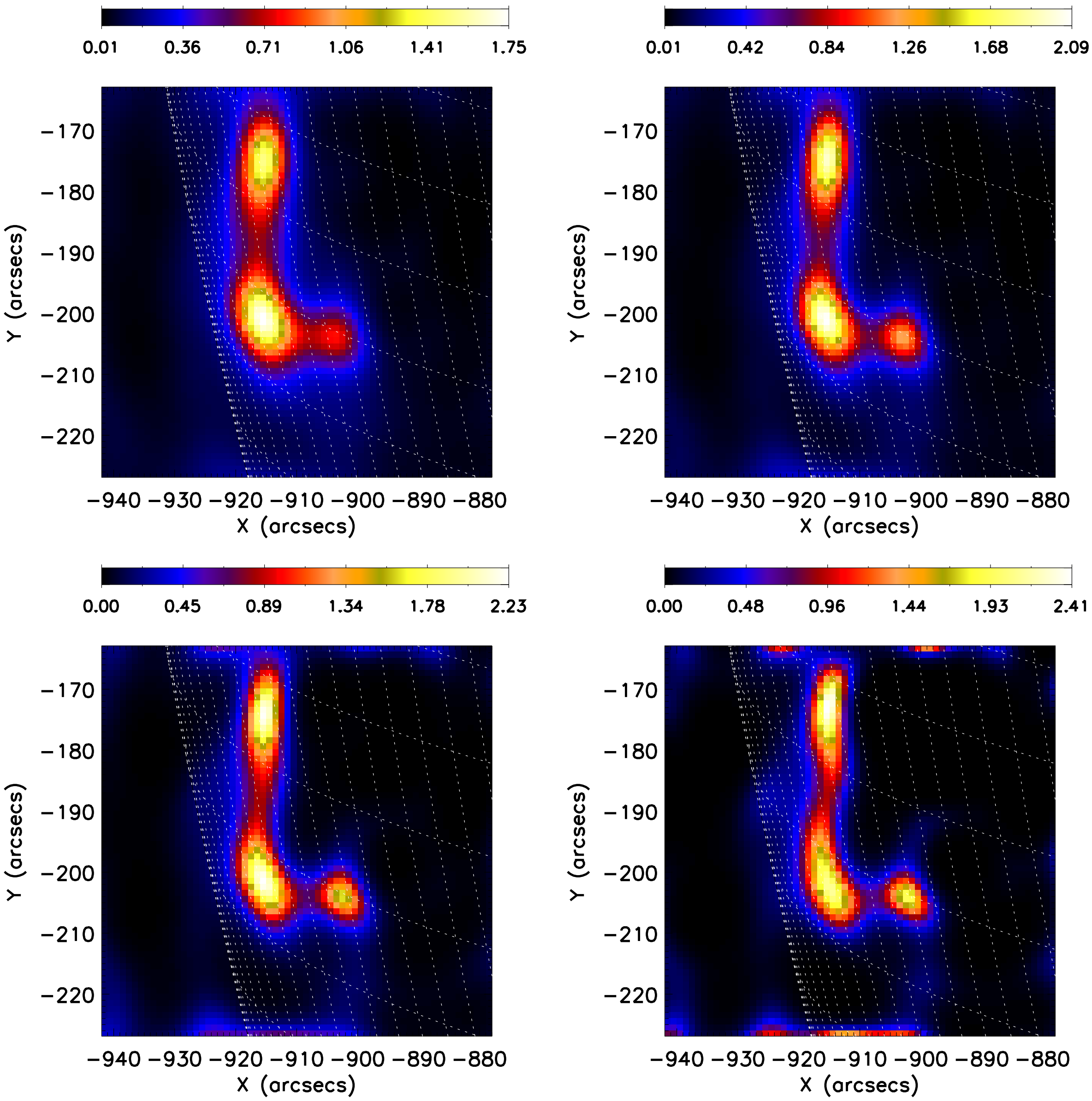

Figure 3. Reconstructions of the September 8 2002 event with the KL-KKT principle for different values of τ. $\tau = 1$ for the upper left, $\tau = 0.1$ for the upper right, $\tau = 0.01$ for the lower left, $\tau = 0.001$ for the lower right.

References

[1] Bardsley J M, 2006 Stopping rules for a nonnegatively constrained iterative method for ill-posed Poisson imaging problems. *BIT Numerical Mathematics*

[2] Barrett H H and Myers K J, 2003 *Foundations of Image Science* (New York: Wiley)

[3] Benvenuto F, Schwartz R, Piana M and Massone A M, 2013 Expectation maximization for hard X-ray count modulation profiles. *Astronomy and Astrophysics* 555 A61

[4] Benvenuto F, Zanella R, Zanni L and Bertero M, 2010 Nonnegative least-squares image deblurring: improved gradient projection approaches. *Inverse Problems* 26 (2)

[5] Bertero M and Boccacci P, 1998 *Introduction to Inverse Problems in Imaging* (Bristol: Institute of Physics Publishing)

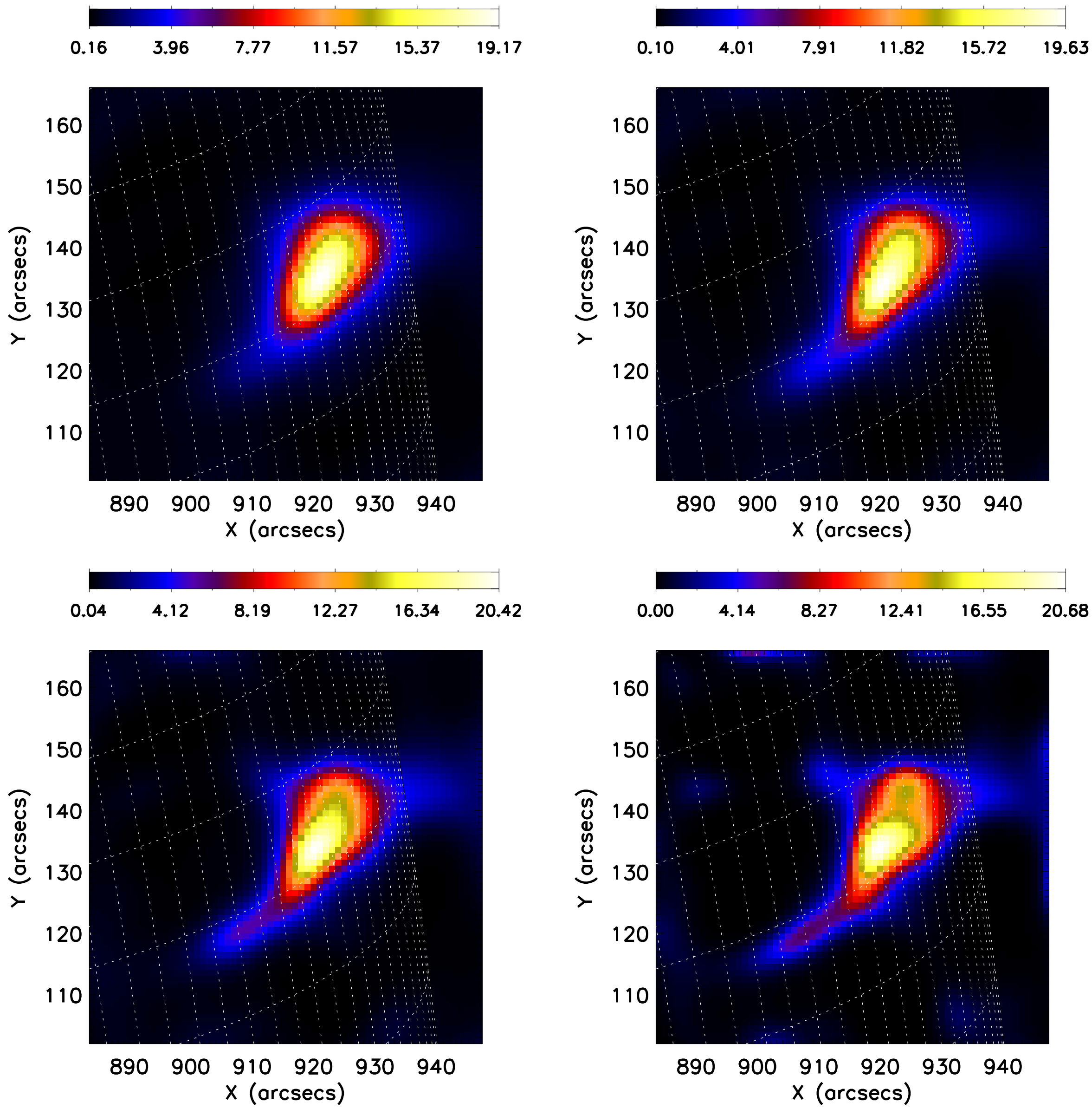

Figure 4. Reconstructions of the November 3 2003 event with the KL-KKT principle for different values of τ. $\tau = 1$ for the upper left, $\tau = 0.1$ for the upper right, $\tau = 0.01$ for the lower left, $\tau = 0.001$ for the lower right.

[6] Bertero M, Boccacci P, Talenti G, Zanella R and Zanni L, 2010 A discrepancy principle for Poisson data. *Inverse Problems* 26 (10)

[7] Bissantz N, Mair B A and Munk A, 2006 A multi-scale stopping criterion for MLEM reconstructions in PET 6 pp. 3376–3379

[8] Boyd S and Vandenberghe L, 2004 *Convex Optimization* (Cambridge: University Press)

[9] Daube Witherspoon M E and Muehllehner G, 1986 An iterative image space reconstruction algorithm suitable for volume ECT. *IEEE Trans. Med. Imaging* 5 pp. 61–66

[10] De Pierro A R, 1987 On the convergence of the iterative image space reconstruction algorithm for volume ECT. *IEEE Trans. Med. Imaging* 6 pp. 124–125

[11] Dempster A P, Laird N M and Rubin D B, 1977 Maximum likelihood from incomplete data via

the EM algorithm. *Journal of the Royal Statistical Society, Series B* 39 (1) pp. 1–38

[12] Engl H W, Hanke M and Neubauer A, 1996 *Regularization of inverse problems* (Dordrecht: Kluwer Academic Publishers)

[13] Högbom J A, 1974 Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. *Astronomy and Astrophysics Supplement* 15 p. 417

[14] Hurford G J, Schmahl E J, Schwartz R A, Conway A J, Aschw M J, Csillaghy A, Dennis B R, Johns-krull C and Smith M, 2002 The RHESSI imaging concept. *Sol. Phys* pp. 61–86

[15] Ishikawa S N, Krucker S, Takahashi T and Lin R P, 2011 On the Relation of Above-the-loop-top and Footpoint Hard X-ray Sources in Solar Flares. In *AAS/Solar Physics Division Abstracts #42*. p. 1104

[16] Joshi B, Manoharan P K, Veronig A M, Pant P and Pandey K, 2006 Multi-wavelength Analysis of an X2.7 Flare on 3 November 2003 from Active Region NOAA 10488. *Sun and Geosphere* 1 (2) pp. 020000–20

[17] Kindermann S, 2013 Discretization independent convergence rates for noise level-free parameter choice rules for the regularization of ill-conditioned problems. *Electronic Transactions on Numerical Analysis* 40 pp. 58–81

[18] Kindermann S and Neubauer A, 2008 On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization. *Inverse Problems and Imaging* 2 pp. 291–299

[19] Kullback S, 1959 *Information Theory and Statistics* (New York: Wiley)

[20] Lin R P, Dennis B R and Hurford G J, 2002 The Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI). *Solar Physics* 210 pp. 3–32

[21] Lucy L B, 1974 An iterative technique for the rectification of observed distributions. *Astronomical Journal* 79 pp. 745–754

[22] Morozov V A, 1984 *Methods for solving incorrectly posed problems* (New York: Springer)

[23] Rao C R and Toutenburg H, 1999 *Linear Models and Generalizations: Least Squares and Alternatives* (Springer Series in Statistics)

[24] Resmerita E, Engl H W and Iusem A N, 2007 The expectation-maximization algorithm for ill-posed integral equations: a convergence analysis. *Inverse Problems* 23 (6) p. 2575

[25] Richardson W H, 1972 Bayesian-Based Iterative Method of Image Restoration. *J. Opt. Soc. Am.* 62 (1) pp. 55–59

[26] Shepp L A and Vardi Y, 1982 Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging* 1 pp. 113–122

[27] Tikhonov A N, 1995 *Numerical methods for the solution of ill-posed problems*, vol. 328 (Springer)

[28] Tikhonov A N and Arsenin V Y, 1977 *Solution of Ill-posed Problems* (Washington: Winston and Sons)