MOROZOV PRINCIPLE FOR KULLBACK-LEIBLER RESIDUAL TERM AND POISSON NOISE

BROUN SIXOU, TOM HOHWEILLER AND NICOLAS DUCROS
CREATIS, CNRS UMR 5220, INSERM U1044, INSA de Lyon, Université de Lyon 1
Université de Lyon, 69621, Villeurbanne Cedex, France

Abstract. We study the properties of a regularization method for inverse problems corrupted by Poisson noise with Kullback-Leibler divergence as data term. The regularization parameter is chosen according to a Morozov type principle. We show that this method of choice of the parameter is well-defined. This a posteriori choice leads to a convergent regularization method. Convergence rates are obtained for this a posteriori choice of the regularization parameter when some source condition is satisfied.

1. Introduction. We consider in this work the linear inverse problem:

\[ AX = Y \]

where \( A : L_1(\Omega) \rightarrow L_1(\Omega) \) is a linear compact operator continuous with respect to the weak topology of \( L_1(\Omega) \), and \( \Omega \) a bounded open subset. The ideal data \( Y \) can be interpreted as a photon density with intensity \( y \) and exposure time \( t \) such that \( Y = yt \). The observed data \( G_t \) is a Poisson process with measure \( yt \), and it can be written \( G_t = \sum_{i=1}^{N_t} \delta_{x_i} \), where the \( x_i \) denotes the positions of the \( N_t \) detected photons.

Poisson inverse problems appears in many astronomical imaging and medical imaging problems like positron emission tomography or multi-spectral computerized tomography [17, 43, 33, 30]. This problem is ill-posed and must be regularized. A regularization functional is usually based on a weighted combination of a data residual term and of a regularization term [21]. When the residual minimizer is a quadratic, a classical choice for the regularization term is some Sobolev space norm as in Tikhonov regularization [44]. For Poisson noise, quadratic data misfit functional have to be replaced by the Kullback-Leibler (KL) functional [29]. In a variational framework, the unknown is obtained with the minimization of a functional which is the sum of a KL data term and of a regularization term which includes a priori knowledge on the solution \( X \). A regularization parameter is balancing the two terms. The Kullback-Leibler divergence is a well-known example of Bregman distance. The Bregman distance is now a widely used tool to study optimization and signal processing algorithms [13]. It is very successful to establish convergence rates [36, 14]. It can also be used to build new data fidelity functionals. The solution obtained can be understood as a maximum a posteriori estimator in a Bayesian framework with exponential families. Inverse problems with the Kullback-Leibler divergence as data term have been been much studied in the context of Poisson

2010 Mathematics Subject Classification. Primary: 65J22, 65J20, 65K10; Secondary: 52A41.
Key words and phrases. Inverse problems, Kullback-Leibler divergence, Morozov principle.
noise [6, 7, 2, 34, 35, 12, 10, 20, 28]. In a deterministic context, this regularization has been investigated in Pöschl et al [34]. For a statistic noise hypothesis, the problem has been investigated by Werner et al. [46, 47]. Some convergence results have been obtained for the mean value of the Bregman divergence associated to the regularization term measuring the distance between the reconstructed solution and the true solution.

As usual for inverse problems, the regularization parameter has to be chosen with care to ensure the convergence of the regularization method. In [46, 47], an a-priori choice of the regularization parameter is investigated, which requires the knowledge of the smoothness of the solution. Some convergence results for a parameter chosen with the Lepskii balancing principle are also presented. Other methods have been proposed for Poisson noise. We can mention the Bardsley et Goldes' approach [6], the Bertero method [9, 48] or the Santos and De Pierro method [38]. The Bertero method is based on the Morozov discrepancy principle and on a finite dimensional approximation of the inverse problem [9] but is not clear if this a posteriori method of choice of the regularization parameter leads to a convergent regularization scheme. Recently, some regularization methods where the regularization parameter and the discretization level are chosen at the same time have been investigated [1]. The aim of this work is to analyze the convergence properties of the Bertero method taking into account the discretization level.

This paper is organized as follows. In the second section, we summarize basic results about Poisson random variables, exponential families, the Bregman distances and the Kullback-Leibler divergence that will be used in the paper. We also present the Bertero approach. In section 3, we show that the Bertero-Morozov type principle is well-defined and that the regularization method for the choice of the regularization parameter is convergent. Some convergence rates are derived in Section 4 under suitable source conditions. Some numerical experiments are presented in Section 5. A conclusion ends this article.

2. Inverse problems with the Kullback-Leibler divergence data term. We summarize in this section well-known results about the Kullback-Leibler divergence that will be used in the following.

2.1. The Kullback-Leibler divergence. The Kullback-Leibler divergence is a Bregman divergence [8, 5] associated with the Poisson noise distribution.

Definition 2.1. A multivariate family of distributions \( \{p_{\theta,b}, \theta \in \text{Dom}(b) \subset \mathbb{R}^d \} \) defined on \( \mathbb{R}^d \) is called a regular exponential family [8] if each probability density is of the form:

\[
p_{\theta,b}(x) = \exp(<x, \theta> - b(\theta))p_0(x) \quad \forall \ x \in \mathbb{R}^d
\]

where \( p_0 \) is a well-defined probability density function. The parameter \( \theta \) is called the natural parameter of the exponential family and the function \( b \) the cumulant function with a domain \( \text{Dom}(b) \).

Definition 2.2. Let \( g : \mathbb{R} \to \mathbb{R} \) a strictly convex and differentiable function, \( G : L_p(\Omega) \to \mathbb{R} \) defined by \( G(v) = \int_\Omega g(v)dx \), and its Bregman domain \( \text{Dom}_B(G) = \{v, \partial G(v) \neq \emptyset\} \). The Bregman distance \( B_G : \text{Dom}(G) \times \text{Dom}(G) \to [0, \infty) \) is defined by:

\[
B_G(v_1, v_2) = \int_\Omega (g(v_1) - g(v_2) - g'(v_2)(v_1 - v_2))dx
\]
The probability density function for exponential families can be expressed with the Bregman distance associated to the Fenchel dual of $b$. This is expressed by the following theorem [34, 8, 5]:

**Theorem 2.3**. Let $p_{\theta,b}$ the probability density function of a regular exponential family distribution and $B_{b^*}$ the Bregman distance derived from $b^*$ the Fenchel dual of $b$. Then $p_{\theta,b}$ is uniquely determined as:

$$p_{\theta,b}(x) = \exp(-B_{b^*}(x))\forall x \in \text{dom}(b^*)$$

where the function $F_{b^*} : \mathbb{R} \to \mathbb{R}^+$ is uniquely determined and where $\mu$ is the expectation of $x$ with respect to $p_{\theta,b}$, $\mu = E[p_{\theta,b}(x)]$.

For Poisson noise distribution, the dual of the cumulant is the function denoted $g$ in the following and defined by $g(x) = b^*(x) = x\ln(x) - x$. The Kullback-Leibler divergence between two functions $u$ and $v$ in its domain is given by:

$$KL(u,v) = \int_{\Omega} (v - u + uln(\frac{u}{v}))dx$$

To avoid divergencies, it is possible to define a regularized distance:

$$KL_\epsilon(u,v) = \int_{\Omega} (v - u + uln(\frac{u+\epsilon}{v+\epsilon}))dx$$

where $\epsilon$ is a small parameter [46, 47].

From the definition of the Kullback-Leibler divergence, we have the following equality for any functions $u$, $v$ and $w$ in the domain of Kullback-Leibler divergence:

$$KL(u,v) + KL(w,u) - KL(w,v) = \int_{\Omega} (g'(u) - g'(v))(u - w)dx$$

The following properties will be useful in the following [36, 20].

**Proposition 2.4**. For any $u,v \in \text{Dom}(g)$, $\|v - u\|_1^2 \leq (\frac{2}{3}\|v\|_1 + \frac{4}{3}\|u\|_1)KL(u,v)$

**Proposition 2.5**. (i) The function $(u,v) \to KL(u,v)$ is convex.

(ii) For any fixed $v$, the function $KL(.,v)$ is lower semi-continuous with respect to the weak topology of $L^1(\Omega)$.

(iii) For any fixed $v$, the function $KL(v,.)$ is lower semi-continuous with respect to the weak topology of $L^1(\Omega)$.

(iv) Let $A : L^1(\Omega) \to L^1(\Omega)$ a compact, linear operator. For any $C > 0$ and any non-negative $u \in L^1(\Omega)$, the following sets are weakly compact in $L^1(\Omega)$:

$$\{x \in L^1(\Omega) : KL(Ax,u) \leq C\}$$

2.2. **Regularization functional with the Kullback-Leibler divergence data term and Morozov principle.** The use of the Kullback-Leibler divergence as data similarity term can then be understood in the Maximum a Posteriori (MAP) estimation framework. Let us denote $Y_i$ be a noisy perturbation of some unknown data $Y = AX$, $p(Y_i|X)$ the conditional probability of $Y_i$ given $X$, $p(X)$ and $p(Y_i)$ the a priori distributions of $X$ and $Y_i$. The continuous MAP estimation problem consists of maximizing the functional:

$$T(X) = \frac{p(Y_i|X)p(X)}{p(Y_i)}$$
with respect to \( X \). Taking the logarithm of \( T \), the problem corresponds to the minimization of the logarithmic MAP estimator

\[
- \log(T(X)) = KL(Y_t, AX) - \ln(p(X))
\]

In a deterministic context, some method for an a priori choice of the regularization parameter have been proposed in [34]. For a statistic noise hypothesis, the problem of the choice of this parameter has been studied by Werner et al. [46, 47]. On the other hand, a Morozov type discrepancy principle has been proposed for Poisson inverse problems. The Bertero discrepancy principle is in close relation with the Morozov discrepancy principle [21, 11, 3, 4].

In its basic formulation for Gaussian noise, the Morozov principle consists to chose the value of the parameter \( \alpha \) such that for a fixed \( \tau > 1 \), the solution \( x^\alpha \) for noise level \( \delta \) satisfies:

\[
\|Ax^\alpha - y^\delta\| = \tau \delta
\]

where \( y^\delta \) is the noisy data. The noise free data \( y \) are assumed to be corrupted by noise in such a way that:

\[
\|y - y^\delta\|_{L^2(\Omega)} \leq \delta
\]

The Bertero principle is the generalization of the method of choice of the regularization parameter for Poisson noise. It is based on a finite dimensional approximation of the inverse problem and on the following lemma: [9, 48]

**Lemma 2.6.** Let \( F(x, \lambda) = 2(x \log(x) - x + \lambda) \) then if \( x \) is a random variable which is Poisson distributed with parameter \( \lambda \), \( x \sim \text{Poisson}(\lambda) \), we have:

\[
E[F(x, \lambda)] = 1 + O\left(\frac{1}{\lambda}\right)
\]

Let assume \( Y_{t,m} = (Y_{i,t})_{1 \leq i \leq m} \) is a m-dimensional Poisson variable, where \( m \) may be the number of pixels of a detector. With the Bertero discrepancy principle the regularization parameter \( \alpha \) is chosen such that the finite dimensional minimizer \( X^\alpha_{t,m} \) of the regularization functional satisfies:

\[
KL(Y_{t,m}, A_m(X^\alpha_{t,m})) = \frac{m}{2}
\]

where \( A_m \) is a finite dimensional approximation of the direct operator \( A \). This principle works well and some numerical examples are detailed [9, 48]. Yet, to our knowledge, this approach has not been justified and the properties of the regularization method with this choice of the regularization parameter have not been investigated. They are studied in the next section.

3. The Morozov discrepancy principle for inverse problems with Poisson noise. In this section, we show that the use of the Morozov principle with the Kullback-Leibler functional is a well-defined and leads to a convergent regularization technique. The proof has many common points with the demonstration in the framework of Banach spaces [3, 4].

3.1. Poisson processes. The observed data \( G_t \) is a Poisson process, and it can be written \( G_t = \sum_{i=1}^{N_t} \delta_{x_i} \), where the \( x_i \) denotes the positions of the \( N_t \) detected photons at time \( t \). It can be seen as a random measure belonging to the space of integer valued Borel measures on \( \Omega \) [26, 37]. For any measurable subset \( A \), \( G_t(A) \) is an integer value random Poisson variable, with mean value denoted as \( \nu_t(A) \). We assume that the mean measure \( \nu_t \) is absolutely continuous with respect to the
Lebesgue measure and proportional to the time of exposure $t$. We have thus $\nu_t = yt$ and:
\begin{equation}
E[G_t(A)] = t \int_A y \, dx = \nu_t(A)
\end{equation}
where $y$ is the scaled photons density. For a measurable function $\phi : \Omega \to \mathbb{R}$ on $\Omega$, the integral of $\phi$ with respect to the Poisson measure is given by:
\begin{equation}
\int_{\Omega} \phi \, dG_t = \sum_{i=1}^{N_t} \phi(x_i)
\end{equation}
We assume that the Poisson measure $G_t$ can be written $G_t = t\tilde{G}_t$ where the random measure $\tilde{G}_t$ converges weakly almost surely as $t \to \infty$ to a measure with the density $y$.

In the following, in order to use the Kullback-Leibler distance, we will define some random intensity $y_t \in L_1(\Omega)$ which is a good approximation of the mean intensity $y$ as $t \to \infty$. The random measure $\tilde{G}_t$ can be considered as a generalized stochastic process. Generalized random processes are the probabilistic version of the distributions of L.Schwartz [41, 27, 23]. We recall that $D(\Omega)$ is the space of infinitely differentiable functions with compact support in $\Omega$ and $D'(\Omega)$ the space of linear continuous functionals on $D(\Omega)$ with the corresponding weak convergence.

The random distribution $\tilde{G}_t \in D'(\Omega)$ can be regularized and approached as limit of random regular functions obtained by truncation and regularization. We set $n = int(t)$, where $int$ is the integer part of $t$, and we define for $n \geq 1$, the compact set:
\begin{equation}
K_n = \{x \in \Omega, \text{dist}(x, \partial \Omega) \geq 1/n, |x| \leq n\}
\end{equation}
For all $n \geq 1$, there exists a function $\chi_n \in C_c^\infty(\Omega)$ such that
\begin{equation}
0 \leq \chi_n \leq 1 \quad \chi_n|K_n = 1 \quad \text{supp}(\chi_n) \subset K_n + B(0, 1/2n)
\end{equation}
On the other hand, there is a function $\xi \in C_c^\infty(\mathbb{R}^2)$ such that:
\begin{equation}
\text{supp}(\xi) \subset B(0, 1), \quad \xi \geq 0, \quad \int_{\mathbb{R}^2} \xi(x) dx = 1
\end{equation}
We consider the function $\xi_n(x) = (4n)^2 \xi(4nx)$ and $y_t = T_n(t) = (\chi_n(t) \tilde{G}_t) * \xi_n(t)$. It can be shown that $\text{supp}(T_n) \subset K_{4n}$, $y_t \in C_c^\infty(\Omega)$ almost surely, where $C_c^\infty(\Omega)$ be the set of $C^\infty(\Omega)$ functions with compact support in $\Omega$. The function $y_t$ defines a positive distribution and a positive measure and we have:
\begin{equation}
\int_{\Omega} \phi y_t \, dw \rightarrow \int_{\Omega} \phi \, d\tilde{G}_t
\end{equation}
for any $\phi \in D(\Omega)$, as $t \to \infty$ almost surely [41, 23, 25].

**Definition 3.1.** Let $\mu$ a positive measure and $A \in \Omega$ such that $\mu(A) = 0$ almost surely, then $A$ is called a stochastic $\mu$ continuity set.

We detail a proposition which is very similar to the classical ‘Portmanteau theorem’ for positive measures.

**Proposition 3.2.** Let $\mu$ a positive distribution, $(\mu_n)_{n \in \mathbb{N}}$ a sequence of positive random distributions on $\Omega$. Then if $\mu_n \to \mu$ in $D'(\Omega)$ almost surely, then:
(1) $\limsup_n \mu_n(F) \leq \mu(F)$ for all closed set $F \subset \Omega$ almost surely.
(2) $\liminf_n \mu_n(G) \geq \mu(G)$ for all open set $G \subset \Omega$ almost surely.
(3) $\lim \mu_n(A) = \mu(A)$ almost surely for all stochastic $\mu$ continuity sets $A$. 

Inverse Problems and Imaging Volume 12, No. 3 (2018), 607–634
Proof. Let us assume \( \mu_n \to \mu \) in \( D'(\Omega) \) almost surely. Let \( F \) be a closed set and \( F_k = \{ x \in \Omega, d(x, F) \leq 1/k \} \). It is possible to find functions \( f_k \) in \( D(\Omega) \) such that \( 1_F \leq f_k \leq 1_{F_k} \) [41, 23, 25] and we have:

\[
\limsup \mu_n(F) = \limsup_n \int 1_F d\mu_n \leq \lim_n \int f_k d\mu_n = \int f_k d\mu \leq \mu(F_k)
\]

for all \( k \geq 0 \) almost surely. As \( k \to \infty \), we obtain (1). By taking the complements we obtain (2). Let \( A \) be a stochastic continuity set of the random measure \( \mu \), \( \text{Int}(A) \) its interior and \( \overline{A} \) its closure, we have almost surely:

\[
\mu(\text{Int}(A)) \leq \liminf_n \mu_n(\text{Int}(A)) \leq \liminf_n \mu_n(A) \leq \limsup_n \mu_n(A) \leq \limsup_n \mu_n(\overline{A}) \leq \mu(\overline{A})
\]

Since \( \mu(\partial \overline{A}) = \mu(\overline{A}) - \mu(\text{Int}(A)) = 0 \), these inequalities shows that \( \lim \mu_n(A) = \mu(A) \) almost surely for all stochastic continuity sets \( A \).

Assuming that \( \Omega \) is a continuity set for the mean measure of \( \tilde{G}_t \), we obtain:

\[
\lim_{t \to \infty} \int \Omega y_t dx = \int \Omega y dx
\]

almost surely. We have thus constructed an approximation \( y_t \) of \( y \) which belongs to the space \( C_c^\infty(\Omega) \subset L_1(\Omega) \) almost surely. The average of \( y_t \) is given by:

\[
E[y_t] = \xi_{n(t)} * y \chi_{n(t)}
\]

As \( t \to \infty \), \( E[y_t] \to y \) almost everywhere.

3.2. Regularization functionals, discretization operators. We assume that the unique minimum norm solution \( x^* \) of the inverse problem without noise belongs to a bounded, closed and convex subset \( B \in L_1(\Omega) \subset L_2(\Omega) \). We will consider the reduced regularization functional defined as:

\[
J_{\alpha, y}(x) = KL(y, Ax) + \alpha \| (x - x_0) \|_1
\]

where \( x_0 \in L_1(\Omega) \) and \( \alpha \) is a regularization parameter. We assume that the operator \( A : B \to L_1(\Omega) \) is linear and continuous.

We consider also a sequence of finite dimensional subspaces \( \{ Y_n \}_{n \geq 0} \) such that

\[
Y_0 \subset Y_n \subset Y_{n+1} \subset ... \subset \bigcup Y_n = L_2(\Omega)
\]

We assume we have continuous approximation operators \( \Pi'_n : B \to B \cap Y_n \) with:

\[
\| (I - \Pi'_n)(x) \|_2 \leq \rho'_n
\]

for all \( x \in B \) with \( \rho'_n \to 0 \) as \( n \to \infty \). For the sake of the simplicity, we will assume in the following that the operator \( \Pi'_n \) is linear.

For some disjoint regions \( \{ A_i \}_{1 \leq i \leq m} \subset \Omega \), we consider the discrete approximation \( \Pi_m \Pi'_n G_i \) of the Poisson process \( G_i \) defined by:

\[
x \to \sum_{1 \leq i \leq m} G_i(x)/|A_i|
\]

where \( G_i(x) = G_i(A_i), I(x, A) = \delta_0(A) = 1_A(x) \) defines the indicator function of the set \( A \) and \( |A_i| \) is the Lebesgue measure of the set \( A_i \). The domain \( \Omega \) can be understood as a countable collection of boxes and \( G_i(x) \) as the number of particles.
in the box $A_i$. For each $i$, $G'_i(\omega)$ is a Poisson random variable. Similarly, we define the discrete approximation $\Pi_m G_t$ of $G_t$ and we have:

\begin{equation}
\int_{\Omega} \Pi_m G_t dx = \sum_{1 \leq i \leq m} G_t(A_i)
\end{equation}

We will also use the notation $g_{t,m} = (g'_i)_{1 \leq i \leq m}$ for the Poisson variables $(\tilde{G}_t(A_i))_{1 \leq i \leq m}$ and $(g^i)_{0 \leq i \leq m}$ for their expectation values. We can thus assume there is a constant $\rho_m$ depending on the measure of $|A - \cup_{1 \leq i \leq m} A_i|$, with $\rho_m \to 0$ as $m \to \infty$ such that almost surely:

\begin{equation}
|\tilde{G}_t(\Omega) - \int_{\Omega} \Pi_m G_t dx| \leq \rho_m/2
\end{equation}

This term corresponds to the discretization error of the Poisson process. The same discretization operator can be defined for $f \in L_1(\Omega)$:

\begin{equation}
\Pi_m f = \sum_{i=1}^{m} I(x, A_i) \int_{A_i} f(x) dx / |A_i|
\end{equation}

This projection operator is a linear and bounded operator. Given the smooth approximation $y_t$ of $G_t$, one can consider $\Pi_m y_t$ and the discrete values $y_{t,m} = (y'_i)_{1 \leq i \leq m} = (\int_{A_i} y_t(x) dx)_{0 \leq i \leq m}$.

For given discretization levels $m$ and $n$ and $x' \in B \cap Y_n$, the discrete regularization functional is given by:

\begin{equation}
J_{\alpha,y_t,m,n}(x) = \frac{1}{K} L(\Pi_m y_t, \Pi_m Ax) + \alpha \|(x - \Pi_n x_0)\|_1
\end{equation}

the function $x \in B \cap Y_t$ can be written as $x = \Pi'_n x'$, with $x' \in B$. We will denote $x'_t$ the minimizer of the regularization functional $J_{\alpha,y_t}$ and $\Pi'_{n,x'_t,m,n}$ the one of $J_{\alpha,y_t,m,n}$. The proofs for the existence of these solutions are detailed in [36, 34]. They are based on the weak compactness with respect to the $L_1(\Omega)$ norm of the sublevels of the regularization functionals. We will assume that $y$ and $y_t$ are bounded below and above by two strictly positive constants $y_{\min}$ and $y_{\max}$. This assumption is also reasonable for $y_t$ since it is a smoothed version of the Poisson process. It implies that enough data have been collected with a long enough sampling time. Similar assumptions are done in [36, 34] to ensure that the regularization method with the Kullback-Leibler divergence as discrepancy term is well-defined. To avoid divergencies, it is also possible to use the regularized Kullback-Leibler distance defined in Section 2.1 for the data term. In the following, we denote $R(x) = \|x - x_0\|_1$, $R_{\min} = \min\{R(x), x \in L_1(\Omega)\}$ and $X_{\min} = \{x \in L_1(\Omega), R(x) = R_{\min}\}$. For the regularization functional considered, $R_{\min} = 0$ and $X_{\min}$ is the set of constant functions equal to $x_0$ almost everywhere. We denote $D_R^\alpha$ the Bregman distance associated to $R$ evaluated at $x^\star$.

With this setting the Kullback-Leibler distance $KL(y_t, Ax'^\alpha_t)$ between the function $y_t$ and the image of the reconstructed solution $x'^\alpha_t$ is a random variable. In the next subsections, we show that this discrepancy term and its finite dimensional approximation satisfies some bounds almost surely as $t \to \infty$ so that the Bertero principle is well-defined. In section 4, we derive convergence rates for the expectation value, $E[KL(\Pi_m y_t, \Pi_m A\Pi_n x'_t,m,n)]$ and $E[D_R^\alpha(\Pi'_n, x'_t,m,n, \Pi_n x^\star)]$. 

In [36, 34] to ensure that the regularization method with the Kullback-Leibler divergence as discrepancy term is well-defined. To avoid divergencies, it is also possible to use the regularized Kullback-Leibler distance defined in Section 2.1 for the data term. In the following, we denote $R(x) = \|x - x_0\|_1$, $R_{\min} = \min\{R(x), x \in L_1(\Omega)\}$ and $X_{\min} = \{x \in L_1(\Omega), R(x) = R_{\min}\}$. For the regularization functional considered, $R_{\min} = 0$ and $X_{\min}$ is the set of constant functions equal to $x_0$ almost everywhere. We denote $D_R^\alpha$ the Bregman distance associated to $R$ evaluated at $x^\star$.
3.3. Morozov principle for Poisson noise. Considering these reduced variable \( y_t \), we will study the following discrepancy principle to chose the regularization parameter \( \alpha \) which is directly derived from the Bertero principle. It is defined starting from the discrete version of the inverse problem.

For given constants \( \tau_3 > \tau_2 > 1 \), and given time exposure \( t \) we will chose the discretization levels \( m(t) \) and \( n(t) \) and the regularization parameter \( \alpha(t) \) such that \( x_{\alpha,t,m,n} \) satisfies:

\[
\frac{\tau_2 m(t)}{2t} \leq KL(\Pi_m y_t, \Pi_m A \Pi_n' x_{\alpha,t,m,n}) \leq \frac{\tau_3 m(t)}{2t}
\]

with \( \frac{m(t)}{t} \to 0 \) as \( t \to \infty \).

We consider the set \( L \) of \( R \)-minimizing solutions for data without noise \( y \):

\[
L = \{ x^* \in L_1(\Omega), \ Ax^* = y, \text{ such that } R(x^*) \leq R(x) \forall x \in L_1(\Omega), \ Ax = y \}
\]

For \( \alpha > 0 \), we define the set of functions minimizing the functional \( J_{\alpha,y} \):

\[
M_{\alpha,y} = \{ x \in L_1(\Omega), J_{\alpha,y}(x) \leq J_{\alpha,y}(z) \text{ for all } z \in L_1(\Omega) \}
\]

and also the set of functions for which the functional \( J_{\alpha,y} \) is bounded by a constant \( c \):

\[
M_{\alpha,y}(c) = \{ x \in L_1(\Omega), J_{\alpha,y}(x) < c \}
\]

These sublevel sets \( M_{\alpha,y}(c) \) are weakly sequentially compact in \( L_1(\Omega) \) [36, 34].

We will also use several times the following lemma:

**Lemma 3.3.** The sublevel sets \( M_{\alpha,y,m,n}(c) = \{ z \in B \cap Y_n, J_{\alpha,y,m,n}(z) < c \} \) are weakly sequentially compact.

**Proof.** The set \( B \cap Y_n \) is closed and convex and thus weakly closed. The proposition 2.5 (iv) and weak continuity of KL and \( R \) with respect to the \( L_1(\Omega) \) norm conclude the proof.

In the following, we will assume \( (H_1) \) that \( X_{\alpha,y} \cap L = \emptyset \). This assumption means that there is no constant function solution of the inverse problem for non noisy data. Let \( Y_{\alpha,y} = AX_{\alpha,y} \) and \( y_0 = Ax_0 \):

\[
KL(y, Y_{\alpha,y}) = \inf_{x \in X_{\alpha,y}} KL(y, Ax) = KL(y, y_0)
\]

For \( \tau_3 > 1 \), we also assume \( (H_2) \) that the smoothed noisy data \( \Pi_m y_t \) satisfy for all \( m \) and \( n \) large enough:

\[
\frac{\tau_3 m(t)}{2t} < KL(\Pi_m y_t, \Pi_m A \Pi_n' x_0)
\]

This assumption is used to ensure the existence of a regularization parameter satisfying the discrepancy principle.

We also make the third assumption, \( (H_3) \):

\[
x \in M_{\alpha,y} \text{ and } x \in L \Rightarrow x \in X_{\alpha,y}
\]
This assumption means that a function $x \in L_1(\Omega)$ that minimizes the continuous regularization functional for the non noisy data and that is a minimum norm solution is a constant function. With assumption $(H_1)$, it follows $M_{\alpha,y} \cap L = \emptyset$.

The assumption $(H_1)$, $(H_2)$ and $(H_3)$ are similar to the ones used for Banach spaces [3, 4].

For $t$, $n$ and $m$ fixed, we define the following functions of $\alpha$:

$$E(\alpha) = KL(\Pi_m y_t, \Pi_m A\Pi_n x_{t,m,n}) + \alpha R(\Pi_n x_{t,m,n})$$

$$\Phi(\alpha) = KL(\Pi_m y_t, \Pi_m A\Pi_n x_{t,m,n})$$

$$\Psi(\alpha) = \alpha R(\Pi_n x_{t,m,n})$$

3.4. Well-posedness of the regularization method for the Morozov principle. We now show that the Morozov principle of Eq.31 is well-defined for $m$ and $n$ sufficiently large. In the first part of this subsection, we bound $KL(\Pi_m \tilde{G}_t, \Pi_m A\Pi_n x^*)$.

Let us consider the m-dimensional projections $\pi_m \tilde{G}_t$ of $\tilde{G}_t$ with components $\tilde{g}_{t,m} = (\tilde{g}_i^t)_{1 \leq i \leq m}$. The moments of a Poisson variable are characterized by Lemma 7.1 (Appendix). The following proposition bounds the deviation of $KL(g_{t,m}, g_m)$ from $\frac{m}{2t}$.

**Proposition 3.4.** Let $\tau_1 > 0$, there exists a constant $C$ such that:

$$P|KL(g_{t,m}, g_m) - \frac{m}{2t}| \geq \frac{(\tau_1 - 1)m}{2t} \leq C(\frac{1}{m(\tau_1 - 1)^2} + \frac{1}{t^2(\tau_1 - 1)^2})$$

**Proof.** With the Tchebychev inequality we obtain:

$$P(|KL(g_{m,t}, g_m) - \frac{m}{2t}| \geq \frac{(\tau_1 - 1)m}{2t}) \leq \frac{4t^2}{m^2(\tau_1 - 1)^2} \int |KL(g_{t,m}, g_m) - \frac{m}{2t}|^2 dP$$

$$\leq \frac{8t^2}{m^2(\tau_1 - 1)^2} \left| KL(g_{t,m}, g_m) - E[KL(g_{t,m}, g_m)] \right|^2 dP$$

$$+ \frac{8t^2}{m^2(\tau_1 - 1)^2} \int |E[KL(g_{t,m}, g_m)] - \frac{m}{2t}|^2 dP$$

With the definition of the Kullback-Leibler term, for each index $i$ with $1 \leq i \leq m$,

$$KL(g^i_t, g^i) = g^i_t \ln(1 + \frac{g^i_t - g^i}{g^i}) + g^i_t - g^i$$

we obtain by a series development, for each index $i$:

$$KL(g^i_t, g^i) = g^i_t \sum_{1 \leq n \leq \infty} (-1)^{n+1} \frac{1}{n} (\frac{g^i_t - g^i}{g^i})^n + g^i_t - g^i_t$$

$$= (g^i_t - g^i + g^i) \left( \frac{g^i - g^i}{g^i} \right)^n + \sum_{2 \leq n \leq \infty} (-1)^{n+1} \frac{1}{n} (\frac{g^i_t - g^i}{g^i})^n$$

$$\frac{(g^i - g^i)^2}{2g^i} + \sum_{2 \leq n \leq \infty} (-1)^{n+1} \frac{1}{n(n+1)} (\frac{g^i_t - g^i}{g^i})^{n+1}$$

with

$$E\left[ (g^i_t - g^i)^{n+1} \right] = \frac{1}{t} E\left[ (G^i_t - G^i)^{n+1} \right]$$
With lemma 7.1, we have
\[(50)\quad E[(G_i^i - G_i^i)^2] = G_i^i\]
and for \(n \geq 2\)
\[(51)\quad E[\frac{(g_i^i - g_i^i)^{n+1}}{(g_i^i)^n}] = O\left(\frac{1}{t^2}\right)\]
Thus
\[(52)\quad |E[KL(g_{t,m}, g_m)]| = \frac{m}{2t} \leq \frac{Cm}{t^2}\]
Let us consider the finite sum of the KL distances between the data with and without noise:
\[(53)\quad J = \sum_{i=1}^{m} KL(g_i^i, g_i^j)\]
The variables \((g_i^i)_{1 \leq i \leq m}\) are assumed to be independent Poisson variables and thus:
\[(54)\quad Var[J] = \sum_{i=1}^{m} Var[KL(g_i^i, g_i^j)]\]
For \(n \geq 1\), and for \(1 \leq i \leq m\):\[(55)\quad Var[\frac{1}{n(n+1)} \frac{(g_i^i - g_i^i)^{n+1}}{(g_i^i)^n}] = \frac{1}{(n(n+1))^2t^2} Var[\frac{(G_i^i - G_i^i)^{n+1}}{(G_i^i)^n}]\]
\[(56)\quad = \frac{1}{(n(n+1))^2t^2} (E[\frac{(G_i^i - G_i^i)^{2n+2}}{(G_i^i)^{2n}}] - E[\frac{(G_i^i - G_i^i)^{n+1}}{(G_i^i)^n}])^2\]
using the lemma 7.1, we obtain for \(n \geq 1\):
\[(57)\quad E[\frac{(G_i^i - G_i^i)^{2n+2}}{(G_i^i)^{2n}}] = O(1)\]
Moreover, for \(n \geq 2\)
\[(58)\quad E[\frac{(G_i^i - G_i^i)^{n+1}}{(G_i^i)^n}]^2 = O(1/t^2)\]
and for \(n = 1\)
\[(59)\quad E[\frac{(G_i^i - G_i^i)^{n+1}}{(G_i^i)^n}]^2 = O(1)\]
Therefore, there exists a constant \(C_1\) such that for each index \(i\):
\[(60)\quad Var[KL(g_i^i, g_i^j)] \leq \frac{C_1}{t^2}\]
and thus
\[(61)\quad Var[J] \leq \frac{mC_1}{t^2}\]
With Eq.44, we obtain there is a constant \(C\) such that:
\[(62)\quad P[|KL(g_{t,m}, g_m) - \frac{m}{2t}| \geq \frac{(\tau_1 - 1)m}{2t}] \leq C\left(\frac{1}{m(\tau_1 - 1)^2} + \frac{1}{t^2(\tau_1 - 1)^2}\right)\]
In the following, we will denote \(P_{m,t} = C\left(\frac{1}{m(\tau_1 - 1)^2} + \frac{1}{t^2(\tau_1 - 1)^2}\right)\).
Corollary 1. Let $\tau_1 > 0$, there exists a constant $C$ such that:

$$P[|KL(g_{t,m}, g_m) - E[KL(g_{t,m}, g_m)]| \geq (\tau_1 - 1)E[KL(g_{t,m}, g_m)]] \leq \frac{C}{m(\tau_1 - 1)^2}$$

with $E[KL(g_{t,m}, g_m)] = O(\frac{m}{\tau_1})$

We are now bounding the KL distance between $\Pi_m\tilde{G}_t$ and $\Pi_mA\Pi_n^*x^*$.

Proposition 3.5. Let $\rho'_n$ defined in Section 3.2, there is a positive constant $D$ such that

$$KL(\Pi_m\tilde{G}_t, \Pi_mA\Pi_n^*x^*) \leq D\rho'_n + \frac{\tau_1m}{2t}$$

with a probability higher than $\mathcal{P} = 1 - P_{m,t}$.

Proof. From the definition of the Kullback-Leibler distance, with $y = Ax^*$:

$$KL(\Pi_m\tilde{G}_t, \Pi_mA\Pi_n^*x^*) = KL(\Pi_m\tilde{G}_t, \Pi_my) + I_1 + I_2 + I_3$$

where $I_1, I_2, I_3$ are the integrals:

$$I_1 = \int_\Omega (g(\Pi_my) - g(\Pi_mA\Pi_n^*x^*))dx$$

$$I_2 = \int_\Omega (g'(\Pi_my) - g'(\Pi_mA\Pi_n^*x^*))\Pi_myt dx$$

$$I_3 = \int_\Omega [g'(\Pi_mA\Pi_n^*x^*)(\Pi_mA\Pi_n^*x^*)dx - g'(\Pi_my)(\Pi_my)]dx$$

One can assume that the derivative $g'(y)$ is bounded. The former integrals depend on the discretization level and they can be bounded from above by $\rho'_n$,

$$I_1 \leq C||\Pi_mAx^* - \Pi_mA\Pi_n^*x^*|| \leq C||\Pi_m||A||\rho'_n$$

Similarly, we obtain:

$$I_2 \leq C_2\rho'_n, I_3 \leq C_3\rho'_n$$

With proposition 3.4, it follows that there is a positive constant $D$ such that:

$$KL(\Pi_m\tilde{G}_t, \Pi_mA\Pi_n^*x^*) \leq D\rho'_n + \frac{\tau_1m}{2t}$$

with a probability higher than $\mathcal{P} = 1 - P_{m,t}$.

It should be noted that this result is also true for non linear lipschitzian operators. This bound can be generalized to the smoothed approximation $y_t$.

Proposition 3.6. Let $\rho'_n$ defined in Section 3.2, there are positive constants $D$ and $E$ such that

$$KL(\Pi_my_t, \Pi_mA\Pi_n^*x^*) \leq D\rho'_n + \frac{\tau_1m}{2t} + E\rho_n + ||\tilde{G}_t - y_t||_1$$

with a probability higher than $\mathcal{P} = 1 - P_{m,t}$. 

Inverse Problems and Imaging Volume 12, No. 3 (2018), 607–634
Proof. From the definition of the Kullback-Leibler distance, it follows:
\[
KL(\Pi_m y_t, \Pi_m A\Pi'_n x^*) = KL(\Pi_m \hat{G}_t, \Pi_m A\Pi'_n x^*)
\]
(73) \[+ \int_{\Omega} g(\Pi_m y_t) - g(\Pi_m \hat{G}_t) - g'(\Pi_m A\Pi'_n x^*)(\Pi_m y_t - \Pi_m \hat{G}_t) \, dx \]

Assuming that the derivative \(g'(\Pi_m A\Pi'_n x^*)\) is bounded, the integral term can be bounded as follows:
\[
\int_{\Omega} g(\Pi_m y_t) - g(\Pi_m \hat{G}_t) - g'(\Pi_m A\Pi'_n x^*)(\Pi_m y_t - \Pi_m \hat{G}_t) \, dx
\]
(74) \[\leq \|\Pi_m y_t - \Pi_m G_t\|_1 \]
(75) \[\leq \|\Pi_m y_t - \Pi_m \hat{G}_t\|_1 + \|\Pi_m \hat{G}_t - G_t\|_1 \]
(76) \[\leq 2\rho_m + \|\hat{G}_t - y_t\|_1
\]
with proposition 3.5, we obtain the result. \(\square\)

**Corollary 2.** Assume we can find \(\tau'_1 > \tau_1\) and we can choose the discretization levels \(m(t), n(t)\) and \(t\) such that \(\|\hat{G}_t - y_t\|_1 \to 0\) almost surely as \(t \to \infty\). Then:
\[
KL(\Pi_m y_t, \Pi_m A\Pi'_n x^*) \leq \frac{\tau'_1 m(t)}{2t}
\]
with a probability higher than \(P = 1 - P_{m(t)}, t\). This probability \(P\) converges to 1 as \(t \to \infty\).

**Proof.** It follows from proposition 7.7 (Appendix) that \(\|\hat{G}_t - y_t\|_1 \to 0\) almost surely as \(t \to \infty\) and thus \(m(t), n(t)\) and \(t\) can be chosen such that Eq. 77 holds. \(\square\)

The former conditions are satisfied for \(m(t) = t^\beta\) for \(\beta < 1\) and then for \(n(t)\) and \(t\) large enough. The value of \(n\) has to be set as large as possible. The interpretation of this corollary is that \(m, n, t\) have to be chosen such that the discretization errors and the approximation error of \(\hat{G}_t\) by \(y_t\) are negligible with respect to the error due to the Poisson noise statistics.

**Proposition 3.7.** For fixed constants \(\tau_3 > \tau_2 > \tau'_1 > 1\), and discretization levels \(m(t)\) and \(n(t)\) chosen according to corollary 2, there exists a regularization parameter \(\alpha(t)\) such that:
\[
\frac{\tau_2 m(t)}{2t} \leq KL(\Pi_m y_t, \Pi_m A\Pi'_n x^\alpha_{t,m,n}) \leq \frac{\tau_3 m(t)}{2t}
\]
with a probability higher than \(P\).

**Proof.** With proposition 7.3 (Appendix) and assumption \((H_2)\), we obtain that with a probability higher than \(P\) we have:
\[
\lim_{\alpha \to 0} KL(\Pi_m y_t, \Pi_m A\Pi'_n x^\alpha_{t,m,n}) = \inf_{u \in \mathcal{B}} KL(\Pi_m y_t, \Pi_m A\Pi'_n u)
\]
(79) \[\leq KL(\Pi_m y_t, \Pi_m A\Pi'_n x^*) \]
(80) \[\leq \frac{\tau_1 m(t)}{2t} \leq \frac{\tau_2 m(t)}{2t} < \frac{\tau_3 m(t)}{2t}
\]
(81) \[< KL(\Pi_m y_t, \Pi_m A\Pi'_n x_0) \]
(82) \[\leq \lim_{\alpha \to \infty} KL(\Pi_m y_t, \Pi_m A\Pi'_n x^\alpha_{t,m,n})
\]
(83) \[\leq \frac{\tau_3 m(t)}{2t}
\]
For a given data realization, the KL term is a continuous function of $\alpha$ (see Appendix) and thus there exists a value $\alpha$ satisfying the Morozov discrepancy principle. As $t \to \infty$, $m(t) \to \infty$, $P \to 1$ and the Bertero method of choice of the regularization parameter based on the Morozov principle is well-defined.

3.5. Convergence of the regularization method for the Morozov principle.
In this section, we show the convergence of the regularization method for the former a posteriori choice of the regularization parameter. The proof is based on the asymptotic properties of the smoothed approximation $y_t$ as $t \to \infty$. We will use the following propositions proved in the Appendix:

**Proposition 3.8.** Let us consider a sequence $\{t_n\}$ with $t_n \to \infty$. The $L_1$ norm $\|y_{t_n} - y\|_1 \to 0$ almost surely as $t_n \to \infty$.

**Proposition 3.9.** Let us consider a sequence $\{t_n\}$ with $t_n \to \infty$. There is a constant $\rho_m \to 0$ such that $\|y_{t_n} - \Pi_m G_{t_n}\|_1 \leq \rho_m$ almost surely as $t_n \to \infty$.

The following proposition is independent of the discretization levels.

**Proposition 3.10.** Let $A : L_1(\Omega) \to L_1(\Omega)$ a continuous operator and $\{t_n\}$ a sequence with $t_n \to \infty$. Let $\{x_n\}$ a sequence in $L_1(\Omega)$ such that there exist $x^* \in L$ with:

\[
\lim_{t_n \to \infty} KL(y, Ax_n) = 0
\]

\[
\limsup_{t_n \to \infty} R(x_n) \leq R(x^*)
\]

then $\{x_n\}$ is convergent to an element of $L$ and $\lim_{t_n \to \infty} R(x_n) = R(x^*)$ almost surely as $t_n \to \infty$.

**Proof.** Let $t_n \to \infty$, the sequence $KL(y, Ax_n)$ is bounded and thus, there exists $x \in L_1(\Omega)$, and a subsequence of $\{x_n\}$ also denoted $\{x_n\}$ weakly converging to $x$ in $L_1(\Omega)$.

With the continuity of the operator $A$ and the weak lower-semicontinuity of $KL$, we obtain:

\[
KL(y, Ax) \leq \liminf_{n \to \infty} KL(y, Ax_n)
\]

Since $KL(y, Ax_n) = KL(y_{t_n}, Ax_n) + \delta_n$ with $\delta_n = \int_\Omega g(y) - g(y_{t_n}) - g'(Ax_n)(y - y_{t_n}) \, dx$, we have:

\[
KL(y, Ax) \leq \liminf_{n \to \infty} KL(y_{t_n}, Ax_n) + \delta_n
\]

The derivative $g'$ is bounded on $[y_{\min}, y_{\max}]$ and thus we obtain that there exists a positive constant $C_1$ such that:

\[
|\delta_n| \leq C_1 \|y - y_{t_n}\|_1
\]

From proposition 3.8, we see that $\delta_n \to 0$ as $t_n \to \infty$ almost surely and $KL(y, Ax) = 0$. With the weak lower-semicontinuity of $R$ we obtain:

\[
R(x) \leq \liminf_{n \to \infty} R(x_n)
\]

\[
\leq \limsup_{n \to \infty} R(x_n)
\]

\[
\leq R(x^*)
\]
The limit \( x^* \) is a \( R \) minimizing solution and thus \( R(x^*) = R(x^*) \), and \( x \in L \). Therefore, we obtain \( R(x_n) \to R(x^*) \). With a subsequence argument, we see that the whole sequence converges weakly to a minimum \( R \) norm solution for the data without noise. \( \square \)

**Proposition 3.11.** Let us assume \( x^+ \in L \), and \( x^+ \in M_{\alpha,y} \), then \( x^+ \in X_{\min} \).

**Proof.** Let 0 < \( t < 1 \), and \( x^+ \in L \cap M_{\alpha,y} \):

\[(92) \quad \alpha R(x^+) \leq J_{\alpha,y}(x^+) \]
\[(93) \quad \leq J_{\alpha,y}((1-t)x^+ + tx_0) \]
\[(94) \quad \leq KL(y, A((1-t)x^+ + tx_0)) + \alpha R((1-t)x^+ + tx_0) \]

The regularization term \( R \) is convex and thus:

\[(95) \quad R((1-t)x^+ + tx_0) \leq (1-t)R(x^+) + tR(x_0) \leq (1-t)R(x^+) \]
and thus

\[(96) \quad \alpha t R(x^+) \leq \|A(x^+ + t(x_0 - x^+)) - A(x^+)|^2 \]

dividing by \( t \) and taking the limit as \( t \to 0 \), we obtain:

\[(97) \quad \alpha R(x^+) \leq 0 \]
and thus \( R(x^+) = 0 \), \( x^+ \in X_{\min} \). \( \square \)

**Proposition 3.12.** There are positive constants \( D_1 \) and \( D_2 \) such that for all \( m \) and \( n \), and for \( u \in B \):

\[(98) \quad KL(y_t, Au) \leq Kl(\Pi_m y_t, \Pi_m A\Pi_n' u) + D_1 \rho_m + D_2 \rho_n' \]

**Proof.** With the definition of the \( KL \) distance,

\[(99) \quad KL(y_t, Au) = Kl(\Pi_m y_t, \Pi_m A\Pi_n' u) + I_1 + I_2 + I_3 + I_4 + I_5 \]

\[(100) \quad I_1 = \int_\Omega (g(y_t) - g(\Pi_m y_t))dx \]
\[(101) \quad I_2 = \int_\Omega (g(Au) - g(\Pi_m A\Pi_n' u))dx \]
\[(102) \quad I_3 = \int_\Omega g'(Au)(y_t - \Pi_m y_t - Au + \Pi_m A\Pi_n' u)dx \]
\[(103) \quad I_4 = \int_\Omega \Pi_m y_t(g'(Au) - g'(\Pi_m A\Pi_n' u))dx \]
\[(104) \quad I_5 = \int_\Omega \Pi_m A\Pi_n u(g'(Au) - g'(\Pi_m A\Pi_n' u))dx \]

From the identity:\n
\[(105) \quad A - \Pi_m A\Pi_n' = A - \Pi_m A + \Pi_m A(I - \Pi_n') \]
it follows: \( |I_1| \leq C_1 \rho_m \), \( |I_2| \leq C_2 \rho_n' \), \( |I_i| \leq C_i \rho_m + C_i' \rho_n \) for \( i \geq 3 \). \( \square \)

**Proposition 3.13.** For all \( u \in Y_m \)

\[(106) \quad KL(\Pi_m y_t, u) \leq Kl(\Pi_m y, u) + \delta_t \]
with \( \delta_t \to 0 \) as \( t \to \infty \) almost surely.
Proof. From the definition of the Kullback-Leibler distance, it follows:

\[ KL(\Pi_m y_t, u) = KL(\Pi_m y, u) + I_1 + I_2 \]

\[ I_1 = \int_{\Omega} g(\Pi_m y_t) - g(\Pi_m y) dx \]

\[ I_2 = \int_{\Omega} g'(u)(\Pi_m y_t - \Pi_m y) dx \]

Thus we have:

\[ |I_1| \leq C \| y_t - y \|_1 \]

\[ |I_2| \leq C \| y_t - y \|_1 \]

The proposition then follows from proposition 3.8. \( \square \)

**Proposition 3.14.** The regularization parameter \( \alpha \) chosen according to the Morozov principle satisfy \( \alpha \to 0 \) almost surely as \( t \to \infty \).

Proof. Let \( t_k \to \infty \), \( m_k = m(t_k) \), \( n_k = n(t_k) \) and \( \alpha_k = \alpha(t_k) \) chosen according to the Morozov principle for the regularization functional \( J_{\alpha_k y, m_k, n_k} \). Assume there exist some constant \( \bar{\alpha} \) such that for some subsequence also denoted \( \{\alpha_k\} \), \( \alpha_k \geq \bar{\alpha} \), and denote \( x_k \) the corresponding minimizer of the regularization functional \( J_{\alpha_k y, m_k, n_k} \). Let us note \( \{u_k\} \) the sequence defined by:

\[ u_k = \arg\min_{f \in Y} \{KL(\Pi_{m_k} y_t, \Pi_{m_k} A f) + \alpha R(f)\} \]

With the Morozov principle and proposition 8.2:

\[ KL(\Pi_{m_k} y_t, \Pi_{m_k} A \Pi_{n_k} u_k) \leq KL(\Pi_{m_k} y_{n_k}, \Pi_{m_k} A \Pi_{n_k} x_k) \leq \tau_1 \frac{m(t_k)}{2 t_k} \]

almost surely and \( \frac{m(t_k)}{2 t_k} \to 0 \).

\[ \limsup_{k \to \infty} \alpha R(\Pi_{n_k} u_k) \leq \limsup_{k \to \infty} KL(\Pi_{m_k} y_t, \Pi_{m_k} A \Pi_{n_k} u_k) + \alpha R(\Pi_{n_k} u_k) \]

From proposition 3.12, we obtain :

\[ \lim_{k \to \infty} KL(y_t, A \Pi_{n_k} u_k) = 0 \]

and

\[ \limsup_{k \to \infty} \alpha R(\Pi_{n_k} u_k) \leq \alpha R(x^*) \]

Therefore, we can apply proposition 3.10 to \( \{\Pi_{m_k} u_k\} \) and we can assume that \( \{\Pi_{m_k} u_k\} \) is convergent to \( z \in L \). From the weak lower semi-continuity of the functional \( R \) and of the KL data term, we obtain, for any \( f \in L_1(\Omega) \)

\[ KL(y, Az) + \alpha R(z) \leq \liminf KL(\Pi_{m_k} y_{n_k}, \Pi_{m_k} A \Pi_{n_k} u_k) + \alpha R(\Pi_{n_k} u_k) \]

\[ \leq \liminf KL(\Pi_{m_k} y_{n_k}, \Pi_{m_k} A f) + \alpha R(f) \]

\[ \leq KL(y, Af) + \alpha R(f) \]

The last inequality is satisfied almost surely as \( k \to \infty \) using proposition 3.13. Thus \( z \in M_{\alpha, y} \). With assumption \((H_3)\), we have \( z \in L \cap X_{\min} \). We have assumed that \( L \cap X_{\min} = \emptyset \), and thus we have reached a contradiction and thus \( \alpha \to 0 \) as \( t \to \infty \) almost surely. \( \square \)
Proposition 3.15. Under the assumption of the preliminary section, the parameter $\alpha$ chosen according to the Morozov principle satisfies:

$$\frac{m(t)}{\alpha} \rightarrow 0 \text{ as } t \rightarrow \infty$$

**Proof.** Let $t > 0$, $m$ and $n$ and $\alpha$ chosen according to the Morozov principle, using the minimizing properties of $\Pi^*_n x^\alpha_{t,m,n}$, we obtain:

$$KL(\Pi^*_m y_t, \Pi^*_m A\Pi^*_n x^\alpha_{t,m,n}) + \alpha R(\Pi^*_n x^\alpha_{t,m,n})$$

(121) \[ \leq KL(\Pi^*_m y_t, \Pi^*_m A\Pi^*_n x^\alpha) + \alpha R(\Pi^*_n x^\alpha) \leq \frac{m(t)}{2t} + \alpha R(\Pi^*_n x^\alpha) \]

where $x^\alpha$ is solution of $Ax^\alpha = y$. The last inequality is satisfied almost surely as $t \rightarrow \infty$. This implies that:

$$KL(\Pi^*_m y_t, \Pi^*_m A\Pi^*_n x^\alpha_{t,m,n}) \leq \frac{m(t)}{2t} + \alpha R(\Pi^*_n x^\alpha)$$

(122) \[ \text{thus from the former proposition, we obtain:} \]

$$KL(\Pi^*_m y_t, \Pi^*_m A\Pi^*_n x^\alpha_{t,m,n}) \rightarrow 0, \text{ a.s as } t \rightarrow \infty$$

(123)

This implies also that

$$\frac{\tau(t)}{2t} + \alpha R(\Pi^*_n x^\alpha_{t,m,n}) \leq \frac{\tau(t)}{2t} + \alpha R_1(\Pi^*_n x^\alpha)$$

(124) \[ \text{as } t \rightarrow \infty, \text{ almost surely.} \]

$$0 \leq \frac{(\tau - \tau_1)m(t)}{2t} \leq \alpha(R_1(\Pi^*_n x^\alpha) - R_1(\Pi^*_n x^\alpha_{t,m,n}))$$

(125)

and thus:

$$\limsup_{t \rightarrow \infty} R(\Pi^*_n x^\alpha_{t,m,n}) \leq R(x^\alpha) \text{ a.s } t \rightarrow \infty$$

(126)

Therefore, with proposition 3.10, $\lim R(\Pi^*_n x^\alpha_{t,m,n}) = R(x^\alpha)$ as $t \rightarrow \infty$ almost surely, and thus with Eq.125, the parameter $\alpha$ chosen according to the Morozov principle satisfies the condition:

$$\frac{m(t)}{\alpha} \rightarrow 0$$

(127)

The regularization method with the regularization parameter chosen according to the Morozov discrepancy principle is a convergent regularization method:

**Proposition 3.16.** Let $t_k \rightarrow \infty$, $\alpha_{k}$ a sequence of parameters chosen according to the Morozov principle, and $\Pi^*_n x^\alpha_{t_k,m_k,n_k}$ minimizing the regularization functional $J_{y_{t_k}, x^\alpha_{t_k,m_k,n_k}}$. The sequence $(\Pi^*_n x^\alpha_{t_k,m_k,n_k})$ has a convergent subsequence with respect to the weak $L^1$ topology. A limit of each convergent subsequence is an $R$-minimizing solution.

**Proof.** With the definition of $x^\alpha_{t_k,m_k,n_k}$, it follows that:

$$KL(\Pi^*_m y_{t_k}, \Pi^*_m A\Pi^*_n x^\alpha_{t_k,m_k,n_k}) + \alpha_k R(\Pi^*_n x^\alpha_{t_k,m_k,n_k})$$

(128) \[ \leq KL(\Pi^*_m y_{t_k}, \Pi^*_m A\Pi^*_n x^\alpha) + \alpha_k R(\Pi^*_n x^\alpha) \]

where $x^\alpha \in L$. With proposition 3.13,

$$\lim_{k \rightarrow \infty} KL(\Pi^*_m y_{t_k}, \Pi^*_m A\Pi^*_n x^\alpha_{t_k,m_k,n_k}) \rightarrow 0 \text{ a.s } t_k \rightarrow \infty$$

(130)
It follows:

\begin{equation}
\limsup_{k \to \infty} R(\Pi_{m_k} x_{tk,m_k,n_k}^\alpha) \leq R(x^*) \text{ a.s } t_k \to \infty
\end{equation}

From proposition 3.12, we obtain:

\begin{equation}
\lim_{k \to \infty} KL(y_{tk}, \Pi_{m_k} x_{tk,m_k,n_k}^\alpha) \to 0 \text{ a.s } t_k \to \infty
\end{equation}

With proposition 3.10, we conclude the proof.

4. Convergence rate of the a posteriori method of choice of the parameter with variational source condition. In this section, we do a statistical convergence analysis for the inverse problem and study the error estimates for approximate solutions in the limit $t \to \infty$. We derive convergence rates for the expectation value, $E[KL(\Pi_{m_y} y, \Pi_{m} x)]$ and $E[KL(\Pi_{m} y, \Pi_{m} x)]$.

Several types of source conditions are considered in the literature to obtain some convergence rate [30]. We assume the following condition (C) holds: there is $\omega \in L^1(\Omega)$ such that $A\omega \in \partial R(x^*)$ with $\|\omega\|_1 < \infty$. Under the following assumption, a convergence rate can be obtained. We will use the following lemmas:

**Lemma 4.1.** Let $Y_t \sim \text{Poisson}(Y)$, and $h(Y_t)$ be such that $E[h(Y_t)]$ is well-defined, then for all $n \geq 1$:

\begin{equation}
E[h(Y_t) Y_t^n] = Y E[h(Y_t) Y_t^{n-1}] + \frac{d}{dt} E[h(Y_t) Y_t^{n-1}]
\end{equation}

The proof can be found in [18]:

**Lemma 4.2.** Let $Y_t \sim \text{Poisson}(Y)$, then $E[\ln(Y_t)] = \ln(Y) + O(1/t)$

**Proof.** With the Taylor development:

\begin{equation}
\ln(Y_t) = \ln(Y(1 + \frac{Y_t - Y}{Y})) = \ln(Y) + \ln(1 + \frac{Y_t - Y}{Y})
\end{equation}

\begin{equation}
= \ln(Y) + \sum_{1 \leq n \leq \infty} (-1)^{n+1} \frac{1}{n} (\frac{Y_t - Y}{Y})^n
\end{equation}

and lemma 7.1, it follows that:

\begin{equation}
E[\ln(Y_t)] = \ln(Y) + O(1/t)
\end{equation}

**Proposition 4.3.** Let assume the wave source condition (C) is satisfied and $\rho_m = O(\frac{m(t)}{t})$, then the following convergence rate is obtained:

\begin{equation}
E[KL(\Pi_{m_y} y, \Pi_{m} x)] = O((\frac{m(t)}{t})^{1/4})
\end{equation}

**Proof.** From the equality for the Bregman divergence, Eq.6, we have:

\begin{equation}
KL(\Pi_{m_y} y, \Pi_{m} A\Pi_{m} x) + KL(\Pi_{m} y, \Pi_{m_y} y) - KL(\Pi_{m} y, \Pi_{m} A\Pi_{m} x) =
\end{equation}

\begin{equation}
\int_{\Omega} (g'(\Pi_{m_y} y) - g'(\Pi_{m} A\Pi_{m} x))(\Pi_{m} y - \Pi_{m_y} y) dx
\end{equation}
and
\[ KL(\Pi_{m}y, \Pi_{m}y_{t}) + KL(\Pi_{m}y_{t}, \Pi_{m}y) = \]
\[ - \int_{\Omega} g'(\Pi_{m}y)(\Pi_{m}y - \Pi_{m}y_{t}) - g'(\Pi_{m}y_{t})(\Pi_{m}y_{t} - \Pi_{m}y)dx \]
\[ = \int_{\Omega} (g'(\Pi_{m}y_{t}) - g'(\Pi_{m}y))(\Pi_{m}y_{t} - \Pi_{m}y)dx \]
\[ (139) \]

thus
\[ KL(\Pi_{m}y, \Pi_{m}A\Pi'_{n}x_{t,m,n}^{\alpha}) = - KL(\Pi_{m}y, \Pi_{m}y) + KL(\Pi_{m}y_{t}, \Pi_{m}A\Pi'_{n}x_{t,m,n}^{\alpha}) \]
\[ - \int_{\Omega} (g'(\Pi_{m}y_{t}) - g'(\Pi_{m}A\Pi'_{n}x_{t,m,n}^{\alpha}))((\Pi_{m}y_{t} - \Pi_{m}y))dx \]
\[ + \int_{\Omega} (g'(\Pi_{m}y_{t}) - g'(\Pi_{m}y))(\Pi_{m}y_{t} - \Pi_{m}y)dx \]
\[ (140) \]

Taking the expectation we obtain:
\[ E[KL(\Pi_{m}y, \Pi_{m}A\Pi'_{n}x_{t,m,n}^{\alpha})] \leq -E[KL(\Pi_{m}y_{t}, \Pi_{m}y)] + \frac{\tau_{3}m(t)}{2t} \]
\[ + E[\int_{\Omega} (g'(\Pi_{m}y_{t}) - g'(\Pi_{m}A\Pi'_{n}x_{t,m,n}^{\alpha}))((\Pi_{m}y_{t} - \Pi_{m}y))dx] \]
\[ + E[\int_{\Omega} (g'(\Pi_{m}y_{t}) - g'(\Pi_{m}y))(\Pi_{m}y_{t} - \Pi_{m}y)dx] \]
\[ (141) \]

On the one hand, with Eq.\[51 \]
\[ | - E(KL(\Pi_{m}y_{t}, \Pi_{m}y)) + \frac{\tau_{3}m(t)}{2t} | \leq \frac{(\tau_{3} - 1)m(t)}{2t} + |E(KL(\Pi_{m}y_{t}, \Pi_{m}y)) - \frac{m(t)}{2t}| \]
\[ \leq \frac{(\tau_{3} - 1)m(t)}{2t} + \frac{Cm(t)}{t^2} \]
\[ (143) \]

From lemma 4.1 and 4.2 with \( n = 1 \), we obtain:
\[ E[\int_{\Omega} (g'(\Pi_{m}y_{t}) - g'(\Pi_{m}y))(\Pi_{m}y_{t} - \Pi_{m}y)dx] = \frac{1}{t}E[\int_{\Omega} (ln(\Pi_{m}Y_{t}) - ln(t))(\Pi_{m}Y_{t} - \Pi_{m}Y)dx] \]
\[ (145) \]
\[ = \frac{1}{t}E[\int_{\Omega} ln(\Pi_{m}Y_{t})(\Pi_{m}Y_{t} - \Pi_{m}Y)dx] \leq \frac{Cm(t)}{yt^2} \]
\[ (146) \]

For each Poisson variable, the second derivative \( g''(Y_{t}^i) = 1/Y_{t}^i \) is bounded from above by a constant \( C_{1} \). We can also assume \( \|y_{t} - \Pi_{m}A\Pi'_{n}(x_{t,m}^{\alpha})\|_{\infty} < C \), where \( C \) is a constant independent of \( m \) and \( n \).

Thus with the Cauchy-Schwarz inequality we obtain:
\[ E[\int_{\Omega} (g'(\Pi_{m}y_{t}) - g'(\Pi_{m}A\Pi'_{n}x_{t,m,n}^{\alpha}))(\Pi_{m}y_{t} - \Pi_{m}y)dx] \leq C_{1}E[\int_{\Omega} (\Pi_{m}y_{t} - \Pi_{m}A\Pi'_{n}x_{t,m,n}^{\alpha}))(\Pi_{m}y_{t} - \Pi_{m}y)dx] \]
\[ \leq C_{1}E[\int_{\Omega} (\Pi_{m}y_{t} - \Pi_{m}A\Pi'_{n}x_{t,m,n}^{\alpha})^2dx]^{1/2}\int_{\Omega} (\Pi_{m}y_{t} - \Pi_{m}y)^2dx \]
\[ (147) \]
Using Eq.155, we get:

\begin{align}
(150) & \quad \leq 2C_1 E\left[ \int_{\Omega} (\Pi_m y_t - \Pi_m A \Pi_n' x_{t,m,n}^\alpha)^2 dx \right] + 2C_1 E\left[ \int_{\Omega} (\Pi_m y_t - \Pi_m y)^2 dx \right] \\
(151) & \quad \leq C_3 E[||\Pi_m y_t - \Pi_m A \Pi_n' (x_{t,m,n}^\alpha)||_1] + C_4 \frac{m}{t}
\end{align}

where the last inequality is obtained with Eq.42 Using the $L_1$ bounding property of the Kullback-Leibler distance (proposition 2.5) and the Mororov principle we obtain:

\begin{align}
(152) & \quad E\left[ \int_{\Omega} g'(\Pi_m y_t) - g'(\Pi_m A \Pi_n' x_{t,m,n}^\alpha). (\Pi_m y_t - \Pi_m y) dx \right] \\
(153) & \quad \leq C_1 E[KL(\Pi_m y_t, \Pi_m A \Pi_n' x_{t,m,n}^\alpha)] + \frac{mC_4}{t} \\
(154) & \quad \leq C_6 \left( \frac{m(t)}{t} \right)^{1/2} + \frac{C_4 m}{t}
\end{align}

With Eq.143, 146 and 154 we obtain:

\begin{align}
(155) & \quad E[KL(\Pi_m y, \Pi_m A \Pi_n' (x_{t,m,n}^\alpha))] = O\left( \left( \frac{m(t)}{t} \right)^{1/2} \right)
\end{align}

From the definition of $x_{t}^\alpha$, we have almost surely as $t \to \infty$ :

\begin{align}
(156) & \quad KL(\Pi_m y_t, \Pi_m A \Pi_n' x_{t,m,n}^\alpha) + \alpha R(\Pi_n' x_{t,m,n}^\alpha) \\
& \quad \leq KL(\Pi_m y_t, \Pi_m A \Pi_n' x^*_{t,m,n}) + \alpha R(\Pi_n' x^*) \\
& \quad \leq KL(\Pi_m y_t, \Pi_m A \Pi_n' x_{t,m,n}^\alpha) + \alpha R(\Pi_n' x_{t,m,n}^\alpha)
\end{align}

with $x^* \in L$ and

\begin{align}
(157) & \quad \tau_2 \frac{m(t)}{2t} + \alpha R(\Pi_n' x_{t,m,n}^\alpha) \leq KL(\Pi_m y_t, \Pi_m A \Pi_n' x_{t,m,n}^\alpha) + \alpha R(\Pi_n' x_{t,m,n}^\alpha)
\end{align}

and thus it follows:

\begin{align}
(158) & \quad R(\Pi_n' x_{t,m,n}^\alpha) \leq R(\Pi_n' x^*)
\end{align}

this inequality is satisfied almost surely as $t \to \infty$. We now assume that the source condition holds, then we obtain:

\begin{align}
(159) & \quad D_{R}^{*} (\Pi_n' x_{t,m,n}^\alpha, \Pi_n' x^*) \leq R(\Pi_n' x_{t,m,n}^\alpha) - R(\Pi_n' x^*) \\
(160) & \quad - < A^* \omega, \Pi_n' x_{t,m,n}^\alpha - \Pi_n x^* > \\
(161) & \quad \leq - < A^* \omega, \Pi_n' x_{t,m,n}^\alpha - \Pi_n x^* >
\end{align}

With $A^* \omega = A^* \omega - A^* \Pi_m \omega + A^* \Pi_m \omega$ and $||A^* \omega - A^* \Pi_m \omega|| \leq \rho_m$, it follows:

\begin{align}
(162) & \quad D_{R}^{*} (\Pi_n' x_{t,m,n}^\alpha, \Pi_n' x^*) \leq - < \omega, \Pi_m A(\Pi_n' x_{t,m,n}^\alpha) - \Pi_m y > + C \rho_m \\
(163) & \quad \leq ||\omega||_2 ||\Pi_m A(\Pi_n' x_{t,m,n}^\alpha) - \Pi_m y||_2 + C \rho_m \\
(164) & \quad \leq C_1 ||\Pi_m A(\Pi_n' x_{t,m,n}^\alpha) - \Pi_m y||_1 + C \rho_m \\
(165) & \quad \leq C_2 KL(\Pi_m A \Pi_n' x_{t,m,n}^\alpha, \Pi_m y)^{1/2} + C \rho_m
\end{align}

Using Eq.155, we get:

\begin{align}
(166) & \quad E[D_{R}^{*} (\Pi_n' x_{t,m,n}^\alpha, \Pi_n' x^*)] = O(\left( \frac{m(t)}{t} \right)^{1/4})
\end{align}
5. A numerical experiment: Spectral computerized tomography. Multi-energy Computerized Tomography, or spectral CT, is an imaging modality where the information received depends on the energy of each photon hitting the detector. This information is provided by a photon counting detector which gives an image for each energy bin. These projections allow us to separate the contribution of the different materials (e.g. soft tissues and bone) to the projections [30, 40, 19, 42]. The measured projections are corrupted by Poisson noise. For weak non-linearities, the spectral CT problem sets a linear inverse problem that has to be regularized to obtain stable results for material decomposition. In the first subsection, we present the forward model relating the projected mass of the material and the number of photons detected and the regularization functional. In the second section, we detail some numerical experiments.

5.1. Forward model for spectral CT. In this section, we present the forward model for spectral CT based on the attenuation model. We denote Ω the space that contains the 3-dimensional (3-D) object studied and \( x \) the voxel position. This object is imaged on a 2-D detector (\( S \)). The number of photon transmitted at energy \( E \) for each pixel \( u \) of the detector is denoted \( n(E,u) \). The attenuation follows a Beer-Lambert law:

\[
(167) \quad n(E,u) = n^0(E) \exp \left( - \int_{\mathcal{L}(u)} \mu(E,x) dx \right)
\]

where \( n^0(E) \) is source spectrum, \( \mathcal{L}(u) \) is line defining by the X-ray beam and \( \mu(E,x) \) is the linear attenuation coefficient at energy \( E \) for the voxel \( x \). The number of photons detected by the detector can be written as:

\[
(168) \quad s(E,u) = \int_{\mathcal{L}(u)} d(E,E)n(E,u) dE
\]

where \( d(E,E) \) is the detector response function. It describes the probability that a photon with the energy \( E \) is detected at the energy \( E \). An integrated circuit counts the number of photons that are detected in the \( i \)th energy bin \([E_i, E_{i+1}]\). The number of photons detected in the \( i \)th is:

\[
(169) \quad s_i(u) = \int_{\mathcal{L}(u)} d_i(E)n(E,u) dE
\]

where

\[
(170) \quad d_i(E) = \int_{E_i}^{E_{i+1}} d(E,E) dE
\]

is the response function of the \( i \)th energy bin of the detector. We assume we can write the attenuation coefficient as a sum of \( M \) basis functions that are separable in energy and space:

\[
(171) \quad \mu(E,x) = \sum_{m=1}^{M} \rho_m(x) \tau_m(E)
\]

where \( \rho_m(x) \) is the mass of the material \( m \) at the voxel \( x \) and \( \tau_m(E) \) is a well-defined function describing the attenuation effects in the material \( m \). Using (167), (171) and (169) we obtain the following relation:

\[
(172) \quad s_i(u) = \int_{\mathcal{L}(u)} d_i(E)n^0(E) \exp \left( - \sum_{m=1}^{M} a_m(u) \tau_m(E) \right) dE
\]
where
\begin{equation}
(173) \quad a_m = \int_{\mathcal{L}(u)} \rho_m(x) dx
\end{equation}
is the projection of all the mass of the material \(m\) along the X-ray path \(\mathcal{L}(u)\). For weak non linearities, the exponential can be replaced by a linear approximation. We denote as \(s = \mathcal{F}(a)\) the corresponding forward model relating the projections of the masses of the materials \((a_m)\) an the number of photons detected in the various energy bins, \((s_i)\).

After discretization, the inverse problem considered is a finite dimensional one. We consider our detector has \(P\) pixels and \(I\) energy bins, the projections data are thus the vector \(s \in \mathbb{R}^{IP} = (s_{ij})_{1 \leq i \leq I, 1 \leq j \leq P}\). We want to recover the set of the projected masses of the materials \((a)\) an the number of photons detected in the various energy bins, \((s_i)\). After discretization, the inverse problem considered is a finite dimensional one. We consider our detector has \(P\) pixels and \(I\) energy bins, the projections data are thus the vector \(s \in \mathbb{R}^{IP} = (s_{ij})_{1 \leq i \leq I, 1 \leq j \leq P}\). We want to recover the set of the projected masses of the materials \((a)\) an the number of photons detected in the various energy bins, \((s_i)\).

5.2. Numerical results. In the following, we will test the Morozov principle for different discretization levels. We can not be sure that the ground truth satisfies the source condition \((C)\). But is is possible to illustrate qualitatively the former results. In order to minimize the regularization functional, we have used Expectation-Minimization type iterative methods [16, 15, 22, 45]. In each case investigated, we will study the decrease of the Kullback-Leibler functional as a function of the iterations.

The algorithm will be tested on a mouse phantom with two materials, bone and soft tissues. The number of incident photons is \(N = 10^3\). The reconstruction algorithm will be tested with \(M = 2\) materials and with various numbers of energy bins \(I\), and various number of pixels \(P\). Let \(a^*\) be the ground truth projected mass map and \(\hat{a}\) the estimated projected mass map. In order to evaluate the quality of the reconstruction, the relative mean square reconstruction error is calculated with
\begin{equation}
(174) \quad J(s, \mathcal{F}(a)) = KL(s, \mathcal{F}(a)) + \alpha \mathcal{R}(a)
\end{equation}
\(\alpha\) the regularization parameter and \(\mathcal{R}(a)\) the regularization term to stabilize the solution. For the numerical simulations, the \(L_1\) norm is replaced by a Huber type penalty functional.

The first case investigated corresponds to \(I = 3\) and \(P = 6820\) pixels. The evolution of the Kullback-Leibler distance with the iterations is displayed on Figure 1 and the reconstructed image on Figure 2. In this case, the discretization level is low and the discretization error \(\rho'_n\) can not be neglected with respect to the error related to the Poisson noise. The corollary 2 shows that the Morozov principle is not well-defined. At the end of the iterations the RMSE error between \(a^*\) and \(\hat{a}\) is 9.75. With this low discretization level, the reconstruction error is very large.

The second case investigated corresponds to a high discretization level \(I = 10\) and \(P = 1737984\) pixels (Figure 3 and Figure 4). In this case, the discretization errors can be neglected and it is possible to chose a regularization parameter \(\alpha\) that fulfills the Morozov principle. A very small regularization parameter \(\alpha = 10^{-5}\) can be chosen. At the end of the iterations, the RMSE error is \(RMSE(a^*, \hat{a}) = 2.39\). With this high discretization level, the reconstruction error is still large. On many pixels, a low number of photons is detected and the Morozov principle is not an efficient way to choose the regularization parameter. According to proposition 4.3,
if the source condition (C) is satisfied, the average reconstruction error scales as $(\frac{m(t)}{t})^{1/4}$ and it is large for large values of $m$.

**Figure 1.** Evolution of the Kullback-Leibler functional for $I = 3$ and $P = 6820$. The value $m/2$ is displayed for comparison.

**Figure 2.** Maps of the reconstructed bone and soft tissues projected mass for $I = 3$ and $P = 6820$.

**Figure 3.** Evolution of the Kullback-Leibler functional for $I = 10$ and $P = 1737984$. The value $m/2$ is displayed for comparison.

**Figure 4.** Maps of the reconstructed bone and soft tissues projected mass for $I = 10$ and $P = 1737984$. 

The last case studied corresponds to the parameters $I = 3$ and $P = 108624$. The evolution of the Kullback-Leibler distance with the iteration and the reconstructed images are displayed in Figure 5 and 6. In this case, the regularization parameter is 6.56. The final reconstruction error is $RMSE(a^*, \hat{a}) = 0.78$. This choice of the discretization levels is the best one. The discretization error $\rho'_n$ can be neglected and the smaller discretization level $m$ leads to a smaller reconstruction errors.

Figure 5. Evolution of the Kullback-Leibler functional for $I = 3$ and $P = 108624$. The value $m/2$ is displayed for comparison.

Figure 6. Maps of the reconstructed bone and soft tissues projected mass for $I = 3$ and $P = 108624$.

6. **Conclusion.** In this work, we have studied a method of choice of the regularization parameter for Poisson inverse problems with the Kullback-Leibler divergence as data term in a finite dimensional setting based on the Morozov-Bertero principle. We have shown that this method of choice is well-defined when the number of Poisson variables is large and when the discretization error on the unknown can be neglected. The convergence of the regularization method with this a posteriori choice is also demonstrated when the number of random variables is large. Assuming some source conditions, convergence rates in expectation are obtained for the Bregman distance between the true solution and the reconstructed solution. Some numerical experiments show that the choice of the discretization levels is crucial to obtain good reconstruction images.

7. **Appendix.** In this appendix, we detail the proof of some propositions used in the paper. The following lemma characterizes the moments of the Poisson distribution [18]:

**Lemma 7.1.** Let $X \sim \text{Poisson}(\lambda)$, a one dimensional Poisson variable then for any $n \geq 1$

$$E_\lambda (X - \lambda)^n = \sum_{k=0}^{n} a_{k,n} \lambda^n \quad (175)$$
where
\begin{equation}
(a_{k,n} = \sum_{i=0}^{k} (-1)^{i} \binom{n}{i} S_{2}(n - i, k - i)
\end{equation}
where for any \( p, q \geq 0 \), \( S(p, q) \) denote the Stirling second number defined as the number of partitions of a set of \( p \) elements into \( q \) nonempty disjoint subsets.

For any \( n \geq 1 \), the leading coefficient \( a_{n,n} = 0 \).
For any \( n \geq 3 \), \( a_{n-1,n} = 0 \).

The following proposition deals with some monotonicity properties and limit properties of \( E, \Phi \) and \( \Psi \).

**Proposition 7.2.** For \( y \) fixed, and \( 0 < \alpha < \alpha' \), \( m, n \) fixed
\begin{equation}
KL(\Pi_{m} y, \Pi_{m} A_{\alpha}^{n} x_{t,m,n}^{\alpha}) \leq KL(\Pi_{m} y, \Pi_{m} A_{\alpha'}^{n} x_{t,m,n}^{\alpha'})
\end{equation}
\begin{equation}
\lim_{\alpha \to 0} KL(\Pi_{m} y, \Pi_{m} A_{\alpha}^{n} x_{t,m,n}^{\alpha}) = \inf_{x \in Y} KL(\Pi_{m} y, \Pi_{m} A_{\alpha}^{n} x)
\end{equation}
\begin{equation}
\lim_{\alpha \to 0} KL(\Pi_{m} y, \Pi_{m} A_{\alpha}^{n} x_{t,m,n}^{\alpha}) = \inf_{u \in X_{m,n}} KL(\Pi_{m} y, \Pi_{m} A_{\alpha}^{n} u)
\end{equation}
\begin{equation}
0 \leq \liminf_{\alpha \to \infty} R(x_{t,m,n}^{\alpha}) \leq \limsup_{\alpha \to \infty} R(x_{t,m,n}^{\alpha}) \leq KL(\Pi_{m} y, \Pi_{m} A x_{0}) \leq 0
\end{equation}
and thus we obtain (c)

We have \( x_{t,m,n}^{\alpha} \in M_{\alpha, y, m, n}(x_{0}) \) for \( c \) large enough for all \( \alpha \). For given \( \alpha \), from the compact property mentioned in section 3.3, we can thus construct some sequence \( (\alpha_{k}) \to \infty \), \( \alpha_{k} \geq \alpha \), such that \( x_{t,m,n}^{\alpha_{k}} \) converges weakly towards \( x_{n} \in Y_{n} \) as \( k \to \infty \).

With the weak lower semicontinuity properties of \( R \) we obtain:
\begin{equation}
R(x_{t,m,n}^{\alpha}) \leq \liminf_{\alpha_{k} \to \infty} R(x_{t,m,n}^{\alpha_{k}}) = 0
\end{equation}
thus \( x_{t,m,n}^{\alpha} \) is equal to \( x_{t,m,n}^{\alpha} \). From the lower semicontinuity of the \( KL \) functional with respect to the weak \( L_{1} \) topology (Proposition 2.5) we obtain:
\begin{equation}
KL(\Pi_{m} y, \Pi_{m} A_{\alpha}^{n} x_{0}) \leq KL(\Pi_{m} y, \Pi_{m} A x_{0})
\end{equation}
\begin{equation}
\leq \liminf_{k \to \infty} KL(\Pi_{m} y, \Pi_{m} A_{\alpha}^{n} x_{t,m,n}^{\alpha_{k}})
\end{equation}
\begin{equation}
= \lim_{\alpha \to \infty} KL(\Pi_{m} y, \Pi_{m} A_{\alpha}^{n} x_{t,m,n}^{\alpha})
\end{equation}
The last limit exists because the discrepancy term is a monotonous function of \( \alpha \) as stated in a). On the other hand,
\begin{equation}
\lim_{\alpha \to \infty} KL(\Pi_{m} y, \Pi_{m} A_{\alpha}^{n} x_{t,m,n}^{\alpha}) \leq J_{\alpha, y, m, n}(x_{t,m,n}^{\alpha})
\end{equation}
\begin{equation}
\leq KL(\Pi_{m} y, \Pi_{m} A x_{0})
\end{equation}
With Eq. 187 and Eq. 189, we obtain proposition d). \qed

**Proposition 7.3.** For a given realization of the data \( y_t \), the functions \( \Phi, \Psi \) and \( E \) are continuous functions of \( \alpha \).

**Proof.** Let \( \{ \alpha_k \} \) a positive sequence with \( \alpha_k \to \alpha > 0 \). For each values \( \alpha_k \), we consider the minimizer, \( \Pi_{n,t,m,n}^{x_0^k} \) in \( Y_n \). Then \( \sup_{k \geq 0} J_{\alpha,y_t,m,n}(\Pi_{n,t,m,n}^{x_0^k}) < \infty \) and there is a convergent subsequence again denoted \( \Pi_{n,t,m,n}^{x_0^k} \) converging towards \( \Pi_n^{x_0^k} \). With the continuity of the operator \( A \) and with the weak lower-semicontinuity of KL and \( R \) we obtain:

\[
\text{(190)} \quad J_{\alpha,y_t,m,n}(\Pi_n^{x_0^k}) \leq \liminf_{k \to \infty} KL(\Pi_{m,t}^{x_0^k}, \Pi_{m,t}^\alpha) + \alpha_k R(\Pi_{n,t,m,n}^{x_0^k})
\]

\[
\leq \limsup_{k \to \infty} KL(\Pi_{m,t}^{x_0^k}, \Pi_{m,t}^\alpha) + \alpha_k R(\Pi_{n,t,m,n}^{x_0^k})
\]

\[
\leq \lim_{k \to \infty} KL(\Pi_{m,t}^{x_0^k}, \Pi_{m,t}^\alpha) + \alpha_k R(\Pi_{n,t,m,n}^{x_0^k})
\]

\[
\leq KL(\Pi_{m,t}^{x_0^k}, \Pi_{m,t}^\alpha) + \alpha_k R(\Pi_{n,t,m,n}^{x_0^k})
\]

for all \( x \in \mathcal{E} \). And thus \( \Pi_n^{x_0^k} = \Pi_n^{x_0^k,t,m,n} \) and \( E(\alpha_n) \to E(\alpha) \). By a subsequence argument, the same result is true for the whole sequence.

We now show the continuity of \( \Psi \) and of the data term \( \Phi \) with respect to \( \alpha \). Let us assume that \( KL(\Pi_{m,t}^{x_0^k}, \Pi_{m,t}^\alpha) \) does not converge to \( KL(\Pi_{m,t}^{x_0^k}, \Pi_{m,t}^\alpha) \). With the weak lower-semicontinuity of KL, we obtain:

\[
\text{(191)} \quad KL(\Pi_{m,t}^{x_0^k}, \Pi_{m,t}^\alpha) < \limsup_{k \to \infty} KL(\Pi_{m,t}^{x_0^k}, \Pi_{m,t}^\alpha) = c < \infty
\]

With the compactness property mentioned in section 3.3, we can consider a subsequence denoted \( \Pi_{n,t,m,n}^{x_0^k} \) such that \( KL(\Pi_{m,t}^{x_0^k}, \Pi_{m,t}^\alpha) \to c \).

Then:

\[
\text{(192)} \quad \Psi(\alpha_k) = E(\alpha_k) - KL(\Pi_{m,t}^{x_0^k}, \Pi_{m,t}^\alpha)
\]

and thus:

\[
\text{(193)} \quad \liminf_{n \to \infty} \Psi(\alpha_k) = KL(\Pi_{m,t}^{x_0^k}, \Pi_{m,t}^\alpha) - c + \alpha_k R(\Pi_{n,t,m,n}^{x_0^k})
\]

\[
< \alpha_k R(\Pi_{n,t,m,n}^{x_0^k})
\]

which contradicts the weak lower semicontinuity of \( \Psi \), and it follows that \( \Psi \) and \( \Phi \) are continuous functions of \( \alpha \).

In the next propositions, we prove that \( \| y_t - y \|_1 \) and \( \| y_t - \Pi_{m,t}^{x_0^k} \|_1 \) are converging towards zero as \( t \to \infty \). We will use the concentration inequality [37]:

**Proposition 7.4.** For any measurable function \( f \), one has:

\[
\text{(194)} \quad \forall u > 0, P(\int \Omega f(x)(dG_t - Ydx) \geq \sqrt{2u} \int \Omega f(x)^2 Ydx + \frac{1}{\sqrt{3}} \| f \|_{\infty} u) \leq \exp(-u)
\]

**Proposition 7.5.** Let us consider a sequence \( \{ t_n \} \), with \( t_n \to \infty \). \( \int \Omega (dG_{t_n} - ydx) \to 0 \) almost surely as \( t_n \to \infty \).

**Proof.** From the former proposition, it follows that for \( u > 0 \), there is a positive constant \( C_1 \) such that:

\[
\text{(195)} \quad P(\int \Omega (dG_t - ydx) \geq \sqrt{2u} \int \Omega C_1 u + \frac{u}{\sqrt{3}}) \leq \exp(-u)
\]
Let us consider a sequence \( \{t_n\} \), with \( t_n \to \infty \). We can deduce that for all \( u \geq 4/9C_1 \):

\[
P(| \int_\Omega (d\tilde{G}_{t_n} - y)dx | \geq \frac{u}{t_n}) \leq \exp(-u)
\]

For all \( \epsilon > 0 \), and \( u = \epsilon t_n \) it follows:

\[
\sum_n P(| \int_\Omega (d\tilde{G}_{t_n} - y)dx | > \epsilon) < \infty
\]

We can thus apply the Borel-Cantelli theorem and \( \int_\Omega (d\tilde{G}_{t_n} - y)dx \to 0 \) almost surely.

**Proposition 7.6.** Let us consider a sequence \( \{t_n\} \) with \( t_n \to \infty \). The \( L_1 \) norm \( \|y_{t_n} - y\|_1 \to 0 \) almost surely as \( t_n \to \infty \).

**Proof.** We have obtained in section 3.1 that:

\[
\lim_{t_n \to \infty} \int_\Omega y_{t_n} dx = \int_\Omega y dx
\]

almost surely. Since \( y_{t_n} \to y \) almost everywhere and almost surely as \( t_n \to \infty \), we obtain from the Lebesgue dominated convergence theorem that \( \|y_{t_n} - y\|_1 \to 0 \) as \( t_n \to \infty \) almost surely.

**Proposition 7.7.** Let us consider a sequence \( \{t_n\} \) with \( t_n \to \infty \). There is a constant \( \rho_m \to 0 \) such that \( \|y_{t_n} - \Pi_m \tilde{G}_{t_n}\|_1 \leq \rho_m \) almost surely as \( t_n \to \infty \).

**Proof.** Similarly we have:

\[
\|y_{t_n} - \Pi_m \tilde{G}_{t_n}\|_1 \leq 2| \int_\Omega y_{t_n} dx - \tilde{G}_{t_n}(\Omega)| + 2|\tilde{G}_{t_n}(\Omega) - \int_\Omega \Pi_m \tilde{G}_{t_n} dx|
\]

and we have

\[
| \int_\Omega y_{t_n} dx - \tilde{G}_{t_n}(\Omega)| \leq | \int_\Omega y dx - \tilde{G}_{t_n}(\Omega)| + | \int_\Omega y dx - \int_\Omega y_{t_n} dx|
\]

The first term in Eq.206 tends to zero as \( t_n \to \infty \) almost with Proposition 7.5. The second term in Eq.205 tends also towards zero as \( t_n \to \infty \) almost surely with Proposition 7.6. With Eq.25, we can assume that \( t_n \) is chosen such that almost surely:

\[
\|y_{t_n} - \Pi_m \tilde{G}_{t_n}\|_1 \leq \rho_m
\]

Acknowledgment. This work was supported by the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program “Investissements d’Avenir” (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

REFERENCES

[1] V. Albani, A. De Cezaro and J. P. Zubelli, On the choice of the Tikhonov regularization parameter and the discretization level: A discrepancy-based strategy, *Inverse Problems and Imaging*, 10 (2016), 1–25.

[2] A. Antoniadis and J. Bigot, Poisson inverse problems, *Ann. Stat.*, 34 (2006), 2132–2158.

[3] S. Anzengruber and R. Ramlau, Morozov discrepancy principle for Tikhonov-type functionals with nonlinear operators, *Inverse Problems*, 26 (2010), 025001, 17pp.

[4] S. Anzengruber and R. Ramlau, Convergence rates for Morozov’s discrepancy principle using variational inequalities, *Inverse Problems*, 27 (2010), 105007, 18pp.
A. Banerjee, S. Merugu, I. Dhilon and J. Ghosh, Clustering with Bregman divergences, *Journal of Machine Learning Research*, 6 (2005), 1705–1749.

J. M. Bardsley and J. Goldes, Regularization parameter selection methods for ill-posed Poisson maximum likelihood estimation, *Inverse Problems*, 25 (2009), 095005, 18pp.

J. M. Bardsley, A theoretical framework for the regularization of Poisson Likelihood estimation problems, *Inverse Problems and Imaging*, 4 (2010), 11–17.

0. Barndorff-Nielsen, *Information and Exponential Families in Statistical Theory*, Wiley Publishers, 1978.

M. Bertero, P. Boccacci, G. Talenti, R. Zanella and L. Zanni, A discrepancy principle for Poisson data, *Inverse Problems*, 26 (2010), 105004, 20 pp.

G. L. Besnerais, J. F. Bercher and G. Demoment, A new look at entropy for solving linear inverse problems, *IEEE Transactions on Information Theory*, 45 (1999), 1566–1578.

T. Bonesky, Morozov’s discrepancy principle and Tikhonov-type functionals, *Inverse Problems*, 25 (2009), 015015, 11pp.

J. M. Borwein and A. S. Lewis, Convergences of best entropy estimates, *SIAM Journal on Optimization*, 1 (1991), 191–205.

L. M. Bregman, The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming, *USSR Computational Mathematics and Mathematical Physics*, 7 (1967), 200–217.

M. Burger and S. Osher, Convergence rates of convex variational regularization, *Inverse Problems*, 20 (2004), 1411–1421.

C. Byrne, *Iterative Optimization in Inverse Problems*, Taylor and Francis, 2014.

C. Byrne and P. Eggermont, EM algorithms, *Handbook of Mathematical Methods in Imaging*, Vol. 1, 2, 3, 305–388, Springer, New York, 2015.

L. Cavalier and J. Y. Koo, Poisson intensity estimation for tomographic data using a wavelet shrinkage approach, *IEEE Trans. on Information Theory*, 48 (2002), 2794–2802.

A. Das Gupta, *On a Differential Equation and one Step Recursion for Poisson Moments*, Purdue University, Technical Report 98-02.

N. Ducros, J. F. P. Abascal, B. Sixou, S. Rit and F. Peyrin, Regularization of nonlinear decomposition of spectral X-ray projection images, *Med. Phys.*, 44 (2017), 157–1576.

H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers: Dordrecht, 1996.

P. Favati, G. Lotti, O. Menchi and F. Romani, Performance analysis of maximum likelihood methods for regularization problems with nonnegativity constraints, *Inverse Problems*, 26 (2010), 085013, 18pp.

I. M. Gelfand and G. E. Shilov, *Generalized Functions*, Academic Press, San Diego, 1964.

R. M. Gray, *Probability, Random Processes and Ergodic Properties*, Second edition. Springer, Dordrecht, 2009.

G. Grubb, *Distributions and Operators*, Springer-Verlag, New York, 2009.

T. Hohage and F. Werner, Inverse problems with Poisson data: Statistical regularization theory, applications and algorithms, *Inverse Problems*, 32 (2016), 093001, 56pp.

K. Ito, *Fundations of Stochastic Differential Equations in Infinite Dimensional Spaces*, SIAM, 1984.

J. Y. Koo and W. C. Kim, Inhomogeneous Poisson intensity estimation via information projections onto wavelet subspaces, *Journal of Korean Statistical Society*, 31 (2002), 343–357.

S. Kullback and R. A. Leibler, *On information and sufficiency*, *Annals of Mathematical Statistics*, 22 (1951), 79–86.

Y. Long and J. A. Fessler, Multi-material decomposition using statistical image reconstruction for spectral CT, *Trans. Med. Imaging*, 33 (2014), 1614–1626.

R. D. Nowak and E. D. Kolaczyk, A Bayesian multiscale framework for Poisson inverse problems, *IEEE Trans. on Information Theory*, 46 (2000), 1811–1825.

M. H. Neumann, Absolute regularity and ergodicity of Poisson count processes, *Bernoulli*, 17 (2011), 1268–1284.

J. M. Ollinger and J. A. Fessler, Positon-emission tomography, *IEEE Signal Processing Magazine*, 14 (1997), 43–55.

C. Pöschl, *Tikhonov Regularization with General Residual Term*, PhD Thesis, Univeristat Innsbruck, 2008.
[35] E. Resmerita and R. S. Anderssen, Joint additive Kullback-Leibler residual minimization and regularization for linear inverse problem, Mathematical Methods in the Applied sciences, 30 (2007), 1527–1544.
[36] E. Resmerita, Regularization of ill-posed inverse problems in Banach spaces: Convergence rates, Inverse Problems, 21 (2005), 1301–1314.
[37] P. Reynaud-Bouret, Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities, Probab. Theory Relat. Fields, 126 (2003), 103–153.
[38] R. J. Santos and A. R. De Pierro, A new parameters choice method for ill-posed problem with Poisson data and its application to emission tomographic imaging, International Journal of Tomography and Statistics, 11 (2009), 33–52.
[39] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmaier and F. Lenzen, Variational Methods in Imaging, Springer Verlag, New York, 2009.
[40] J. P. Schlomka, E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. Bumer, C. Herrmann, R. Steadman, G. Zeitler, A. Livne and R. Preska, Experimental feasibility of multi-energy photon-counting k-edge imaging in pre-clinical computed tomography, Physics in Medicine and Biology, 53 (2008), p4031.
[41] L. Schwartz, Théorie des Distributions, Hermann, Paris, 1966.
[42] Spray toolbox, https://www.creatis.insa-lyon.fr/~ducros/WebPage/spray.html
[43] J. L. Stark and F. Murtagh, Astronomical Image and Data Analysis, Springer Verlag, New York, 2006.
[44] A. N. Tikhonov and V. Y. Arsenin, Solutions to Ill-Posed Problems, Winston-Wiley, New York, 1977.
[45] H. Wang and P. C. Miller, Scaled heavy-ball acceleration of the Richardson-Lucy algorithm for 3D microscopy image restoration, IEEE Transactions on Image Processing, 23 (2014), 848–854.
[46] F. Werner and T. Hohage, Convergence rate in expectation of Tikhonov-type regularization of inverse problems with Poisson data, Inverse Problems, 28 (2012), 104004, 15pp.
[47] F. Werner, Inverse problems with Poisson data: Tikhonov-type regularization and Iteratively Regularized Newton Methods, PhD thesis, University of Göttingen, 2012.
[48] R. Zanella, P. Boccacci, L. Zani and M. Bertero, Efficient gradient projection methods for edge-preserving removal of poisson noise, Inverse Problems, 25 (2009), 045010, 24pp.

Received June 2017; revised November 2017.

E-mail address: bruno.sixou@insa-lyon.fr
E-mail address: Tom.Hohweiller@creatis.insa-lyon.fr
E-mail address: nicolas.ducros@insa-lyon.fr