Regularization of constrained maximum likelihood iterative algorithms by means of statistical stopping rule

Federico Benvenuto and Michele Piana
DIMA, Università di Genova, Via Dodecaneso 35, 16100, Italy
E-mail: benvenuto@dima.unige.it

Abstract. In this paper we propose a new statistical stopping rule for constrained maximum likelihood iterative algorithms applied to ill-posed inverse problems. To this aim we extend the definition of Tikhonov regularization in a statistical framework and prove that the application of the proposed stopping rule to the Iterative Space Reconstruction Algorithm (ISRA) in the Gaussian case and Expectation Maximization (EM) in the Poisson case leads to well defined regularization methods according to the given definition. We also prove that, if an inverse problem is genuinely ill-posed in the sense of Tikhonov, the same definition is not satisfied when ISRA and EM are optimized by classical stopping rule like Morozov’s discrepancy principle, Pearson’s test and Poisson discrepancy principle. The stopping rule is illustrated in the case of image reconstruction from data recorded by the Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI). First, by using a simulated image consisting of structures analogous to those of a real solar flare we validate the fidelity and accuracy with which the proposed stopping rule recovers the input image. Second, the robustness of the method is compared with the other classical stopping rules and its advantages are shown in the case of real data recorded by RHESSI during two different flaring events.
1. Introduction

Maximum likelihood (ML) is a standard approach to parameter estimation in statistics, and provides estimates of the solution even for statistical inverse problem. Once the statistical distribution of data is known, its application is straightforward and is widely used in many different fields [23], from astronomical image reconstruction to medical imaging where the data distribution is often Poisson; from geophysical to elastic waveforms inverse problem, where the data are mainly modeled by Gaussian noise.

ML estimates can be effectively constrained to be non negative. In the case of Poisson noise, the constrained ML approach leads to the Expectation Maximization (EM) algorithm [11]. In the case of white Gaussian noise, the constrained ML approach leads to the Iterative Space Reconstruction Algorithm (ISRA) [8]. Although both these algorithms have the convergence property, the properties of the limit solution have not yet been fully investigated. What is known is that such solution, in the case of experimental data, although non negative, is not acceptable from a physical viewpoint, since the intrinsic ill-posedness of the inverse problem induces noise amplification. There are two ways to regularize a statistical inverse problem: first, with an add of information on the solution realized by a prior probability in a Bayesian framework, or, second, without adding information on the solution but simply stopping the iterative approximation process before getting the limit solution [17]. This paper focuses on this second approach.

Among the many different stopping criteria available in literature, Morozov’s discrepancy principle [16], Pearson’s cumulative test [26, 12] and Generalized Cross validation [7, 18] can be applied to rather general schemes, while the updating coefficients method [25] and the more recent Poisson discrepancy criterion [5] have been specifically designed for EM. Some of these criteria may not work when the inverse problem is genuinely ill-posed in a Tikhonov sense [24], i.e. when the data do not belong to the range of the forward operator. The object of this paper is to introduce an optimal stopping rule for constrained iterative ML methods which is characterized by two properties. First, it is motivated by statistical arguments; second it works under completely general conditions (i.e. including the case of genuinely ill-posed problems). We apply this criterion to EM and ISRA and we show its effectiveness in the case of both synthetic (but realistically simulated) and real data. Finally we prove that this stopping rule makes these two iterative algorithms well defined regularization methods, according to a regularization definition which extends the classical Tikhonov definition to a statistical framework.

In Section 2 we describe the constrained ML approach to a statistical inverse problem, showing how to derive a general iterative algorithm given the likelihood, and we complete this overview applying the method to the Gaussian and Poisson cases. In Section 3 we introduce a definition of regularization for constrained ML problems extending the Tikhonov one to a statistical framework and we also show that classical stopping rules satisfy this definition only when the problem is not genuinely ill-posed. In
Section 4 we propose the new statistical criterion and we prove that it is a well-defined regularization method both in the Gaussian and the Poisson case. In Section 5 we perform some numerical applications with Gaussian and Poisson noise respectively, and we also show an application to a real case reconstructing solar X-ray images starting from count data collected by an on-orbit satellite.

2. Constrained ML problem

We denote by $x = \{x_j\}_{j=1,...,M}$ the unknown parameters, where $j$ is, in general, a multi-index. Moreover, we denote by $y = \{y_i\}_{i=1,...,N}$ the detected signal, where $i$ can be in general again a multi-index. $N$ and $M$ are, respectively, the number of data and unknown parameters of the reconstruction problem. Finally, let the relation between unknown parameters and the data be described by a linear system. We denote by $H$ the matrix describing the transformation from the parameter space to the signal space. We suppose each element of $H$ is positive, i.e.

$$H_{ij} > 0 \quad \forall \; i = 1, \ldots, N, \; j = 1, \ldots, M .$$

Then, the forward model can be written in the form

$$y = Hx .$$

A standard statistical approach for estimating the parameter $x$ given $y$, is the Maximum Likelihood (ML) method. It is based on the assumption that the data vector $y$ is an observed value of random vector $Y$ with mean $Hx$. In other words, denoting by $\eta$ a generic random process, we can write

$$Y_i \sim \eta_i(Hx) .$$

The density function of the vector $Y$ is given by the joint probability density function $p_\eta(y, Hx)$. When this density is thought of as a function of $x$ given $y$ we call it the likelihood and we write

$$L_y(x) = p_\eta(y, Hx) .$$

Once the matrix $H$ and the data $y$ are given, ML obtains the solution of $x$ when the likelihood reaches its maximum value. However, in the majority of cases physical motivations regarding the nature of the problem imply that the components of parameter $x$ has to be non-negative. Consequently, the ML estimator is constrained as

$$\hat{x} = \arg \max_{x \in C} \mathcal{L}_y(x) ,$$

where $C = \{x \mid x_j \geq 0 , \; \forall \; j = 1, \ldots, M\}$ is the non-negative orthant.

Usually, it is more convenient to minimize the negative logarithm of the likelihood instead of maximizing it. The constrained ML problem is therefore equivalent to

$$\hat{x} = \arg \min_{x \in C} L_y(x) ,$$

where $L_y(x) = -\log(\mathcal{L}_y(x))$. Indeed, when both the negative logarithm of the likelihood function and the constraint are convex, the necessary and sufficient conditions for $x$ to
be the constrained ML estimator are the Karush-Khun-Tucker (KKT) conditions [6], which in this particular case take the form
\[ x \nabla L_y(x) = 0 \text{ , with } x \geq 0 \] (7)
where multiplication and inequality between vectors are done element-by-element.

Splitting the gradient into the positive part \( V(x) \) and the negative part \( U(x) \) transforms (7) into a fixed point equation [13]; then applying the successive approximation method leads to the multiplicative iterative algorithm
\[ x^{(k+1)} = x^{(k)} \frac{U(x^{(k)})}{V(x^{(k)})} \] (8)

Even if there is no general proof of convergence for these algorithms, in the particular convex cases we will discuss this technique leads to well-known algorithms for which proof of convergence has been done. It has been conjectured [2] that the minimizer of the constrained ML problem should be sparse in a pixel space, but the problem of giving a parametric form of this solution is still open. Moreover, in [3] authors have shown with a numerical experiment that this conjecture is verified when the data are perturbed by Gaussian noise.

When \( \eta \) is a vector of independent and identically distributed Gaussian variables with mean given by \( Hx \) and variance equal to \( \sigma^2 \), the constrained ML problem (5) is equivalent to the minimization of the Least Squares function
\[ D_{LS}(y, x) = \| Hx - y \|^2 \] (9)
under the non-negativity constraint [4]. Consequently, KKT conditions (7) lead to the algorithm
\[ x^{(k+1)} = x^{(k)} \frac{H^T y}{H^T H x^{(k)}} \] (10)
which is known as Iterative Space Reconstruction Algorithm (ISRA), it has been introduced [8] as an acceleration of EM and it is convergent to the constrained minimum of \( D_{LS} \) [9].

Analogously, when \( \eta \) is a vector of independent and identically distributed Poisson variables with parameter given by \( Hx \), the negative logarithm of the likelihood (6) is equivalent to the Kullback Leibler divergence
\[ D_{KL}(y, x) = \sum_{i=1}^{N} y_i \log \frac{y_i}{(Hx)_i} + (Hx)_i - y_i \] (11)
Hence, the constrained ML problem (5) is equivalent to the minimization of \( D_{KL} \) under the non-negativity constraint [4]. In the Poisson case the KKT conditions (7) lead to the following iterative algorithm
\[ x^{(k+1)} = x^{(k)} \frac{H^T y}{H^T H x^{(k)}} \] (12)
which is known as Expectation Maximization [11], or also as Richardson Lucy algorithm [20, 15] when \( H \) represents a convolution operator and it is convergent to the constrained minimum of \( D_{KL} \) [21].
The limit solutions reached by EM and ISRA satisfy two analogous properties. In fact, let us first observe that, if

$$H(C) := \{ y' \in \mathbb{R}^N \mid y' = \sum_{j=1}^{M} a_j H_{ij}, \ a_j \geq 0 \} ,$$

then

**Lemma 2.1.** The minimum of the functions $D_{LS}$ and $D_{KL}$ is zero if and only if $y \in H(C)$.

**Proof.** We start by remarking that $D_{LS}$ and $D_{KL}$ are nonnegative functions. So if they assume the zero value, this is the minimum. For both $D_{LS}$ and $D_{KL}$ it is evident that the zero value is assumed if and only if $Hx = y$. Therefore, the minimum of the functions $D_{LS}$ and $D_{KL}$ is zero if and only if there exists at least one point $\bar{x} \in C$ such that $H\bar{x} = y$. Such a point exists if and only if the data $y$ belongs to $H(C)$.

Since EM and ISRA converge to a minimum of the functionals $D_{KL}$ and $D_{LS}$ respectively, Lemma 2.1 implies the following:

**Proposition 2.1.** If $y \not\in H(C)$ then the algorithms EM and ISRA converge to a solution $x^{(\infty)}$ such that

$$D_{KL}(y, x^{(\infty)}_{EM}) > 0 \quad \text{and} \quad D_{LS}(y, x^{(\infty)}_{ISRA}) > 0$$

respectively.

### 3. Regularization for a constrained ML problem

The noise corrupting the data $y$ makes the constrained ML solution not physically acceptable. In this case one can get an estimation of the unknown signal exploiting regularization. However, the definition of regularization given by Tikhonov [24] does not explicitly use statistical concepts, but takes place in the framework of functional analysis where the noise is modeled as a generic perturbation of the data in a metric space, instead of as a random variable. In order to provide a statistical definition of regularization we utilize the concept of coefficient of variation [10].

Given a random variable $N$ with mean $\mu$ and standard deviation $\sigma$, the coefficient of variation is defined as the ratio $\delta = \sigma/\mu$, which is an inversely proportional measure of the signal to noise ratio. Let us denote with $y_{\delta} = N_\delta(\mu)$ a realization of the random variable of a given coefficient of variation $\delta$, and let the mean $\mu$ be modeled by the action of a linear operator on the unknown parameter $x$, i.e. $\mu = Hx$.

**Definition 3.1.** When $\mathcal{L}(x, y_{\delta})$ is the likelihood associated with $y_{\delta} = N_\delta(Hx)$ and $C$ is a convex subset of $\mathbb{R}^M$, an operator $R_\alpha : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}^M$ is said to be a regularizing operator for the constrained ML problem

$$\arg\max_{x \in C} \mathcal{L}(x, y_{\delta})$$

if the following conditions hold:
1) \( R_\alpha \) is defined on the range of \( H \);

2) there exists a function \( \alpha = \alpha(\delta) \) such that, when the coefficient of variation tends to zero, \( R_{\alpha(\delta)}(y_\delta) \) tends to a constrained solution \( \hat{x} \) of problem (15), i.e.

\[
\lim_{\delta \to 0} R_{\alpha(\delta)}(y_\delta) = \hat{x}.
\]

(16)

In the case of iterative methods, when an algorithm \( x^{(k)} \) converges to a constrained ML solution \( \hat{x} \), condition 2) can be restated as

2a) there exists a function \( k = k(\delta) \) such that,

\[
\lim_{\delta \to 0} k(\delta) = \infty.
\]

(17)

We note that when \( \mathcal{N} \) is a Gaussian variable, \( H \) represents an invertible linear operator, and \( \mathcal{C} = \mathbb{R}^N \), the constrained ML problem becomes equivalent to the unconstrained least square problem, and so this definition coincides with the well-known Tikhonov definition (specifically, as \( \delta \) tends to zero, the unconstrained ML solution \( \hat{x} \) tends to the so-called generalized solution). When \( \mathcal{N} \) is a Poisson variable the second statement of Definition 3.1 requires that the solution provided by the regularization operator \( R_{\alpha}(y) \) tends to a constrained ML solution \( \hat{x} \) as the mean of the Poisson variable tends to infinite. Hence the properties 2) and 2a) have to hold asymptotically.

As a first issue, we discuss the conditions when traditional stopping rules define a regularization algorithm in the sense of Definition 3.1. We consider the Morozov’s discrepancy principle for the white Gaussian noise case and three different rules for Poisson noise, i.e.

i) Adapted Morozov’s discrepancy principle,

ii) Pearson’s cumulative test,

iii) Poisson discrepancy criterion.

The most widely known and used criterion is the Morozov’s discrepancy principle. It was first developed in the case of signal corrupted by Gaussian noise [16], but an adapted version can be restated for the Poisson case [11, 22]. Pearson’s cumulative test [26, 12] becomes the same as Morozov’s discrepancy principle when the noise is white Gaussian, so it has been considered in the Poisson case. Poisson discrepancy criterion is a recently formulated stopping rule appropriate for Poisson noise [3]. We will show now that only if \( y \in H(\mathcal{C}) \) these stopping rules satisfy Definition 3.1 with the consequence that, only when that condition is satisfied, ISRA and EM supplied with one of these stopping rules become well-defined regularization methods.

3.1. Gaussian case

Given the ISRA iterative process, Morozov’s discrepancy principle says that a reliable estimate of the solution is obtained by choosing the first iteration \( k \) such that

\[
\| H x^{(k)} - y \|^2 \leq \tau N \sigma^2,
\]

(18)

for some fixed \( \tau > 0 \). For this rule the following proposition holds true.
Theorem 3.1. ISRA supplied with the Morozov’s discrepancy principle becomes a well-defined regularization method if and only if \( y \in H(C) \).

Proof. We have to verify the second condition of Definition 3.1. When \( \sigma = 0 \), the stopping criterion is equivalent to the requirement that ISRA converges to the zero of the constrained LS functional. From Lemma 2.1 this follows if and only if \( y \in H(C) \). \( \square \)

3.2. Poisson case

We first introduce two general properties of EM. The first one can be summarized as follows: EM produces scaled reconstructions when the input data are scaled. In fact,

Lemma 3.1. Given a data \( y \), let us consider a scalar \( L > 0 \) and the corresponding scaled data \( y_L = Ly \). Let \( x^{(k)} \) indicate the \( k \)-th EM iteration with entry data \( y \) as in (12), and \( x^{(k)}_L \) indicates the \( k \)-th EM iteration with entry data \( y_L \). The following relation holds true

\[
x_L^{(k)} = Lx^{(k)} .
\] (19)

Proof. Let \( x^{(0)} \) the algorithm initialization. At the first iteration, with entry data \( y_L \), we have

\[
x^{(1)}_L = \frac{x^{(0)}_L}{H^T_1} H^T_1 Ly H x^{(0)}_L = Lx^{(1)} .
\] (20)

From the second iteration onwards (\( k \geq 1 \)) we have

\[
x^{(k+1)}_L = \frac{Lx^{(k)}_L}{H^T_1} H^T_1 - \frac{Ly}{L H x^{(k)}_L} = Lx^{(k+1)} ,
\] (21)

and hence the thesis holds true. \( \square \)

We will also use the following well-known property of EM:

Lemma 3.2. For each iteration \( k \) the relation

\[
\sum_{i=1}^{N} (H x^{(k)})_i = \sum_{i=1}^{N} y_i
\] (22)

holds.

Proof. The thesis follows directly from computations

\[
\langle H x^{(k)} , 1 \rangle = \left\langle \frac{x^{(k-1)}}{H^T_1} H^T_1 \frac{y}{H x^{(k-1)}} , H^T_1 \right\rangle = \left\langle \frac{y}{H x^{(k-1)}} , H x^{(k-1)} \right\rangle = \langle y , 1 \rangle .
\] \( \square \)
In the Poisson framework, the adapted version of Morozov’s discrepancy principle says that a reliable estimate of the solution is obtained by choosing the first iteration \( k \) such that
\[
\| H x^{(k)} - y \|^2 \leq \tau \sum_{i=1}^{N} (Hx^{(k)})_i ,
\] (23)
where \( x^{(k)} \) represents the EM iteration and \( \tau \) is a fixed positive number.

**Theorem 3.2.** *EM supplied with the adapted Morozov’s discrepancy principle becomes a well-defined regularization method if and only if \( y \in H(\mathcal{C}) \).*

**Proof.** We have to prove that the two conditions of Definition 3.1 are satisfied. The first one is obvious. For the second one, we recall that given a Poisson random variable with coefficient of variation \( \delta \), for any realization \( y_\delta \) the component \((y_\delta)_i\) tends to \( \infty \) as \( \delta \to 0 \) and hence \( \lim_{\delta \to 0} \sum_{i=1}^{N} (y_\delta)_i = \infty \). As a consequence of that, when \( k \) can be written as a function of some parameter proportional to \( \sum_{i=1}^{N} (y_\delta)_i \), i.e. \( k = k(L) \), with \( L = \gamma \sum_{i=1}^{N} (y_\delta)_i \), \( \gamma > 0 \), we can rewrite condition 2a) in Definition 3.1 as
\[
\lim_{L \to \infty} k(L) = \infty .
\] (24)

Now, in order to prove the thesis we will show that \( k \) can be written as \( k(L) \) and we will verify condition (24). Let \( x^{(k)} \) be the \( k \)-th EM iteration (12) with data entry \( y \) and let \( x^{(k)}_L \) be the \( k \)-th EM iteration with data entry \( y_L = Ly \) with \( L > 0 \). The adapted Morozov’s discrepancy principle for \( x^{(k)}_L \) consists in finding the first iteration \( k \) such that
\[
\| H x^{(k)}_L - y_L \|^2 \leq \tau L \sum_{i=1}^{N} (y_L)_i .
\] (25)
This relation can be written in terms of \( y \) and the corresponding EM iteration \( x^{(k)} \) using Lemma 3.1 as
\[
\| H x^{(k)} - y \|^2 \leq \frac{\tau}{L} \sum_{i=1}^{N} y_i .
\] (26)
For \( L \to \infty \) we have
\[
\| H x^{(k)} - y \|^2 \leq 0 .
\] (27)
From Lemma 2.1 and the convergence property of EM, the l.h.s. of (27) converges to zero as \( k \) tends to \( \infty \) if and only if \( y \in H(\mathcal{C}) \). Therefore condition (24) is satisfied and the thesis is proved.

Given the EM iterative process, Pearson’s cumulative test says that a reliable estimate of the solution is obtained by choosing the first iteration \( k \) such that
\[
\sum_{i=1}^{N} \frac{(Hx - y)_i^2}{(Hx)_i} \leq \tau N ,
\] (28)
where \( N \) represents the number of data points, as discussed in [19], and \( \tau \) is a fixed positive number.
Theorem 3.3. *EM supplied with the Pearson’s cumulative test becomes a well-defined regularization method if and only if \( y \in H(C) \).*

Proof. Let \( x^{(k)} \) be the EM iteration with data entry \( y \). By taking \( y_L = Ly \) with \( L > 0 \) the stopping criterion for EM applied to the data \( y_L \) takes the form

\[
\sum_{i=1}^{N} \frac{(Hx^{(k)}_i - y)^2}{(Hx^{(k)}_i) \leq \tau \frac{N}{L}}
\]

and the thesis follows as in Theorem 3.2.

Given the EM iterative process, Poisson discrepancy criterion says that a reliable estimate of the solution is obtained by choosing the first iteration \( k \) such that

\[
\frac{2}{N} \sum_{i=1}^{N} y_i \log \frac{y_i}{(Hx^{(k)}_i)} + (Hx^{(k)})_i - y_i \leq \tau ,
\]

for some positive number \( \tau \). Poisson discrepancy criterion with \( \tau = 1 \) has been proposed both for choosing the regularization parameter in an EM scheme with penalty term and for stopping the unpenalized EM iterative process [5].

Theorem 3.4. *EM supplied with the Poisson discrepancy criterion becomes a well-defined regularization method if and only if \( y \in H(C) \).*

Proof. Let \( x^{(k)} \) be the EM iteration with data entry \( y \). By taking \( y_L = Ly \), with \( L > 0 \), the stopping criterion for EM applied to the data \( y_L \) takes the form

\[
\frac{2}{N} \sum_{i=1}^{N} y_i \log \frac{y_i}{(Hx^{(k)}_i)} + (Hx^{(k)})_i - y_i \leq \tau \frac{N}{L}
\]

and the thesis follows as in Theorem 3.2.

These theorems point out that condition \( y \in H(C) \) is crucial for these classical stopping rules to define a regularization algorithm for ISRA and EM. Tikhonov defines the cases where this property is not satisfied as genuinely ill-posed. In practice, what very often happens is that data do not belong to the convex cone \( H(C) \). Indeed, if the true object \( x_T \) contains even one zero value, the corresponding signal \( y_T = Hx_T \) belongs to the frontier of \( H(C) \), and any little variation due to noise fluctuation can move the data \( y \) outside \( H(C) \). Moreover, in such cases the standard stopping rules could not work. We will see an example in Section 5.

In the next Section we will introduce a stopping criterion defining a regularization algorithm for constrained ML problems that is valid even in the case of genuinely ill-posed problems.
4. Constrained stopping criterion

Our aim is to realize a stopping rule which gives rise to a regularization algorithm for every \( y \in \mathbb{R}^N \). To do this we first observe that the regularization methods discussed in the previous section are based on equations of the kind

\[
    r(x^{(k)}, y) \leq \tau \mathbb{E}(r(x^{(k)}, y)),
\]

where \( r(x, y) \) is a function specific for each rule and \( \tau \) is a fixed positive number.

Choosing \( r(x, y) = D_{LS}((x, y)) \) we find the Morozov’s discrepancy principle; choosing \( r(x, y) = \| (Hx - y)/\sqrt{H}x \|^2 \) we find the Pearson’s test criterion; choosing \( r(x, y) = D_{KL}(y, x) \) we find the Poisson discrepancy principle.

The main drawbacks of these criteria is that for genuinely ill-posed problems, i.e. when \( y \notin H(C) \), the l.h.s. of (32) does not converge to 0 for \( k \to \infty \) (see Lemma 2.1), and therefore it may happen that the stopping rule is never applied. A choice of \( r(x, y) \) that for sure overtakes this difficulty is a function converging to 0 for \( k \to \infty \) and whose expected value is positive. In this paper we propose to choose

\[
    r(x, y) = \| x \nabla L_y(x) \|^2.
\]

In fact, this choice ensures that given an algorithm \( x^{(k)} \) converging to the constrained minimum of (6), then \( \lim_{k \to \infty} r(x^{(k)}, y) = 0 \). Let us call \( r(x, y) \) the constrained backprojected residual (CBR) and therefore the associated stopping rule (32) is the CBR criterion. In the following we describe the CBR criterion for the Gaussian and Poisson cases and prove that, if applied to ISRA and EM, the stopping rule leads to two regularization algorithms in the sense of Definition 3.1, without any restriction on the input data. In the next section, we will also show that the CBR criterion works in a very reliable way, using applications to astronomical image processing.

4.1. Gaussian case

In the Gaussian case the CBR takes the form

\[
    r(x, y) = \| x H^T (Hx - y) \|^2,
\]

and the following result holds true:

**Proposition 4.1.** In the Gaussian case, the expected value of \( r(x, y) \) is

\[
    \mathbb{E}(r(x, y)) = \sum_{j=1}^{M} x_j^2 (H_2^T \sigma^2)_j,
\]

where \( (H_2)_{ij} = (H_{ij})^2 \).

**Proof.** The expected value of each component is

\[
    \mathbb{E} \left[ x_j^2 (H^T (Hx - y))^2_j \right] = x_j^2 \mathbb{E} \left[ (H^T (Hx - y))^2_j \right].
\]
Stopping criterion for constrained maximum likelihood algorithms

Since the noise is independently distributed, the expected value of the product \( (Hx - y)_i(Hx - y)_h \) is different from zero only if \( i = h \). Then

\[
\mathbb{E} \left[ x_j^2 (H^T (Hx - y))_j^2 \right] = x_j^2 (H_2^T \sigma^2)_j ,
\]

since \( \mathbb{E} [(Hx - y)_i^2] = \sigma^2 \) for all \( i \). By summing up all the components we get the thesis. \( \square \)

Now, we can explicitly write the CBR criterion for the Gaussian case. Giving an algorithm \( x^{(k)} \) converging to a constrained LS solution, the criterion says that a reliable estimate of the solution is obtained by choosing the first iteration \( k \) such that:

\[
\| x^{(k)} H^T (Hx^{(k)} - y) \|^2 \leq \tau \sum_{i=1}^{N} (x^{(k)}_i)^2 (H_2^T \sigma^2)_i ,
\]

with \( \tau \) fixed positive number.

**Theorem 4.1.** ISRA supplied with the CBR criterion \( (38) \), becomes a well-defined regularization method.

**Proof.** The first condition of Definition 3.1 is obvious. As for the second condition, in the Gaussian case the coefficient of variation \( \delta \) tends to 0 if and only if the standard deviation \( \sigma \) tends to 0. Therefore to compute the limit in the second condition of Definition 3.1 we can set \( \sigma = 0 \) and so the CBR criterion \( (38) \) for \( \delta \rightarrow 0 \) becomes

\[
\| x^{(k)} H^T (Hx^{(k)} - y) \|^2 \leq 0 .
\]

Since the algorithm is convergent, the l.h.s. in (39) converges to 0. \( \square \)

4.2. Poisson case

In the Poisson case, the CBR takes the form

\[
r(x, y) = \left\| x H^T \left( 1 - \frac{y}{Hx} \right) \right\|^2 ,
\]

and the following result holds true:

**Proposition 4.2.** In the Poisson case, the expected value of \( r(x, y) \) is

\[
\mathbb{E}(r(x, y)) = \sum_{j=1}^{M} x_j^2 \left( H_2^T \frac{1}{Hx} \right)_j ,
\]

where \( (H_2)_{ij} = (H_{ij})^2 \).

**Proof.** The expected value of each component is

\[
\mathbb{E} \left[ x_j^2 \left( H^T \left( 1 - \frac{y}{Hx} \right) \right)_j^2 \right] = x_j^2 \mathbb{E} \left[ \left( H^T \left( 1 - \frac{y}{Hx} \right) \right)_j^2 \right] = x_j^2 \mathbb{E} \left[ \left( \frac{1}{Hx} (Hx - y) \right)_j^2 \right] .
\]
As in the Gaussian case, noise is independently distributed. Therefore we can exploit
again the fact that the expected value of the product \((Hx - y)_i(Hx - y)_h\) is different
from zero only if \(i = h\). Then
\[
\mathbb{E} \left[ x_j^2 \left( H^T \left( 1 - \frac{y}{Hx} \right) \right)_j \right] = x_j^2 \left( H^T \left( \frac{1}{(Hx)^2} \mathbb{E} [(Hx - y)^2] \right) \right)_j.
\]
(42)
Since noise is Poisson, \(\mathbb{E} [(Hx - y)^2] = Hx\). Summing up all the components we have
the thesis.

Now, we can explicitly write the CBR criterion for the Poisson case. The criterion
says that a reliable estimate of the solution is obtained by choosing the first iteration \(k\)
such that
\[
\| x^{(k)} H^T \left( 1 - \frac{y}{Hx^{(k)}} \right) \|^2 \leq \tau \sum_{i=1}^{N} (x^{(k)})_i^2 \left( H^T \frac{1}{Hx^{(k)}} \right)_i,
\]
where \(x^{(k)}\) is the EM iteration and \(\tau\) is a fixed positive number. To prove the
regularization property for this stopping rule applied to EM, we need the following
Lemma.

**Lemma 4.1.** Let \(x^{(k)}\) be the \(k\)-th iteration of EM. The following inequality holds:
\[
\sum_{i=1}^{N} (x^{(k)})_i^2 \left( H^T \frac{1}{Hx^{(k)}} \right)_i \leq \sum_{i=1}^{N} y_i.
\]
(44)

**Proof.** For simplicity of notation we use \(x\) instead of \(x^{(k)}\). We begin by noting that
\[
\sum_{i=1}^{M} x_j^2 \left( H^T \frac{1}{Hx} \right)_j = \sum_{i=1}^{N} (H^T \frac{1}{Hx})_i.
\]
(45)
Then
\[
\sum_{i=1}^{N} \sum_{j=1}^{M} h_{ij} x_j^2 \leq \sum_{i=1}^{N} \sqrt{\sum_{j=1}^{M} h_{ij}^2 x_j^2} \leq \sum_{i=1}^{N} \sum_{j=1}^{M} h_{ij} x_j \leq \sum_{i=1}^{N} y_i
\]
having used the relation \(\sqrt{\sum_{j=1}^{M} h_{ij}^2 x_j^2} \leq \sum_{j=1}^{M} h_{ij} x_j\) and Lemma [3.2]

**Theorem 4.2.** EM supplied with the CBR criterion \((43)\) becomes a well-defined
regularization method.

**Proof.** The first condition of Definition [3.1] is obvious. For the second one, we consider
a given data \(y\) and the data \(y_L = Ly\) scaled by a number \(L > 0\). We will prove that
when \(L\) tends to infinity the number of iterations needed to satisfy the CBR criterion
tends to infinity. According to the remark at the beginning of the proof of Theorem [3.2]
this prove the condition 2a) of the Definition [3.1]
Stopping criterion for constrained maximum likelihood algorithms

Given the data $y_L$ and the corresponding EM algorithm $x_L^{(k)}$ the CBR criterion stops the iterative process at the first iteration $k$ such that

$$\left\| x_L^{(k)} H^T \left( 1 - \frac{y_L}{H x_L^{(k)}} \right) \right\|^2 \leq \tau \sum_{i=1}^{M} (x_L^{(k)})^2_j \left( H^2_2 \frac{1}{H x_L^{(k)}} \right)_j,$$

(46)

for a fixed $\tau > 0$. This equation can be written by replacing $y_L$ with $y$ and the corresponding EM algorithm $x_L^{(k)}$ with $x^{(k)}$ by using Lemma 3.1 and picking up the $L$ factor, i.e.

$$\left\| x^{(k)} H^T \left( 1 - \frac{y}{H x^{(k)}} \right) \right\|^2 \leq \frac{\tau}{L} \sum_{i=1}^{M} (x^{(k)})^2_j \left( H^2_2 \frac{1}{H x^{(k)}} \right)_j \leq \frac{\tau}{L} \sum_{i=1}^{N} y_i ,$$

(47)

where the second inequality holds by Lemma 4.1. For $L \to \infty$ we have

$$\left\| x^{(k)} H^T \left( 1 - \frac{y}{H x^{(k)}} \right) \right\|^2 \leq 0 .$$

(48)

Since the algorithm is convergent, the l.h.s. in (48) converges to 0.

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) [14] mission has been launched by NASA in February 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.

We first validate the stopping rule introduced in this paper for ISRA and EM, in the case of synthetic data mimicking the modulation introduced by RHESSI grids and affected by Gaussian and Poisson noise, respectively. Then we investigate the behavior of EM equipped with the new criterion when applied to a set of experimental RHESSI observations.

5.1. Simulated data

Solar flares are sudden and intense explosions occurring high in the solar corona and accelerating electrons down to the thicker chromospheric part of the solar atmosphere. While diving into the plasma driven by the flow lines of intense magnetic fields, these electrons emit hard X-rays by collisional bremsstrahlung. As a result, typical hard X-ray source configurations observed by RHESSI present one, two or a higher number of bright footpoints on a weak background. In this example we consider a simulated source configuration mimicking the structure and the physical properties of a real flare observed...
by RHESSI on July 23 2002 (Figure 1). Such configuration is contained in a 64 by 64 pixel image. The sources have Gaussian form and are arranged according to the following scheme: the left source (source L) is located at coordinates (16, 32) with variance 0.64 and amplitude 1.28, the center source (source C) is located at coordinates (32, 32) with variance 0.64 and amplitude 1.6, the upper right source (source UR) is located at coordinates (42, 45) with variance 0.48 and amplitude 0.6 and the lower right source (source LR) is located at coordinates (42, 19) with variance 0.64 and amplitude 1.28. Using the routines of Solar SoftWare (SSW) [http://www.mssl.ucl.ac.uk/surf/sswdoc/] we reproduced the RHESSI acquisition process when the grids are reached by the photon flux emitted by such simulated source constellation and simulated two different sets of count modulation profiles using detectors from 3 through 8.

The first set is affected by white Gaussian noise with standard deviation \( \sigma = 10 \). We applied ISRA against these synthetic data and we stopped the iterations both in correspondence with the minimum of the L2-norm of the difference between the reconstructed and the simulated configuration (L2-norm criteria) and by means of the CBR criterion, using \( \tau = 1/\sigma^2 \) (such choice has been made to contrast the typical oversmoothing property of discrepancy methods). L2-norm minimization applies at the 518-th iteration and CBR criterion at the 645-th iteration. Figure 2 shows the images obtained with the two stopping rules. The shapes of the sources determined using the two methods are very similar and are also consistent with the Gaussian source structures assumed in the original configuration. On the other hand the image provided by the CBR rule shows a better separation for the four sources. We used the images obtained by the two stopping rules to compute the total flux emitted by the four sources within a square with side length 13 pixels and located around the center of the sources of the original configuration. Then we computed the rate of the flux reconstructed by the two criteria for each sources. Table 1 shows that for source C the difference in
Table 1. Comparison between the photometry of the solutions provided by ISRA stopped using the CBR and the L2-norm. The first column represents the original flux integrated in a 13 by 13 square located around the center of the sources. The second and the third columns indicate the fluxes reconstructed by the two methods within the same square, and the fourth and fifth columns show the rates of the reconstructed fluxes.

| Source position | flux | L2-norm | CBR | L2-norm % | CBR % |
|-----------------|------|---------|-----|-----------|-------|
| C               | 5.884| 4.025   | 4.057| 68.4      | 68.9  |
| L               | 4.691| 3.882   | 4.048| 82.7      | 86.2  |
| BR              | 4.762| 4.265   | 4.456| 89.5      | 93.5  |
| UR              | 1.824| 1.454   | 1.654| 79.7      | 90.6  |

Figure 2. From left to right: the reconstruction with ISRA stopped in correspondence with the minimum of L2-norm (518-th iteration) and the reconstruction with ISRA stopped according to CBR criterion (645-th iteration).

reconstructing the original photometry between CBR criterion and L2-minimization is negligible. Conversely, the CBR rule has a better photometry in the case of the other three sources and, particularly, in the case of source UR, which is the weakest one in the simulated configuration.

The second set of synthetic profiles has been affected by Poisson noise and EM was applied to obtain images in Figure 3. The left panel of Figure 3 shows the image obtained stopping the iterations using the L2-norm, and the right panel shows the image obtained using the CBR (coherently with the Gaussian case, here we chose \( \tau = N/\sum_{i=1}^{N} y_i \)). The L2-norm rule stops EM at the 133-th iteration and the CBR criterion at the 229-th iteration. Both reconstructions show that source UR has a smaller support with respect to what obtained by ISRA in the case of white Gaussian noise. This is a direct consequence of the fact that Poisson noise is signal dependent and hence the weakest source has smaller signal to noise ratio with respect to the other sources. However, Table 2 shows that for all sources and specifically for source UR, the CBR criterion reconstructs the photometry in a much more accurate way.
5.2. Real data

We then studied the behavior of EM regularized by the CBR criterion 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 \(1.38 \times 10^6\), the number of data is \(N = 3816\). During the second event the total number of counts collected is about \(7.45 \times 10^4\), the number of data is \(N = 3168\). In both cases the reconstructed field of view is a square of 80 arcseconds side length corresponding to a 64 by 64 pixel image.

For these two events EM regularized with the CBR criterion provides the two reconstruction in Figure 4. In the left panel case, the stopping rule applies after 498 iterations while in the right pane case after 822 iterations. In order to compare the behaviors of the CBR criterion and of the other criteria for Poisson noise described in Section 4, in Figure 5 we computed the discrepancy equations corresponding to the
Morozov’s discrepancy (3.2), Pearson’s test (3.3) and Poisson discrepancy (3.4). For five cases over six, the criteria never apply since the left hand sides of the discrepancy equations never intersect the corresponding right hand sides. There is just one case that works (Pearson’s test on the reconstruction of the September 08 2002 event) but just after 1000 iterations and the corresponding reconstruction is clearly undersmoothed.

6. Conclusions

In the present paper we have formulated a general stopping criterion for constrained ML algorithms. This new criterion, called CBR criterion, is based on the statistical properties of the signal and takes into account the constraint on solution. We have generalized the Tikhonov definition of regularization for constrained ML problems and we have proved that ISRA and EM procedures, equipped with the CBR criterion, are well-defined regularization algorithms. Moreover, we have also proved that the traditional stopping rules, applied to any convergent constrained ML algorithm, do not define regularization algorithms in the case of genuinely ill-posed problems.

We have illustrated the method first against synthetic count modulation profiles simulated in the framework of an X-ray solar mission. Specifically an analysis of photometry in specific region of interest showed the accuracy of this new stopping rule. We also considered two real observations and reconstructed the X-ray sources with the new method, obtaining reliable flaring configurations. Finally, we point out that other stopping rules traditionally applied in the case of Poisson data do not properly work.
A systematically validation of this approach is under construction in both astronomical and medical imaging applications, in the case the noise is Poisson, the algorithm adopted is EM and the model equations is genuinely ill-posed.

Acknowledgments

The work has been supported by the EU FP7 HESPE grant no. 263086.

References

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

[2] Barrett H H and Myers K J, 2003 Foundations of Image Science (New York: Wiley)
[3] Benvenuto F, Zanella R, Zanni L and Bertero M, 2010 Nonnegative least-squares image deblurring: improved gradient projection approaches. Inverse Problems 26 (2)
[4] Bertero M and Boccacci P, 1998 Introduction to Inverse Problems in Imaging (Bristol: Institute of Physics Publishing)
[5] Bertero M, Boccacci P, Talenti G, Zanella R and Zanni L, 2010 A discrepancy principle for Poisson data. Inverse Problems 26 (10)
[6] Boyd S and Vandenberghe L, 2004 Convex Optimization (Cambridge: University Press)
[7] Coakley K J and Llacer J, 1991 The use of cross-validation as a stopping rule and reconstruction of emission tomography images. Proceeding of SPIE medical imaging V: Image Physics pp. 226–233
[8] Daube-Witherspoon M E and Muehllehner G, 1986 An iterative image space reconstruction algorithm suitable for volume ECT. IEEE Transactions on Medical Imaging 5 (2) pp. 61–66
[9] De Pierro A R, 1987 On the convergence of the iterative image space reconstruction algorithm for volume ECT. IEEE Trans. Med. Imaging 6 pp. 174–175
[10] DeGroot M H and Schervish M J, 2002 Probability and Statistics (Pearson)
[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] Hebert T J, 1990 Statistical stopping criteria for iterative maximum likelihood reconstruction of emission images. Phys. Med. Biol. 35 pp. 1221–1232
[13] Lantéri H, Roche M and Aime C, 2002 Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms. Inverse Problems 18
[14] 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
[15] Lucy L B, 1974 An iterative technique for the rectification of observed distributions. Astronomical Journal 79 pp. 745–754
[16] Morozov V A, 1984 Methods for solving incorrectly posed problems (New York: Springer)
[17] Natterer F, 1983 On the order of regularization methods. In Improperly posed problems and their numerical treatment, vol. 63 of Internat. Schriftenreihe Numer. Math. (Basel: Birkhäuser), pp. 189–203
[18] Perry K M and Reeves S J, 1995 Generalized cross-validation as a stopping rule for the Richardson-Lucy algorithm. International Journal of Imaging Systems and Technology 6 pp. 387–391
[19] Puetter R, Gosnell T and Yahil A, 2005 Digital image reconstruction: Deblurring and Denoising. Annu. Rev. Astron. Astrophys (43) pp. 139–194
[20] Richardson W H, 1972 Bayesian-Based Iterative Method of Image Restoration. J. Opt. Soc. Am. 62 (1) pp. 55–59
[21] Shepp L A and Vardi Y, 1982 Maximum Likelihood Reconstruction for Emission Tomography. IEEE Transactions on Medical Imaging 1 (2) pp. 113–122
[22] Staglianò A, Boccacci P and Bertero M, 2011 Analysis of an approximate model for Poisson data reconstruction and a related discrepancy principle. Inverse Problems 27 (12)
[23] Tarantola A, 2004 Inverse Problem Theory and Methods for Model Parameter Estimation (Philadelphia: Society for Industrial and Applied Mathematics)
[24] Tikhonov A N and Arsenin V Y, 1977 Solution of Ill-posed Problems (Washington: Winston and Sons)
[25] Tzanakos G and Kontaxakis G, 1994 Statistical behavior of the updating coefficients in the EM algorithm for PET. In Nuclear Science Symposium and Medical Imaging Conference, 1994, vol. 3. pp. 1310–1314
[26] Veklerov E and Llacer J, 1987 Stopping rule for the maximum likelihood estimation algorithm based on statistical hypothesis testing. IEEE Trans. Med. Imag. 6 pp. 313–319