Tomographic reconstruction from Poisson distributed data: a fast and convergent EM-TV dual approach

Voichita Maxim, Yuemeng Feng, Hussein Banjak, Elie Bretin

Tomographic reconstruction from Poisson distributed data: a fast and convergent EM-TV dual approach

Voichiţa Maxim[1], **Yuemeng Feng**[1], **Hussein Banjak**[1], **Elie Bretin**[2]

(1) Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69621, Villeurbanne, France
(2) Univ Lyon, INSA de Lyon, CNRS UMR 5208, Institut Camille Jordan, 20 avenue Albert Einstein, F-69621 Villeurbanne Cedex, France

E-mail: maxim@creatis.insa-lyon.fr, elie.bretin@insa-lyon.fr

Abstract.

This paper focuses on tomographic reconstruction for nuclear medicine imaging, where the classical approach consists to maximize the likelihood of Poisson distributed data using the iterative Expectation Maximization algorithm. In this context and when the quantity of acquired data is low and produces low signal-to-noise ratio in the images, a step forward consists to incorporate a total variation a priori on the solution into a MAP-EM formulation. The novelty of this paper is to propose a convergent and efficient numerical scheme to compute the MAP-EM optimizer, based on a splitting approach which alternates an EM step and a dual-TV-minimization step. The main theoretical result is the proof of stability and convergence of this scheme. Moreover, we also present some numerical experiments in which our algorithm appears at least as efficient and accurate as some other reference algorithms from the literature.

Keywords: image reconstruction, total variation, Poisson noise, tomography, MAP-EM, MLEM, dual approach

1. Introduction

Emission tomography works by detecting radiation emitted from within the patient, enabling clinicians to identify the position and size of tumours, to control the quality of a treatment or to conduct diagnostic procedures such as coronary perfusion. Important issues relating to the invasive character of this type of imaging is the reduction of the tracer dosage and acquisition time. As a consequence, the clinical emission data uses to be strongly affected by Poisson noise. This noise propagates through the tomographic reconstruction process and leads to images with low signal-to-noise ratio. Early stopping of the iterations or post-smoothing reduces variance of the noise but also blurs the edges and may mask small sources. Moreover, it is known that early stopping of the iterations conducts to biased images and that the variance of the noise is intensity-dependent,

making it difficult to choose the smoothing kernel. Finally, stopping the iterations does not ensure that the reconstructed image will have the desired properties from the point of view of the user's a priori information, and post-smoothing is likely to reduce concordance between solution and data.

An appealing alternative is to include a priori information in the reconstruction process, a priori which ensures a best compromise between smoothness and agreement to the data. The likelihood being shift-variant, shift invariant priors give unsatisfactory results in terms of bias and resolution. Shift-variant quadratic priors were proposed *e.g.*, in [1] and [2] and were shown to give better results. More recently priors based on the total variation norm have been introduced but encountered strong computational difficulties due to the more complex and non-differentiable expression of the total variation norm. However the use of the total variation prior is appealing when one attempts to obtain images with smooth regions separated by sharp edges. The resolution of linear inverse problems with total variation regularization was tackled in the context of the Gaussian noise and ℓ_2-norm (see *e.g.*, [3]) but is still an issue when the data is Poisson distributed and thus a Kullback-Leibler distance is better suited. In this work we address the last topic, and we propose a solution in which the classical maximum likelihood expectation maximization algorithm is used in conjunction with a custom-designed total-variation regularization.

1.1. Principle of MLEM algorithm

The volume containing the source of gamma photons is divided in J spatial locations called pixels or voxels, indexed on $j = 1, \ldots, J$. The emission counts at the various locations are independent and the value x_j of the j^{th} pixel is a realization of a Poisson random variable with mean value λ_j. The values $\mathbf{x}$ cannot be observed directly with a detector placed outside the object. For a given pixel j and a given detector element i, t_{ij} is the probability that a photon emitted at pixel j will be detected at detector i. The system matrix of the detection system (also called transition matrix) is then $T = (t_{ij})$. The element s_j, where $j = 1, \ldots, J$, of the sensitivity vector $\mathbf{s}$ is defined as the probability of a photon emitted at pixel j to be detected somewhere in the camera. For each pixel j, we obviously have the relation:

$$s_j = \sum_{i=1}^{I} t_{ij}. \tag{1}$$

For $i = 1, \ldots, I$, the numbers y_i of events detected by the detector elements also follow Poisson distributions with mean values μ_i and the observations are independent. The vector $\mathbf{y}$ of detected events approximately matches $T\mathbf{x}$, more precisely the means verify the equation

$$\boldsymbol{\mu} = T\boldsymbol{\lambda}. \tag{2}$$

In [4] the non observable variable $\mathbf{x}$ is called complete data and the observed $\mathbf{y}$ are called incomplete data. The aim of the MLEM algorithm is to estimate the means $\boldsymbol{\lambda}$ of the complete data from the observed values $\mathbf{y}$ of the incomplete data.

Given the observations $\mathbf{y} = (y_i)_{i=1,\dots,I}$, an estimator of the vector $\boldsymbol{\lambda} = (\lambda_j)_{j=1,\dots,J}$ is the solution of the maximum of the log-likelihood function

$$\ell(\boldsymbol{\lambda}|\mathbf{y}) = -\sum_{i=1}^{I}\sum_{j=1}^{J} t_{ij}\lambda_j + \sum_{i=1}^{I} y_i \ln\left(\sum_{j=1}^{J} t_{ij}\lambda_j\right) - \sum_{i=1}^{I} \ln(y_i!). \tag{3}$$

The elements of the Hessian matrix $\mathscr{H}_\ell(\boldsymbol{\lambda}|\mathbf{y})$ of ℓ are

$$\frac{\partial^2 \ell}{\partial\lambda_j \partial\lambda_k}(\boldsymbol{\lambda}|\mathbf{y}) = -\sum_{i=1}^{I} \frac{t_{ij}t_{ik}y_i}{\mu_i^2} \tag{4}$$

and for all $\mathbf{u} \in \mathbb{R}^J$,

$$\mathbf{u}^T \mathscr{H}_\ell(\boldsymbol{\lambda}|\mathbf{y})\mathbf{u} = -\sum_{i=1}^{I}\left(\frac{\sqrt{y_i}}{\mu_i}\sum_{j=1}^{J} t_{ij}u_j\right)^2 \leq 0. \tag{5}$$

The concavity of the log-likelihood implies that its maximum is global. When $\text{rank}(T) = J$, the linear application of matrix T is injective and thus the Hessian is negative definite. The likelihood has then a strict maximum and an unique maximum solution. When $\text{rank}(T) < J$, the Hessian is negative semi-definite and there may be multiple solutions for the maximum.

In practice, an iterative method is called to approximate the minimum, and often the Expectation Maximization (EM) algorithm from [4] is used. One thus obtain the maximum likelihood expectation maximization algorithm that was proposed for nuclear imaging applications in [5, 6]. Starting from an arbitrary vector $\boldsymbol{\lambda}^{(0)} \in (\mathbb{R}_+^*)^J$, the $(n+1)^{th}$ estimation of the maximum likelihood is given by:

$$\lambda_j^{(n+1)} = \lambda_j^{(n)} \frac{1}{s_j}\sum_{i=1}^{I} t_{ij}\frac{y_i}{\sum\limits_{k=1}^{J} t_{ik}\lambda_k^{(n)}}, \tag{6}$$

and the EM algorithm always decreases the value of the likelihood function. The convergence of the algorithm to the maximum of the likelihood was proved in [5] for strictly concave likelihood function and was proved in the general case in [7].

Both simple and efficient, the MLEM is widely used and sometimes even when the data are not Poisson distributed. Its convergence speed is related to the system matrix. When the system matrix only accounts for the geometry of detection the convergence may be relatively fast and the result noisy. For this reason the iterations must be stopped before too much high frequencies are introduced. On the contrary, when the system matrix also models some convolution introduced by *e.g.*, the detectors or the attenuation in the patient, the convergence is slow and the higher frequencies are hardly recoverable.

Introduction of a priori information helps in both cases to obtain a smooth and precise image in a reasonable time. This a priori information can be added easily in the Bayesian formalism and leads to maximum a posteriori (MAP) algorithms. When the prior is a smooth function, the proof of convergence of the algorithm naturally fits the EM developments from [4]. This is not any more the case when the prior is not smooth. Moreover, numerical algorithms are needed to solve for the MAP estimator.

1.2. A priori information and discrete total variation

Total variation denoising is known to promote smoothness in images while still conserving sharp edges. It tends to produce almost homogeneous regions separated by sharp frontiers, which makes that the images have a cartoon aspect. This type of regularization has a strong interest in particular for low dose acquisitions where the goal is to identify the shape of the objects in the volume.

To facilitate the presentation and without loss of generality, all the developments hereafter are presented for the two-dimensional case. The three-dimensional case can then be easily derived and the results remain true with minor changes that we will specify when necessary.

Let Ω be an open subset of $\mathbb{R}^2$. The standard total variation is defined for functions $u \in L^1(\Omega)$ by

$$TV(u) = \sup\left\{ \int_\Omega u(x) \mathrm{div}\,(\varphi(x)) dx : \ \varphi \in C_c^1(\Omega;\mathbb{R}^2),\ |\varphi(x)| \leq 1\ \forall x \in \Omega \right\}. \tag{7}$$

In particular, it is well known ([8]) that the functional TV is finite if and only if the distributional derivative Du of u is a finite measure on Ω. Moreover, if u has a gradient $\nabla u \in L^1(\Omega)$, then $TV(u) = \int_\Omega |\nabla u(x)| dx$. An other interesting particular case is $u = \chi_Q$ the characteristic function of a smooth set Q, when $TV(u)$ can be identified to the perimeter of Q : $TV(u) = \int_{\partial Q} 1 d\sigma$.

From the discrete point of view, we assume that the image $\mathbf{u}$ is a 2-dimensional matrix of size $J = N \times N$ and we then denote by X the Euclidean space $\mathbb{R}^{N\times N} = \mathbb{R}^J$. The discrete total variation of $\mathbf{u}$ is then defined by

$$TV(\mathbf{u}) = \sum_{1\leq i,j\leq N} |(\nabla \mathbf{u})_{i,j}|, \tag{8}$$

where $|y| := \sqrt{y_1^2 + y_2^2}$ for all $y = (y_1, y_2) \in \mathbb{R}^2$. Here, the discrete gradient $\nabla \mathbf{u}$ is a vector in $Y = X \times X$ given by $(\nabla \mathbf{u})_{i,j} = ((\nabla \mathbf{u})^1_{i,j}, (\nabla \mathbf{u})^2_{i,j})$ with

$$(\nabla \mathbf{u})^1_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j}, & \text{if } i < N \\ 0, & \text{if } i = N \end{cases} \text{ and } (\nabla \mathbf{u})^2_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j}, & \text{if } j < N \\ 0, & \text{if } j = N \end{cases}.$$

Notice that an equivalent definition of $TV(\mathbf{u})$ is

$$TV(\mathbf{u}) = \sup\left\{ \langle \boldsymbol{\varphi}, \nabla \mathbf{u} \rangle_Y : \ \boldsymbol{\varphi} \in Y \text{ such that } |\varphi_{i,j}| \leq 1,\ i,j = 1,\ldots,N \right\} \tag{9}$$

where $\langle \mathbf{p}, \mathbf{q} \rangle_Y = \sum_{i,j} (p^1_{i,j} q^1_{i,j} + p^2_{i,j} q^2_{i,j})$. Finally, let us remark that $\langle \boldsymbol{\varphi}, \nabla \mathbf{u} \rangle_Y = -\langle \mathrm{div}\, \boldsymbol{\varphi}, \mathbf{u} \rangle_Y$ as soon as the discrete divergence $\mathrm{div}\, \boldsymbol{\varphi}$ is defined by

$$(\mathrm{div}\, \boldsymbol{\varphi})_{i,j} = \begin{cases} \varphi^1_{i,j} - \varphi^1_{i-1,j}, & \text{if } 1 < i < N \\ \varphi^1_{i,j}, & \text{if } i = 1 \\ -\varphi^1_{i-1,j}, & \text{if } i = N \end{cases} + \begin{cases} \varphi^2_{i,j} - \varphi^2_{i,j-1}, & \text{if } 1 < j < N \\ \varphi^2_{i,j}, & \text{if } j = 1 \\ -\varphi^2_{i,j-1}, & \text{if } j = N \end{cases}.$$

Let us also introduce the proximal operator $\mathrm{prox}_{\tau TV}[\mathbf{u}]$ defined by

$$\mathrm{prox}_{\tau TV}[\mathbf{u}] = (I + \tau \partial TV)^{-1}(\mathbf{u}) = \underset{\mathbf{v}\in X}{\arg\min} \left\{ \frac{1}{2\tau} \|\mathbf{u} - \mathbf{v}\|_X^2 + TV(\mathbf{v}) \right\},$$

where ∂TV is the sub-gradient of the TV function. Its value can be approximated for instance using the efficient dual approach proposed and analyzed by Chambolle in [8].

1.3. State-of-the art of numerical methods for Total Variation regularization for Poisson distributed data

In a recent work, Anthoine et al. [9] proposed some proximal methods for tomographic inverse problem with Poisson intensity and total variation regularization. Their idea consists to use some classical proximal methods to minimize an energy F in the form of:

$$F(\boldsymbol{\lambda}) = L(\boldsymbol{\lambda}) + G(\boldsymbol{\lambda}), \tag{10}$$

where L is the negative log-likelihood function (from which we withdraw the constant term) given in (3), that we re-express as:

$$L(\boldsymbol{\lambda}) = -\ell(\boldsymbol{\lambda}|\mathbf{y}) - \langle \log(\mathbf{y}!), \mathbf{1}\rangle = \langle T\boldsymbol{\lambda} - \mathbf{y}\log(T\boldsymbol{\lambda}), \mathbf{1}\rangle \tag{11}$$

and G is a total variational regularization term

$$G(\boldsymbol{\lambda}) = \alpha TV(\boldsymbol{\lambda}). \tag{12}$$

The operations on vectors from (11) are performed component-wise and the vector $\mathbf{1}$ is the vector of ones having the dimension of the data vector $\mathbf{y}$. A natural approach consists to apply the Chambolle-Pock [10] optimization algorithm that does not require the differentiability of the two operands. Denoting H the application given for all $\boldsymbol{\mu} \in (\mathbb{R}_+)^I$ by

$$H(\boldsymbol{\mu}) = \langle \boldsymbol{\mu} - \mathbf{y}\log(\boldsymbol{\mu}), \mathbf{1}\rangle, \tag{13}$$

that also verifies for all $\boldsymbol{\lambda}$ the equation $L(\boldsymbol{\lambda}) = H(T\boldsymbol{\lambda})$, the algorithm reads:

$$\left\{ \begin{array}{rcl} \boldsymbol{\mu}^{(n+1)} & = & (I + \sigma\partial H^*)^{-1}(\boldsymbol{\mu}^{(n)} + \sigma T\boldsymbol{\lambda}^{(n)}) \\ \overline{\boldsymbol{\lambda}}^{(n+1)} & = & (I + \tau\partial G)^{-1}(\overline{\boldsymbol{\lambda}}^{(n)} - \tau T^*\boldsymbol{\mu}^{(n+1)}) \\ \boldsymbol{\lambda}^{(n+1)} & = & 2\overline{\boldsymbol{\lambda}}^{(n+1)} - \overline{\boldsymbol{\lambda}}^{(n)} \end{array} \right. , \tag{14}$$

and the sequence $(\boldsymbol{\lambda}^{(n)})$ converges to a minimum of F when the descent steps τ and σ verify $\tau\sigma\|T\|^2 < 1$. As F is well defined for $\boldsymbol{\lambda} > 0$, addition of a positivity constraint is also examined in [9]. Even if algorithms as (14) give interesting reconstructions, our approach is different in the sense that we want to exploit the idea and the good behaviour of the MLEM algorithm on highly noisy data.

With the same purpose, Sawatzky et al. [11] proposed to successively apply:

- The (EM) step from (6), which can be expressed in matrix form as:

$$\boldsymbol{\lambda}^{(n+1/2)} = \frac{\boldsymbol{\lambda}^{(n)}}{T^*\mathbf{1}} T^* \left[\frac{\mathbf{y}}{T\boldsymbol{\lambda}^{(n)}}\right], \tag{15}$$

- A weighted TV step

$$\boldsymbol{\lambda}^{(n+1)} = \boldsymbol{\lambda}^{(n+1/2)} - \boldsymbol{\omega}^{(n)}\partial TV(\boldsymbol{\lambda}^{(n+1)}), \quad \text{with } \boldsymbol{\omega}^{(n)} = \alpha\frac{\boldsymbol{\lambda}^{(n)}}{T^*\mathbf{1}}. \tag{16}$$

Indeed, the gradient of $\ell(\boldsymbol{\lambda}|\mathbf{y})$ satisfies

$$\nabla_{\boldsymbol{\lambda}}\ell(\boldsymbol{\lambda}|\mathbf{y}) = T^*\mathbf{1} - T^*\left[\frac{\mathbf{y}}{T\boldsymbol{\lambda}}\right] \tag{17}$$

and the Euler equation associated to the minimization of $L + G$ reads as

$$0 \in T^*\mathbf{1} - T^*\left[\frac{\mathbf{y}}{T\lambda}\right] + \alpha\partial TV(\boldsymbol{\lambda}). \tag{18}$$

The inclusion of the positivity constraint as Karush-Kuhn-Tucker conditions then gives:

$$\boldsymbol{\lambda} \in \frac{\boldsymbol{\lambda}}{T^*\mathbf{1}}T^*\left[\frac{\mathbf{y}}{T\boldsymbol{\lambda}}\right] - \alpha\frac{\boldsymbol{\lambda}}{T^*\mathbf{1}}\partial TV(\boldsymbol{\lambda}). \tag{19}$$

According to this equality, Sawatzky et al. consider the iterative scheme

$$\boldsymbol{\lambda}^{(n+1)} = \frac{\boldsymbol{\lambda}^{(n)}}{T^*\mathbf{1}}T^*\left[\frac{\mathbf{y}}{T\boldsymbol{\lambda}^{(n)}}\right] - \alpha\frac{\boldsymbol{\lambda}^{(n)}}{T^*\mathbf{1}}\partial TV(\boldsymbol{\lambda}^{n+1}), \tag{20}$$

which appears to be equivalent to the previous weighted-TV one. Additionally, the authors explain how the weighted-TV step can be solved by minimizing the functional

$$\begin{aligned}\boldsymbol{\lambda}^{(n+1)} &= \operatorname{prox}_{\boldsymbol{\omega}^{(n)}TV}(\boldsymbol{\lambda}^{(n+1/2)}) \\ &= \arg\min_{\boldsymbol{\lambda}}\left\{\frac{1}{2}\|\sqrt{\boldsymbol{\omega}^{(n)}}(\boldsymbol{\lambda} - \boldsymbol{\lambda}^{(n+1/2)})\|^2 + TV(\boldsymbol{\lambda})\right\},\end{aligned} \tag{21}$$

using a slightly modified version of the Chambolle's algorithm ([8]) where $\boldsymbol{\lambda}^{(n+1)} = \boldsymbol{\lambda}^{(n+1/2)} - \boldsymbol{\omega}^{(n)}\operatorname{div}\boldsymbol{\varphi}$ and $\boldsymbol{\varphi}$ is calculated iteratively as

$$\boldsymbol{\varphi}^{(k+1)} = \frac{\boldsymbol{\varphi}^{(k)} + \tau\nabla\left(\boldsymbol{\omega}^{(n)}\operatorname{div}\boldsymbol{\varphi}^{(k)} - \boldsymbol{\lambda}^{(n+1/2)}\right)}{1 + \tau|\nabla\left(\boldsymbol{\omega}^{(n)}\operatorname{div}\boldsymbol{\varphi}^{(k)} - \boldsymbol{\lambda}^{(n+1/2)}\right)|}, \tag{22}$$

with the descent step τ verifying $0 < \tau < \dfrac{1}{4\|\boldsymbol{\omega}^{(n)}\|_\infty}$. To simplify the presentation we extended the notation $|\cdot|$ to vectors $\boldsymbol{\varphi} \in Y$, where $|\boldsymbol{\varphi}|$ is the vector of X with components $|\varphi_j|$, $j = 1, \ldots, J$.

As we will see later, even if some numerical experiments show that $(\boldsymbol{\lambda}^{(n)})_{n\geq 0}$ seems to decrease the value of H, it is not clear from a theoretical point of view if this is true or not and if the convergence can be asymptotically attained.

1.4. Outline of the paper

Our motivation in this work is to introduce an algorithm which is similar to the one from Sawatzky et al., but using a variational point of view that allows to establish the convergence of the sequence of iterates to a minimizer of F. More precisely, we show in section 2 that an iterative sequence $(\boldsymbol{\lambda}^{(n)})_{n\geq 0}$ defined by

$$\boldsymbol{\lambda}^{(n+1)} \in \arg\min_{\mathbf{u}}\{\langle\mathbf{u}, \mathbf{s}\rangle - \langle\log(\mathbf{u}), \mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}\rangle + G(\mathbf{u})\}, \tag{23}$$

and $\boldsymbol{\lambda}^{(0)} > 0$ convergences to a minimizer of F. Here, as in Sawatzky et al., $\boldsymbol{\lambda}^{(n+1/2)}$ is the result of the EM step (15) and we note $\mathbf{s} = T^*\mathbf{1}$ which is assumed to satisfy

$$\min_j s_j = s_{\min} > 0. \tag{24}$$

In section 3, we focus on the numerical computation of a solution of (23). In particular, we introduce a descent algorithm based on a dual approach and prove its stability under classical assumptions on the step size.

Finally, in the last section, we give numerical comparisons of our algorithm with the ones proposed by Anthoine et al. [9] and Sawatzky et al. [11]. We will see in particular that our approach gives better results in the sense that the energy F attains lower values and decreases faster.

2. General variational framework for the Expectation Maximization approach with smooth or non smooth a priori

In this section we adapt the general framework of the Expectation Maximization algorithm introduced in [4] to non-smooth posterior distributions. The case of exponential families to which the Poisson distributions belong is then derived as a particular case.

Let $q(\mathbf{x}|\boldsymbol{\phi})$ be the family of sampling densities depending on parameter $\boldsymbol{\phi} \in X$ characterizing the complete data, $p(\mathbf{y}|\boldsymbol{\phi})$ the sampling density of the incomplete data and $k(\mathbf{x}|\mathbf{y}, \boldsymbol{\phi})$ the conditional density of $\mathbf{x}$ given $\mathbf{y}$ and $\boldsymbol{\phi}$. As in [4] we denote

$$P(\boldsymbol{\phi}) = \log(p(\mathbf{y}|\boldsymbol{\phi}), \tag{25}$$

$$Q(\boldsymbol{\phi}'|\boldsymbol{\phi}) = \mathbb{E}(\log q(\mathbf{x}|\boldsymbol{\phi}')|\mathbf{y}, \boldsymbol{\phi}), \tag{26}$$

$$K(\boldsymbol{\phi}'|\boldsymbol{\phi}) = \mathbb{E}(\log k(\mathbf{x}|\mathbf{y}, \boldsymbol{\phi}')|\mathbf{y}, \boldsymbol{\phi}), \tag{27}$$

which implies that

$$Q(\boldsymbol{\phi}'|\boldsymbol{\phi}) = P(\boldsymbol{\phi}') + K(\boldsymbol{\phi}', \boldsymbol{\phi}). \tag{28}$$

Assume that the log-likelihood P of the complete data, described in our tomographic application by equation (6), is differentiable, concave and that its negative $-P$ is coercive,

$$\lim_{\|\boldsymbol{\phi}\| \to +\infty} P(\boldsymbol{\phi}) = +\infty. \tag{29}$$

From the Jensen's inequality it follows that

$$\forall (\boldsymbol{\phi}, \boldsymbol{\phi}') \in X \times X \quad K(\boldsymbol{\phi}'|\boldsymbol{\phi}) \leq K(\boldsymbol{\phi}|\boldsymbol{\phi}). \tag{30}$$

and we also suppose that for all $\boldsymbol{\phi} \in X$, the application $\boldsymbol{\phi}' \mapsto Q(\boldsymbol{\phi}'|\boldsymbol{\phi})$ is concave and bounded above.

Instead of searching for an estimation of ϕ as the maximum of the likelihood P, our strategy here consists to calculate the maximum a posteriori associated to some prior density whose negative log-likelihood is G. Equivalently, we search for the minimum of

$$F = -P + G, \tag{31}$$

where G is assumed to be a continuous convex function possibly non smooth.

The idea of the maximum a posteriori Expectation Maximization (MAP-EM) algorithm is to iteratively reach a maximum of F by solving a sequence of minimization problems

$$\boldsymbol{\phi}^{(n+1)} \in \arg\min_{\boldsymbol{\phi} \in X} \{U(\boldsymbol{\phi}|\boldsymbol{\phi}^{(n)})\}, \tag{32}$$

relaxed versions of (31), where the criterion is

$$U(\boldsymbol{\phi}'|\boldsymbol{\phi}) = F(\boldsymbol{\phi}') - K(\boldsymbol{\phi}'|\boldsymbol{\phi}) = G(\boldsymbol{\phi}') - Q(\boldsymbol{\phi}'|\boldsymbol{\phi}). \tag{33}$$

Hereafter we show that the sequence $(\boldsymbol{\phi}^{(n)})$ converges to a minimum of F. For this we revisit the proof given to a similar result in [4], using the non-smooth convex analysis formalism.

2.1. Convergence of the iterative sequence produced by a GEM algorithm

The following lemma characterizes the solution of the minimum of function F from (31).

Lemma 1. *A vector $\boldsymbol{\phi}^*$ is a minimum of F if and only if*

$$\boldsymbol{\phi}^* \in \arg\min \{U(\boldsymbol{\phi}|\boldsymbol{\phi}^*) \,:\, \boldsymbol{\phi} \in X\}. \tag{34}$$

Proof. Let $\boldsymbol{\phi}^*$ be a minimum of F. From (30) we know that for all $\boldsymbol{\phi} \in X$, $K(\boldsymbol{\phi}|\boldsymbol{\phi}^*) \leq K(\boldsymbol{\phi}^*|\boldsymbol{\phi}^*)$ thus $U(\boldsymbol{\phi}|\boldsymbol{\phi}^*) \geq U(\boldsymbol{\phi}^*|\boldsymbol{\phi}^*)$ and equation (34) is verified. Conversely, if $\boldsymbol{\phi}^*$ verifies (34) then $0 \in \partial U(\cdot\,|\boldsymbol{\phi}^*)(\boldsymbol{\phi}^*) = \partial F(\boldsymbol{\phi}^*) - \nabla K(\cdot|\boldsymbol{\phi}^*)(\boldsymbol{\phi}^*)$, where $\nabla K(\cdot|\boldsymbol{\phi}^*)(\boldsymbol{\phi}^*)$ is the gradient of $K(\cdot|\boldsymbol{\phi}^*)$ at $\boldsymbol{\phi}^*$. From Lemma 2 of [4], $\nabla K(\cdot|\boldsymbol{\phi})(\boldsymbol{\phi}) = 0$ for all $\boldsymbol{\phi}$ and in particular for $\boldsymbol{\phi} = \boldsymbol{\phi}^*$, thus $0 \in \partial F(\boldsymbol{\phi}^*)$ and $\boldsymbol{\phi}^*$ is a minimum of F. □

Hereafter we re-use but also re-define the term Generalized Expectation Maximization (GEM) algorithm from Dempster et al. [4].

Definition 1. *For an application U bounded above and such that for all $\boldsymbol{\phi} \in X$, the mapping $\boldsymbol{\phi}' \in X \mapsto U(\boldsymbol{\phi}'|\boldsymbol{\phi})$ is concave, we call Generalized Expectation Maximization (GEM) algorithm a continuous mapping $M : X \to X$ such that if*

$$\boldsymbol{\phi} \notin \arg\min \{U(\boldsymbol{\phi}'|\boldsymbol{\phi}) \,:\, \boldsymbol{\phi}' \in X\}, \tag{35}$$

then $M(\boldsymbol{\phi})$ should verify

$$U(M(\boldsymbol{\phi})|\boldsymbol{\phi}) < U(\boldsymbol{\phi}|\boldsymbol{\phi}), \tag{36}$$

otherwise $M(\boldsymbol{\phi}) = \boldsymbol{\phi}$.

The next theorem shows that each iteration of a GEM algorithm decreases the value of the cost function F except when the minimum is already attained, and establishes the link between the minima of F and the fixed points of the GEM algorithm.

Theorem 1. *For any GEM algorithm the following properties hold.*

(i) For all $\boldsymbol{\phi} \notin \arg\min \{U(\boldsymbol{\phi}'|\boldsymbol{\phi}) \,:\, \boldsymbol{\phi}' \in X\}$ we have $F(M(\boldsymbol{\phi})) < F(\boldsymbol{\phi})$.

(ii) A vector $\boldsymbol{\phi}^*$ is a fixed point of M if and only if

$$\boldsymbol{\phi}^* \in \arg\min \{U(\boldsymbol{\phi}|\boldsymbol{\phi}^*) \,:\, \boldsymbol{\phi} \in X\}. \tag{37}$$

(iii) The set of fixed points of M coincides with the set of points where F attains its minimum:

$$\forall \boldsymbol{\phi}^* \in X \quad [M(\boldsymbol{\phi}^*) = \boldsymbol{\phi}^* \;\Leftrightarrow\; \boldsymbol{\phi}^* \in \arg\min \{F(\boldsymbol{\phi}) \,:\, \boldsymbol{\phi} \in X\}]. \tag{38}$$

Proof. (i) From (33) and (30) it follows that

$$F(M(\boldsymbol{\phi})) = U(M(\boldsymbol{\phi})|\boldsymbol{\phi}) + K(M(\boldsymbol{\phi})|\boldsymbol{\phi}) \leq U(M(\boldsymbol{\phi})|\boldsymbol{\phi}) + K(\boldsymbol{\phi}|\boldsymbol{\phi}). \quad (39)$$

Then from (36) and again (33) we obtain the result.

(ii) If $\boldsymbol{\phi}^*$ is a fixed point of M then $U(M(\boldsymbol{\phi}^*)|\boldsymbol{\phi}^*) = U(\boldsymbol{\phi}^*|\boldsymbol{\phi}^*)$. Equation (37) follows from definition 1. The reciprocal is obvious by the definition of GEM.

(iii) This property immediately follows from (ii) and Lemma 1.

□

Let $(\boldsymbol{\phi}^{(n)})_{n\in\mathbb{N}}$ be a sequence produced by a GEM algorithm, *i.e.*, for all $n \in \mathbb{N}^*$, $\boldsymbol{\phi}^{(n+1)} = M(\boldsymbol{\phi}^{(n)})$ and $\boldsymbol{\phi}^{(0)}$ is some given initial value in X. We show now that the sequence $(\boldsymbol{\phi}^{(n)})_{n\in\mathbb{N}}$ converges to a minimizer of F at least when F is strictly convex. Note that the strict convexity of F was not required in the seminal paper [4], but the proof given therein is flawed. To the best of our knowledge there is no general proof of convergence for the MAP-EM algorithm, and the proof given for the particular case of the Poisson-MLEM algorithm by A.N. Iusem in [7] cannot be adapted easily to MAP. However, the theoretical strict convexity could be obtained by addition of a second regularization term with a very small coefficient that makes its numerical influence on the result negligible.

Theorem 2. *For any sequence $(\boldsymbol{\phi}^{(n)})_{n\in\mathbb{N}}$ produced by a GEM algorithm the following properties hold.*

(i) The sequence $(F(\boldsymbol{\phi}^{(n)}))$ is non-increasing and converges to the minimum of F.

(ii) If $(\boldsymbol{\phi}^{(n_k)})$ is a convergent sub-sequence of $(\boldsymbol{\phi}^{(n)})$ with limit $\boldsymbol{\phi}^*$, then $\boldsymbol{\phi}^* \in \arg\min\{F(\boldsymbol{\phi}) : \boldsymbol{\phi} \in X\}$.

(iii) If F is strictly convex,

$$\lim_{n\to+\infty} \boldsymbol{\phi}^{(n)} = \arg\min\{F(\boldsymbol{\phi}) : \boldsymbol{\phi} \in X\}.$$

Proof. (i) The fact that the sequence $(F(\boldsymbol{\phi}^{(n)}))$ is non-increasing is a direct consequence of theorem 1 (i). It is clear that a GEM sequence is bounded. Indeed, if this would not be the case, a sub-sequence $(\boldsymbol{\phi}^{(n_k)})$ such that $\lim_{k\to+\infty} \|\boldsymbol{\phi}^{(n_k)}\| = +\infty$ may be extracted. Since F is coercive, the sequence $(F(\boldsymbol{\phi}^{(n_k)}))$ would not be bounded either, which comes in contradiction with the fact that $(F(\boldsymbol{\phi}^{(n)}))$ is non-increasing and bounded below by the minimum of F. Let $(\boldsymbol{\phi}^{(n_k)})$ be a convergent sub-sequence of $(\boldsymbol{\phi}^{(n)})$ with limit $\boldsymbol{\phi}^*$. Then the sub-sequence $(\boldsymbol{\phi}^{(n_k+1)})$ is also convergent and tends to $M(\boldsymbol{\phi}^*)$. The sequence $(F(\boldsymbol{\phi}^{(n)}))$ being non-increasing and bounded bellow, it converges and

$$\lim_{n\to+\infty} F(\boldsymbol{\phi}^{(n)}) = \lim_{k\to+\infty} F(\boldsymbol{\phi}^{(n_k)}) = \lim_{k\to+\infty} F(\boldsymbol{\phi}^{(n_k+1)}),$$

thus $F(M(\boldsymbol{\phi}^*)) = F(\boldsymbol{\phi}^*)$. From theorem 1 (i) we then deduce that

$$\boldsymbol{\phi}^* \in \arg\min\{U(\boldsymbol{\phi}|\boldsymbol{\phi}^*) : \boldsymbol{\phi} \in X\},$$

and from the same theorem it results that $\boldsymbol{\phi}^*$ is a fixed point of M and $\boldsymbol{\phi}^* \in \arg\min\{F(\boldsymbol{\phi}) : \boldsymbol{\phi} \in X\}$. Thus

$$\lim_{n\to+\infty} F(\boldsymbol{\phi}^{(n)}) = F(\boldsymbol{\phi}^*) = \min\{F(\boldsymbol{\phi}) : \boldsymbol{\phi} \in X\}.$$

(ii) As an immediate consequence of the proof of (i) we have:

$$F(\boldsymbol{\phi}^*) \lim_{k\to+\infty} F(\boldsymbol{\phi}^{(n_k)}) = \lim_{n\to+\infty} F(\boldsymbol{\phi}^{(n)}) = \min\{F(\boldsymbol{\phi}) : \boldsymbol{\phi} \in X\}.$$

(iii) If F is strictly convex there is an unique $\boldsymbol{\phi}^* \in \arg\min\{F(\boldsymbol{\phi}) : \boldsymbol{\phi} \in X\}$. From (ii), any convergent sub-sequence of $(\boldsymbol{\phi}^{(n)})$ has to converge to $\boldsymbol{\phi}^*$, thus the sequence converges to the same limit. □

A MAP estimator of the parameter $\boldsymbol{\phi}$ of the complete data density function can thus be obtained by choosing some initial value $\boldsymbol{\phi}^{(0)} \in X$ and then constructing a sequence of minimizers following the recursive relation (32). This condition, corresponding to the (M) step of the EM algorithm, is stronger than the GEM algorithm since it requires

$$U(\boldsymbol{\phi}^{(n+1)}|\boldsymbol{\phi}^{(n)}) \leq U(\boldsymbol{\phi}|\boldsymbol{\phi}^{(n)}) \tag{40}$$

for all $\boldsymbol{\phi} \in X$ and not only for $\boldsymbol{\phi} = \boldsymbol{\phi}^{(n)}$. However, when the solution of (32) is too difficult to obtain exactly, an approximate solution that verifies the GEM condition is sufficient to ensure the convergence of $(F(\boldsymbol{\phi}^{(n)}))$ to the minimum value of F.

We showed that a minimizing sequence for F can thus be obtained from (32) where U is defined by the law of the complete data and the regularization term. In the next subsection we apply this result to the particular case of tomography with low number of photons and data following Poisson distributions, a model largely employed in nuclear medicine imaging.

2.2. Application to tomography with counted photons

In emission tomography, the low emission rates makes that the photons are detected individually and their number follows a Poisson law. The same holds at emission level and the likelihood of the complete data belongs to an exponential family with logarithm that may be written as (for details see *e.g.*, [5, 6]):

$$\log q(\mathbf{x}|\boldsymbol{\lambda}) = -\langle \boldsymbol{\lambda}, \mathbf{s}\rangle + \langle \log(\mathbf{s}\boldsymbol{\lambda}), t(\mathbf{x})\rangle + \log b(\mathbf{x}), \tag{41}$$

where $\mathbf{s} = T^*\mathbf{1}$ and the product of two vectors $(\mathbf{s}\boldsymbol{\lambda})$ is a component-wise product. The (E) step of the EM algorithm consists to calculate its expectation conditionally on the observations $\mathbf{y}$ and on the previous estimate $\boldsymbol{\lambda}^{(n)}$ which is:

$$Q(\boldsymbol{\lambda}|\boldsymbol{\lambda}^{(n)}) = -\langle \boldsymbol{\lambda}, \mathbf{s}\rangle + \langle \log(\mathbf{s}) + \log(\boldsymbol{\lambda}), \mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}\rangle + c, \tag{42}$$

where c is a constant and $\boldsymbol{\lambda}^{(n+1/2)}$ corresponds to the MLEM estimation and is calculated from $\boldsymbol{\lambda}^{(n)}$ following (15). To compute the MAP-EM estimate $\boldsymbol{\lambda}^{(n+1)}$ in case the minimum was not already reached in $\boldsymbol{\lambda}^{(n)}$, one must either choose

$$\boldsymbol{\lambda}^{(n+1)} \in \underset{\mathbf{u}\in(\mathbb{R}_+)^J}{\arg\min} \{\langle \mathbf{u}, \mathbf{s}\rangle - \langle \log(\mathbf{u}), \mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}\rangle + G(\mathbf{u})\}, \tag{43}$$

or at least find a $\boldsymbol{\lambda}^{(n+1)}$ such that $U(\boldsymbol{\lambda}^{(n+1)}|\boldsymbol{\lambda}^{(n)}) < U(\boldsymbol{\lambda}^{(n)}|\boldsymbol{\lambda}^{(n)})$.

Equation (43) was previously derived in [12] for the differentiable total variation prior. The same equation results in the smooth case directly from [4] and was also derived in [13]. Here we obtain (43) as a particular case of the MAP-EM estimation with GEM algorithms for non-smooth priors.

Solutions for the numerical calculation of the MAP estimator were already proposed in the literature. In [13], an explicit scheme with complexity similar to the one of the MLEM algorithm was proposed. The (smoothed) total variation regularization comes as a particular case and the same explicit scheme was studied in [14] and [15]. It was however observed that the explicit scheme is unstable and requires very low regularization parameters. The semi-implicit scheme from [12] also seems unstable. A stable explicit-implicit scheme was proposed in [11] although no proof of convergence was provided.

In the following section, we introduce and give the stability analysis for an efficient dual algorithm for the resolution of (43) with non-smooth total variation regularization.

3. A dual algorithm to compute the MAP-EM estimator

The objective of this section is to approximate numerically the minimizer

$$\boldsymbol{\lambda}^{(n+1)} = \mathbf{u}^* = \arg\min_{\mathbf{u}} \left\{ H(\mathbf{u}) + G(\mathbf{u}) \right\},$$

where

- the function H is defined for all $\mathbf{u} \in (\mathbb{R}_+^*)^J$ by
$$H(\mathbf{u}) = \langle \mathbf{u} - \boldsymbol{\lambda}^{(n+1/2)} \log \mathbf{u}, \mathbf{s} \rangle, \tag{44}$$
with $\mathbf{s}$ and $\boldsymbol{\lambda}^{(n+1/2)}$ constants during the minimization procedure,
- the function G is the discrete total variation functional $G(\mathbf{u}) = \alpha TV(\mathbf{u})$, with TV defined in (9).

Some connection can be made between the functions defined by equations (44) and (13). The role of both is to avoid the linear transform induced by the matrix T and thus to simplify the minimization. However the two functions are fundamentally different, since (44) is defined for variables having the dimension of the complete data and makes use of the partial estimation $\boldsymbol{\lambda}^{(n+1/2)}$, whether (13) is defined for variables having the dimension of the incomplete data.

A first natural idea to minimize $H + G$ consists to use a splitting approach with implicit minimization of both H and G. More precisely, we may consider a Douglas-Rachford splitting consisting to iterate for $l > 0$:

$$\left\{ \begin{array}{lcl} \mathbf{u}^{(l)} & = & \text{prox}_{\,\sigma G}[\mathbf{v}^{(l)}] \\ \mathbf{v}^{(l+1)} & = & \mathbf{v}^{(l)} + \lambda_l(\text{prox}_{\,\sigma H}[2\mathbf{u}^{(l)} - \mathbf{v}^{(l)}] - \mathbf{u}^{(l)}) \end{array} \right. \tag{45}$$

Here, $\lambda_l \in [\varepsilon, 2-\varepsilon]$ with $\varepsilon \in (0,1)$ can be viewed as an inertial parameter, $\sigma > 0$ is the descent step and the initial value $\mathbf{v}^{(0)}$ can be set to $\boldsymbol{\lambda}^{(n+1/2)}$. Then the sequence $(\mathbf{u}^{(l)})_{l \geq 0}$ is expected to converge to a minimum of $H + G$.

The proximal operator $\text{prox}_{\sigma H}[\mathbf{u}]$ defined by

$$\text{prox}_{\sigma H}[\mathbf{u}] = (I + \sigma\partial H)^{-1}(\mathbf{u}) = \arg\min_{\mathbf{v}} \left\{ \frac{1}{2\sigma}\|\mathbf{u}-\mathbf{v}\|^2 + H(\mathbf{v}) \right\},$$

can be calculated explicitly as

$$\text{prox}_{\sigma H}[\mathbf{u}] = \frac{1}{2}\left((\mathbf{u}-\sigma\mathbf{s}) + \sqrt{(\mathbf{u}-\sigma\mathbf{s})^2 + 4\sigma\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}} \right) \tag{46}$$

and the proximal operator $\text{prox}_{\sigma G}[\mathbf{u}]$ can be computed numerically using the Chambolle's algorithm ([8]), $\text{prox}_{\sigma G}[\mathbf{u}] = \mathbf{u} - \sigma \operatorname{div}\boldsymbol{\varphi}$ with $\boldsymbol{\varphi}$ calculated iteratively as:

$$\boldsymbol{\varphi}^{(k+1)} = \frac{\boldsymbol{\varphi}^{(k)} + \tau\nabla\left(\operatorname{div}\boldsymbol{\varphi}^{(k)} - \mathbf{u}/\sigma\right)}{1 + \tau|\nabla\left(\operatorname{div}\boldsymbol{\varphi}^{(k)} - \mathbf{u}/\sigma\right)|}, \tag{47}$$

for some descent step chosen as $0 < \tau < 1/4$. Even if this algorithm gives good solutions, its algorithmic cost is important as it requires three iteration loops: one in n for the calculation of the intermediate MLEM solution, one in l for the Douglas-Rachford splitting and one in k for the application of the Chambolle's algorithm.

In this paper, we propose to improve the efficiency of the method by using a complete dual approach to minimize $H + G$. In particular, as we will see later, the dual approach reduces the complexity as only the loops in n and k are necessary.

As the functions H and G are proper convex and lower semi-continuous, the Fenchel-Rockafellar duality theorem states that if H^* and G^* are the convex conjugates of H and G,

$$H^*(\mathbf{p}) = \sup_{\mathbf{u}} \left\{ \langle \mathbf{u}, \mathbf{p}\rangle - H(\mathbf{u}) \right\} \text{ and } G^*(\mathbf{p}) = \sup_{\mathbf{u}} \left\{ \langle \mathbf{u}, \mathbf{p}\rangle - G(\mathbf{u}) \right\}, \tag{48}$$

then

$$\inf_{\mathbf{u}} \left\{ H(\mathbf{u}) + G(\mathbf{u}) \right\} = -\inf_{\mathbf{p}} \left\{ H^*(-\mathbf{p}) + G^*(\mathbf{p}) \right\}. \tag{49}$$

If $\mathbf{u}^*$ is a minimizer of $H + G$ then there exists a solution $\mathbf{p}^*$ of the dual problem such that $-\mathbf{p}^* \in \partial H(\mathbf{u}^*)$ and $\mathbf{u}^* \in \partial G^*(\mathbf{p}^*)$. The minimizer $\mathbf{u}^*$ of the primal problem should verify $\mathbf{u}^* > 0$. Since H is differentiable for $\mathbf{u}^* > 0$, $-\mathbf{p}^* = \nabla H(\mathbf{u}^*)$ and we get

$$\mathbf{u}^* = \frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s} + \mathbf{p}^*}. \tag{50}$$

A minimizer $\mathbf{u}^*$ of $H + G$ can then be deduced from a minimizer $\mathbf{p}^*$ of $H^* + G^*$.

3.1. Formulation of the dual problem

The Fenchel-Legendre transform of H can be calculated explicitly as follows. From the Euler equation associated to the maximization in $\mathbf{u}$ of $\langle \mathbf{u}, \mathbf{p}\rangle - H(\mathbf{u})$ we obtain

$$H^*(\mathbf{p}) = \langle \mathbf{u}^*[\mathbf{p}], \mathbf{p}\rangle - H(\mathbf{u}^*[\mathbf{p}]) \tag{51}$$

with

$$\mathbf{u}^*[\mathbf{p}] = \max\left\{ \frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s}-\mathbf{p}}, 0 \right\}. \tag{52}$$

Thus $H^*(\mathbf{p})$ is infinite for $\mathbf{p} \geq \mathbf{s}$. For $\mathbf{p} < \mathbf{s}$ its value is:

$$\begin{aligned} H^*(\mathbf{p}) &= -\left\langle \frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s}-\mathbf{p}}, \mathbf{s}-\mathbf{p} \right\rangle + \left\langle \log \frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s}-\mathbf{p}}, \mathbf{s}\boldsymbol{\lambda}^{(n+1/2)} \right\rangle \\ &= \left\langle \log(\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}) - \mathbf{1}, \mathbf{s}\boldsymbol{\lambda}^{(n+1/2)} \right\rangle - \left\langle \log(\mathbf{s}-\mathbf{p}), \mathbf{s}\boldsymbol{\lambda}^{(n+1/2)} \right\rangle \\ &= C - \left\langle \log(\mathbf{s}-\mathbf{p}), \mathbf{s}\boldsymbol{\lambda}^{(n+1/2)} \right\rangle, \end{aligned} \tag{53}$$

with C independent from $\mathbf{p}$. On the other side, it was shown in [8] that the Fenchel-Legendre transform of the total variation functional G is

$$G^*(\mathbf{p}) = \chi_{\alpha K}(\mathbf{p}) = \begin{cases} 0 & \text{if } \mathbf{p} \in \alpha K \\ +\infty & \text{otherwise,} \end{cases} \tag{54}$$

where K is the set

$$K = \{\operatorname{div} \boldsymbol{\varphi} : \ \boldsymbol{\varphi} \in Y, \ |\boldsymbol{\varphi}| \leq 1\}.$$

Now from (53) and (54) it follows that the solution of dual problem p^* can be expressed as $p^* = \alpha \operatorname{div} \boldsymbol{\varphi}^*$ where $\boldsymbol{\varphi}^*$ is defined as a solution of

$$\inf\{h(\boldsymbol{\varphi}) : \ \boldsymbol{\varphi} \in Y, \ |\boldsymbol{\varphi}| \leq 1\}, \tag{55}$$

where the function h is defined for $\boldsymbol{\varphi}$ such that $\mathbf{s} + \alpha \operatorname{div} \boldsymbol{\varphi} > 0$ by

$$h(\boldsymbol{\varphi}) = H^*(-\alpha \operatorname{div} \boldsymbol{\varphi}) - C = -\langle \log(\mathbf{s} + \alpha \operatorname{div} \boldsymbol{\varphi}), \mathbf{s}\boldsymbol{\lambda}^{(n+1/2)} \rangle \tag{56}$$

and is infinite elsewhere. Finally, by the duality equation (52), the minimizer $\mathbf{u}^*$ of the primal problem can be obtained from the minimizer $\boldsymbol{\varphi}^*$ of the dual problem as:

$$\mathbf{u}^* = \frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s} + \alpha \operatorname{div} \boldsymbol{\varphi}^*}. \tag{57}$$

Remark 1. *From a discrete point of view, note that if for all $j = 1 \ldots, J$, $|\varphi_j| \leq 1$, we have $\|\operatorname{div} \boldsymbol{\varphi}\|_\infty \leq 4$ and then for $\alpha < s_{\min}/4$ we obtain*

$$\mathbf{s} - \alpha \operatorname{div} \boldsymbol{\varphi} \geq s_{\min} - 4\alpha > 0,$$

where $s_{\min}$ is defined in (24). The functional h is thus finite on $S = \{\boldsymbol{\varphi} \in Y : \ |\boldsymbol{\varphi}| \leq 1\}$. In the rest of the paper we assume that $\alpha < s_{\min}/4$. This condition is also assumed for the proof of convergence given in theorem 3.

Remark 2. *In a generalization from two-dimensional images to three-dimensional volumes we will have $\|\operatorname{div} \boldsymbol{\varphi}\|_\infty \leq 6$ and the condition on the regularization parameter becomes $\alpha < s_{\min}/6$.*

3.2. An iterative scheme to solve the dual problem

In this subsection we derive an algorithm to compute a solution of (55), that is

$$\boldsymbol{\varphi}^* = \arg\min_{\boldsymbol{\varphi} \in S} \ h(\boldsymbol{\varphi}), \tag{58}$$

with $S = \{\boldsymbol{\varphi} \in Y : \ |\boldsymbol{\varphi}| \leq 1\}$. The gradient of h is

$$\nabla h(\boldsymbol{\varphi}) = \alpha \, \nabla \left(\frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}} \right). \tag{59}$$

Now, given $\boldsymbol{\mu} \in X$ the Lagrange multiplier associated to the constraint $|\boldsymbol{\varphi}| \leq 1$, the Karush-Kuhn-Tuker condition reads:

$$\alpha \, \nabla \left(\frac{s\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}} \right) + \boldsymbol{\mu}\boldsymbol{\varphi} = 0, \tag{60}$$

with either $|\varphi_j| = 1$ and $\mu_j > 0$, or $|\varphi_j| < 1$ and $\mu_j = 0$. Moreover, as in the latter case we also have $\left| \left(\nabla \left(\frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}} \right) \right)_j \right| = 0$, we see that

$$\boldsymbol{\mu} = \alpha \left| \nabla \left(\frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}} \right) \right|. \tag{61}$$

Following the idea of Chambolle's algorithm, we consider a semi-implicit gradient descent scheme and we show that it converges to the solution $\boldsymbol{\varphi}^*$ of (58). With some minimization step $\tau > 0$ and the initial value $\boldsymbol{\varphi}^{(0)} = 0$ we then consider the scheme

$$\boldsymbol{\varphi}^{(k+1)} = \boldsymbol{\varphi}^{(k)} - \tau \nabla \left(\frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}^{(k)}} \right) - \tau \left| \nabla \left(\frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}^{k}} \right) \right| \boldsymbol{\varphi}^{(k+1)}, \tag{62}$$

which can be written equivalently as

$$\boldsymbol{\varphi}^{(k+1)} = \frac{\boldsymbol{\varphi}^{(k)} - \tau \mathbf{z}^{(k)}}{1 + \tau |\mathbf{z}^{(k)}|}, \quad \text{with } \mathbf{z}^{(k)} = \nabla \left(\frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}^{(k)}} \right). \tag{63}$$

Hereafter we will show the convergence of the sequence of iterates to the solution of (58), which requires the following auxiliary result.

Lemma 2. *If $\alpha < s_{\min}/4$, the function h defined in (56) is convex and continuously differentiable on its domain. Its gradient is Lipschitz on S with constant*

$$L_h = 8\alpha^2 \frac{\|\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}\|_\infty}{(s_{\min} - 4\alpha)^2}. \tag{64}$$

Proof. The first derivative of h satisfies for $\boldsymbol{\varphi} \in Y$ such that $\mathbf{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi} > 0$ and $\boldsymbol{\psi} \in X$ the equation:

$$h'(\boldsymbol{\varphi})(\boldsymbol{\psi}) = -\alpha \left\langle \frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}}, \mathrm{div}\, (\boldsymbol{\psi}) \right\rangle,$$

thus

$$\begin{aligned}
\|\nabla h(\boldsymbol{\varphi}) - \nabla h(\boldsymbol{\varphi}')\|_2 &= \sup_{\|\boldsymbol{\psi}\|_2 = 1} |h'(\boldsymbol{\varphi})(\boldsymbol{\psi}) - h'(\boldsymbol{\varphi}')(\boldsymbol{\psi})| \\
&= \alpha^2 \sup_{\|\boldsymbol{\psi}\|_2 = 1} \left| \left\langle \frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)} \mathrm{div}\, (\boldsymbol{\varphi} - \boldsymbol{\varphi}')}{(\mathbf{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi})(\mathbf{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}')}, \mathrm{div}\, (\boldsymbol{\psi}) \right\rangle \right| \\
&\leq \alpha^2 \|\mathrm{div}\, (\boldsymbol{\varphi} - \boldsymbol{\varphi}')\|_2 \sup_{\|\boldsymbol{\psi}\|_2 = 1} \left\| \frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)} \mathrm{div}\, \boldsymbol{\psi}}{(\mathbf{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi})(\mathbf{s} + \alpha \mathrm{div}\, \boldsymbol{\varphi}')} \right\|_2.
\end{aligned}$$

As shown in [8], $\|\mathrm{div}\,(\boldsymbol{\varphi}-\boldsymbol{\varphi}')\|_2^2 \leq 8\|\boldsymbol{\varphi}-\boldsymbol{\varphi}'\|_2^2$ and $\|\mathrm{div}\,\boldsymbol{\psi}\|_2^2 \leq 8$, thus

$$\|\nabla h(\boldsymbol{\varphi}) - \nabla h(\boldsymbol{\varphi}')\|_2 \leq 8\alpha^2 \left\| \frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{(\mathbf{s}+\alpha\mathrm{div}\,\boldsymbol{\varphi})(\mathbf{s}+\alpha\mathrm{div}\,\boldsymbol{\varphi}')} \right\|_\infty \|\boldsymbol{\varphi}-\boldsymbol{\varphi}'\|_2.$$

Since on S we have $\|\mathrm{div}\,\boldsymbol{\varphi}\|_\infty \leq 4$ it follows that

$$\|\nabla h(\boldsymbol{\varphi}) - \nabla h(\boldsymbol{\varphi}')\|_2 \leq 8\alpha^2 \frac{\|\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}\|_\infty}{(s_{\min}-4\alpha)^2} \|\boldsymbol{\varphi}-\boldsymbol{\varphi}'\|_2,$$

thus the gradient of h is Lipschitz with constant given in (64).

The second derivative of h is

$$h''(\boldsymbol{\varphi})(\boldsymbol{\psi})(\widetilde{\boldsymbol{\psi}}) = \left\langle \frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{(\mathbf{s}+\alpha\mathrm{div}\,\boldsymbol{\varphi})^2}, \alpha^2 \mathrm{div}\,(\boldsymbol{\psi})\mathrm{div}\,(\widetilde{\boldsymbol{\psi}}) \right\rangle,$$

thus h is convex as

$$h''(\boldsymbol{\varphi})(\boldsymbol{\psi})(\boldsymbol{\psi}) = \left\langle \frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{(\mathbf{s}+\alpha\mathrm{div}\,\boldsymbol{\varphi})^2}, \alpha^2 \mathrm{div}\,(\boldsymbol{\psi})^2 \right\rangle \geq 0.$$

□

Remark 3. *When the coefficient of regularization is sufficiently small, i.e* $\alpha << 1$*, the constant* L_h *can also be approximated by*

$$L_h \simeq 8\alpha^2 \|\boldsymbol{\lambda}^{(n+1/2)}/\mathbf{s}\|_\infty. \tag{65}$$

We can now state the main result of this section.

Theorem 3. *Let* $\alpha < s_{\min}/4$ *and* $\tau < \alpha/L_h$*. Then the sequence* $(h(\boldsymbol{\varphi}^{(k)}))$ *with* $\boldsymbol{\varphi}^{(k)}$ *defined in (63) is decreasing and converges to the minimum of* h *on* S*.*

Proof. It is easy to show by induction that $|\boldsymbol{\varphi}^{(k)}| \leq 1$ for all $k \in \mathbb{N}$. As the functional h is convex and with gradient L_h-Lipschitz on K, it follows from a classical result (see *e.g.*, [16]) that for all $k \in \mathbb{N}$,

$$h(\boldsymbol{\varphi}^{(k+1)}) \leq h(\boldsymbol{\varphi}^{(k)}) + \langle \nabla h(\boldsymbol{\varphi}^{(k)}), \boldsymbol{\varphi}^{(k+1)} - \boldsymbol{\varphi}^{(k)} \rangle + \frac{L_h}{2}\|\boldsymbol{\varphi}^{(k+1)} - \boldsymbol{\varphi}^{(k)}\|_2^2.$$

Let us note $\boldsymbol{\eta} = (\boldsymbol{\varphi}^{(k+1)} - \boldsymbol{\varphi}^{(k)})/\tau$. As $\nabla h(\boldsymbol{\varphi}^{(k)}) = \alpha \mathbf{z}^{(k)}$, we obtain

$$h(\boldsymbol{\varphi}^{(k+1)}) - h(\boldsymbol{\varphi}^{(k)}) \leq \tau \left(\alpha \langle \mathbf{z}^{(k)}, \boldsymbol{\eta} \rangle + \tau \frac{L_h}{2} \|\boldsymbol{\eta}\|_2^2 \right),$$

From (62) it follows that

$$\boldsymbol{\eta} = -\mathbf{z}^{(k)} - |\mathbf{z}^k| \boldsymbol{\varphi}^{(k+1)}, \text{ with } \mathbf{z}^{(k)} = \nabla \left(\frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s}+\alpha\mathrm{div}\,\boldsymbol{\varphi}^{(k)}} \right),$$

which is well defined as $\mathbf{s}+\alpha\mathrm{div}\,\boldsymbol{\varphi}^{(k)}>0$. From the last two equations we then obtain

$$\begin{aligned} h(\boldsymbol{\varphi}^{(k+1)})-h(\boldsymbol{\varphi}^{(k)}) \leq & -\alpha(\|\mathbf{z}^{(k)}\|_2^2+\langle \mathbf{z}^{(k)},|\mathbf{z}^{(k)}|\boldsymbol{\varphi}^{(k+1)}\rangle) \\ & +\tau\frac{L_h}{2}\left(\|\mathbf{z}^{(k)}\|_2^2+2\langle \mathbf{z}^{(k)},|\mathbf{z}^{(k)}|\boldsymbol{\varphi}^{(k+1)}\rangle+\|\mathbf{z}^{(k)}\boldsymbol{\varphi}^{(k+1)}\|_2^2\right). \end{aligned}$$

Now, as $\|\boldsymbol{\varphi}^{(k+1)}\|_\infty\leq 1$, it holds $\|\mathbf{z}^{(k)}\boldsymbol{\varphi}^{(k+1)}\|_2^2\leq\|\mathbf{z}^{(k)}\|_2^2$ thus

$$h(\boldsymbol{\varphi}^{(k+1)})-h(\boldsymbol{\varphi}^{(k)})\leq(-\alpha+\tau L_h)\left(\|\mathbf{z}^{(k)}\|_2^2+\langle \mathbf{z}^{(k)},|\mathbf{z}^{(k)}|\boldsymbol{\varphi}^{(k+1)}\rangle\right).$$

Finally, as $\langle \mathbf{z}^{(k)},|\mathbf{z}^{(k)}|\boldsymbol{\varphi}^{(k+1)}\rangle\geq-\|\mathbf{z}^{(k)}\|_2^2$ and $\alpha-\tau L_h>0$, we have shown that $h(\boldsymbol{\varphi}^{(k+1)})\leq h(\boldsymbol{\varphi}^{(k)})$ for all $k\in\mathbb{N}$.

Let $\ell=\lim_{k\to\infty}h(\boldsymbol{\varphi}^{(k)})$. The bounded sequence $(\boldsymbol{\varphi}^{(k)})$ has a converging subsequence $(\boldsymbol{\varphi}^{(k_m)})$ with limit some $\boldsymbol{\varphi}^*\in Y$. The subsequence $(\boldsymbol{\varphi}^{(k_m+1)})$ also converges to some $\widetilde{\boldsymbol{\varphi}}\in Y$. Passing to the limit in (63) we get

$$\widetilde{\boldsymbol{\varphi}}=\frac{\boldsymbol{\varphi}^*-\tau\mathbf{z}^*}{1+\tau|\mathbf{z}^*|},\qquad \text{with } \mathbf{z}^*=\nabla\left(\frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s}+\alpha\mathrm{div}\,\boldsymbol{\varphi}^*}\right). \tag{66}$$

With the same arguments as above it can be shown that

$$h(\tilde{\boldsymbol{\varphi}})-h(\boldsymbol{\varphi}^*)\leq(-\alpha+\tau L_h)\left(\|\mathbf{z}^*\|_2^2+\langle \mathbf{z}^*,|\mathbf{z}^*|\tilde{\boldsymbol{\varphi}}\rangle\right)\leq 0.$$

Moreover, as $\ell=\lim_{m\to\infty}h(\boldsymbol{\varphi}^{(k_m)})=\lim_{m\to\infty}h(\boldsymbol{\varphi}^{(k_m+1)})$ we deduce that $h(\widetilde{\boldsymbol{\varphi}})=h(\boldsymbol{\varphi}^*)$. Thus $\|\mathbf{z}^*\|_2^2+\langle \mathbf{z}^*,|\mathbf{z}^*|\tilde{\boldsymbol{\varphi}}\rangle=0$, which implies that $\mathbf{z}^*+|\mathbf{z}^*|\tilde{\boldsymbol{\varphi}}=0$. From (66) it follows that $\boldsymbol{\varphi}^*=\widetilde{\boldsymbol{\varphi}}$ and it satisfies the equation

$$\nabla\left(\frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s}+\alpha\mathrm{div}\,\boldsymbol{\varphi}^*}\right)+\left|\nabla\left(\frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s}+\alpha\mathrm{div}\,\boldsymbol{\varphi}^*}\right)\right|\boldsymbol{\varphi}^*=0,$$

which is the Euler equation for the dual problem. Finally,

$$\lim_{k\to\infty}h(\boldsymbol{\varphi}^{(k)})=\ell=\min_{\boldsymbol{\varphi}\in S}h(\boldsymbol{\varphi}).$$

□

Remark 4. *Notice that the case $\alpha>s_{\min}/4$ can be treated numerically using the following scheme*

$$\boldsymbol{\varphi}^{(k+1)}=\frac{\boldsymbol{\varphi}^{(k)}+\tau\mathbf{z}^{(k)}}{1+\tau|\mathbf{z}^{(k)}|},\qquad \text{with } \mathbf{z}^{(k)}=\nabla\left(\max\left\{\frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s}+\alpha\mathrm{div}\,\boldsymbol{\varphi}^{(k)}},0\right\}\right), \tag{67}$$

which seems to stabilize the previous scheme for large value of α. Indeed, this modified dual approach can be viewed as a projection of the primal variable

$$\mathbf{u}^{(k)}=\frac{\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}}{\mathbf{s}+\alpha\mathrm{div}\,\boldsymbol{\varphi}^{(k)}},$$

on $(\mathbb{R}_+)^J$ (see definition of $\mathbf{u}^*[p]$ in (52)).

Remark 5. *For the three-dimensional case, the hypotheses of theorem 3 becomes $\alpha<s_{\min}/6$ and $\tau\leq\alpha/L_h$ where*

$$L_h=12\alpha^2\frac{\|\mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}\|_\infty}{(s_{\min}-6\alpha)^2}. \tag{68}$$

4. Numerical experiments

4.1. The algorithms

In this section, we give some numerical experiments of our global scheme which reads as

- (EM) step from MAP-EM

$$\boldsymbol{\lambda}^{(n+1/2)} = \frac{\boldsymbol{\lambda}^{(n)}}{T^*\mathbf{1}} T^* \left[\frac{\mathbf{y}}{T\boldsymbol{\lambda}^{(n)}}\right]$$

- TV minimization

$$\boldsymbol{\lambda}^{(n+1)} \in \underset{\mathbf{u}\in(\mathbb{R}_+)^J}{\arg\min} \{\langle \mathbf{u}, \mathbf{s}\rangle - \langle \log(\mathbf{u}), \mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}\rangle + G(\mathbf{u})\}.$$

We also test a FISTA version. Indeed, in a certain sense, the up-mentioned scheme can be viewed as an ISTA approach ([17]) which alternates minimization of a smooth and non smooth energy, and could further be accelerated by adding a FISTA step. The algorithm becomes :

- ISTA Scheme

$$\widetilde{\boldsymbol{\lambda}}^{(n+1)} \in \underset{\mathbf{u}\in(\mathbb{R}_+)^J}{\arg\min} \{\langle \mathbf{u}, \mathbf{s}\rangle - \langle \log(\mathbf{u}), \mathbf{s}\boldsymbol{\lambda}^{(n+1/2)}\rangle + G(\mathbf{u})\},$$

  with $\boldsymbol{\lambda}^{(n+1/2)} = \frac{\boldsymbol{\lambda}^{(n)}}{T^*\mathbf{1}} T^* \left[\frac{\mathbf{y}}{T\boldsymbol{\lambda}^{(n)}}\right]$.

- Inertial parameter step

$$t_{n+1} = \frac{1+\sqrt{1+4t_n^2}}{2},$$

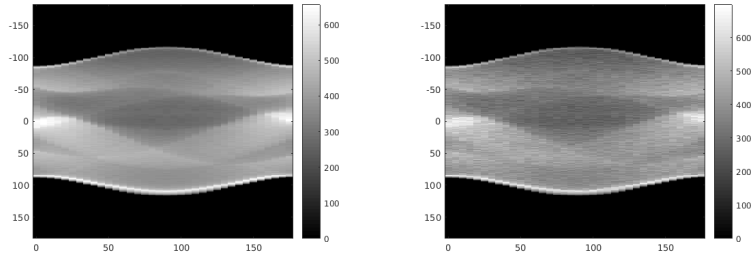
- FISTA update

$$\boldsymbol{\lambda}^{(n+1)} = \widetilde{\boldsymbol{\lambda}}^{(n+1)} + \frac{(t_n-1)}{t_{n+1}}(\widetilde{\boldsymbol{\lambda}}^{(n+1)} - \widetilde{\boldsymbol{\lambda}}^{(n)})$$

where $t_1 = 1$ and $\widetilde{\boldsymbol{\lambda}}^{(0)} = \mathbf{1}$.

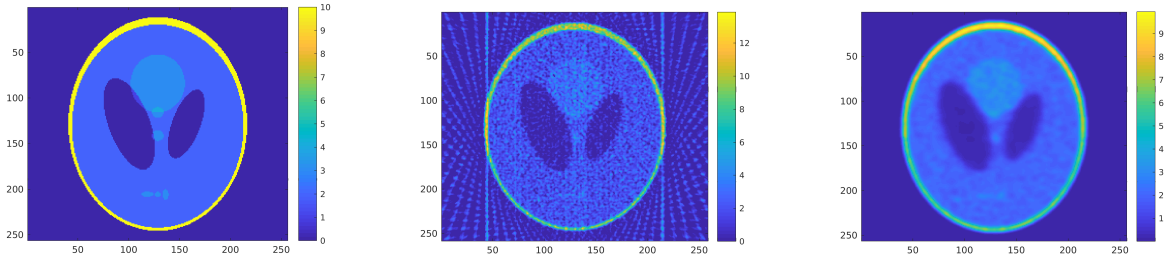
4.2. Results

Ideal Radon projections of the Shepp-Logan phantom from Matlab, multiplied by 10 and sampled in 256^2 pixels, were calculated for angles running from 0° to 175° in steps of 5°. Random values for empirical projections were drawn from a multi-variate Poisson law having mean the theoretical projections. The total number of counts is then about $3*10^6$. Ideal projections and their noisy counterparts are shown for comparison in figure 1. In this test, where the number of projections and the number of counts are low, the quality of the reconstructed image is low too. For comparison, in figure 1 (c) we show the analytic reconstruction including the Hamming filter from Matlab.



(a) Ideal projections (b) Empirical projections

Figure 1. Sinograms for the Shepp-Logan phantom: (a) without noise, (b) with Poisson random noise.



(a) Reference (b) Analytic (c) MLEM

Figure 2. (a) Reference, (b) The analytic reconstruction from noisy projections including Hamming filtering, (c) MLEM solution after 50 iterations and Gaussian smoothing with $\sigma = 2$.

We add a total variation penalty term to the negative log-likelihood and solve numerically for the minimum of the energy with different algorithms mentioned in the paper. Unless otherwise stated, the value of the regularization coefficient α is set to 0.025, where this value gave in a subjective analysis the best compromise between smoothness of the homogeneous regions and image contrast. The noise realizations are different between figures, but identical inside all of the comparison figures. As figures of merit we plot the total energy calculated from equation (10) and the mean squared error,

$$MSE(\widehat{\boldsymbol{\lambda}}) = \frac{1}{J}\|\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}\|_2. \tag{69}$$

We first compare in figure 3 the proposed dual method with the Douglas-Rachford splitting described in section §3 after 200 MLEM iterations. The two methods give very similar results but the computing time is sensibly larger for Douglas-Rachford due to an additional internal loop (see section §3 for details). In our numerical comparisons the computing time was reduced by a factor of ten with the proposed dual algorithm.

The method proposed by Anthoine et al. do not use the capability of the EM algorithm to rapidly increase the likelihood and converges slowly compared to the proposed dual and the Sawatzky et al. methods, as it can be seen in figure 4. After 1000 iterations the reconstructed images and extracted central vertical profiles are relatively

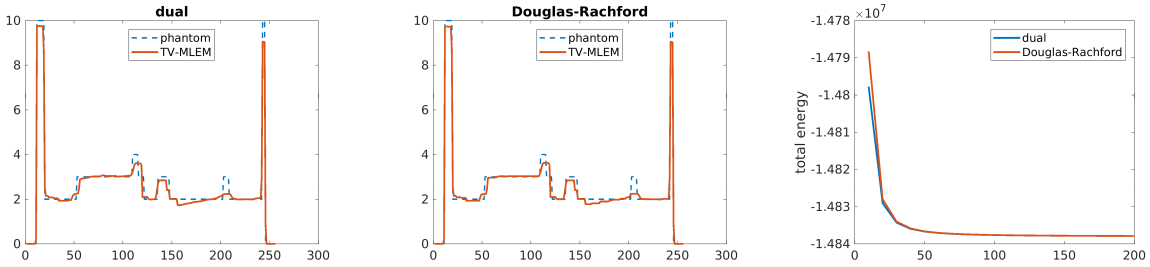


Figure 3. Comparison between Douglas-Rachford splitting and dual algorithm. We show the central vertical profiles from the reconstructed images (left and center) and the total energy curves for the two methods (right). The time necessary to run the 200 iterations is 10 times less for the dual algorithm compared to Douglas-Rachford. The results are very similar and the mean squared error is not shown since the curves look identical.

close to each other.

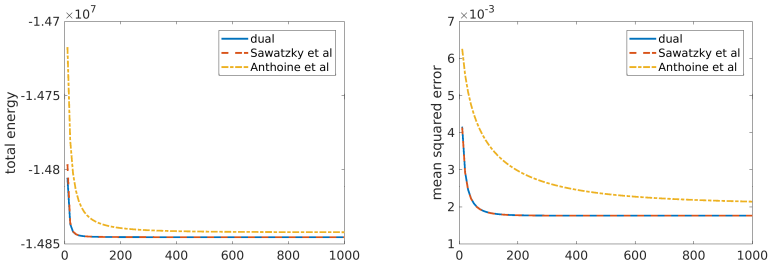
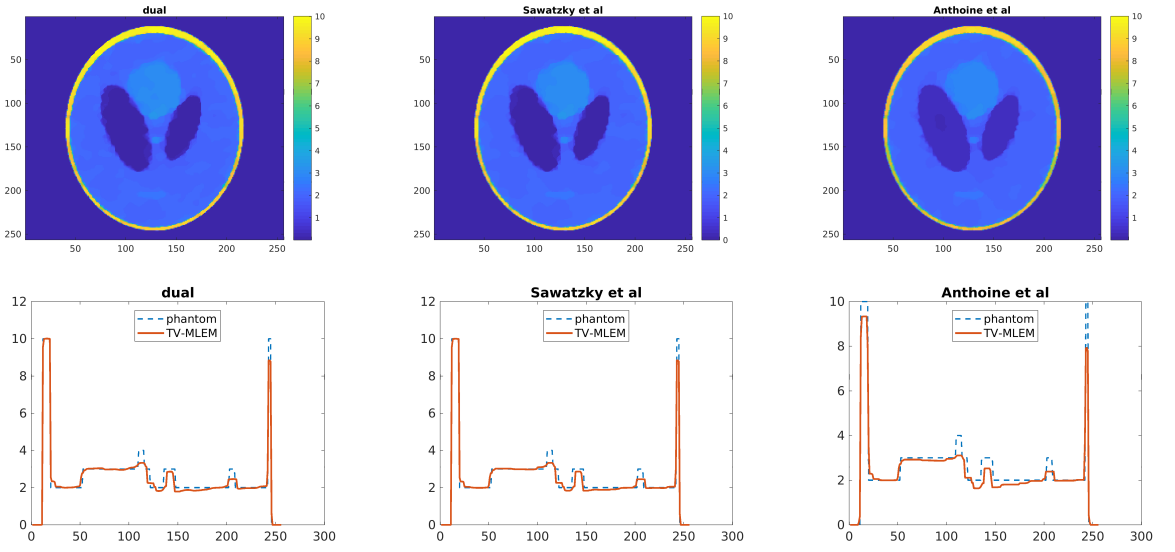


Figure 4. Figures of merit for the comparison between the proposed dual method (blue line), Sawatzky et al. method (red line) and Anthoine et al. method (yellow line).

In figure 6 we evaluate the influence of the FISTA acceleration technique on the convergence speed of the dual algorithm. After 200 iterations (this number was set arbitrarily for homogeneity with the previous comparisons) the results with and without FISTA are the same. However, the FISTA technique allows to reach the numerical convergence in about 30 iterations whereas it requires more than 50 iterations for the dual method without FISTA technique.

5. Conclusions

In this paper we adapt the maximum-a-posteriori expectation maximization framework to non-smooth convex priors, aiming to maximize an energy composed of a likelihood function and a prior distribution. Its specificity is to split the search for the optimal value in an expectation step that allows to move from an optimization problem in the domain of incomplete data (in our case the projections) to a simpler one in the domain of complete data (in our case the image), and a maximization step for the new criterion. We then deduce consistency and convergence results for the MAP-EM algorithms. Total



(a) Dual method (b) Sawatzky et al. method (c) Anthoine et al. method

Figure 5. Comparison between the proposed dual method, Sawatzky et al. and Anthoine et al. methods after 1000 MLEM iterations. We recall that the noise realization is different from figure 3 and thus the two figures should not be compared.

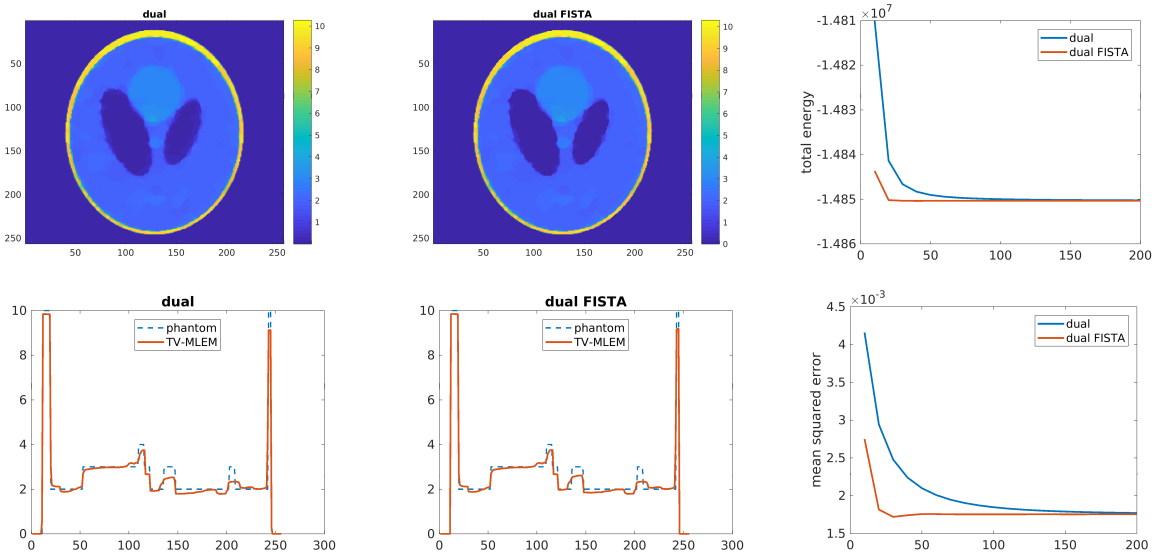


Figure 6. FISTA acceleration strongly improves the convergence of the dual algorithm.

variation regularization of tomographic images calculated from Poisson distributed data is then derived from as a particular case. For the maximization step we lean on the Fenchel-Rockafellar duality principle and we propose a simple and effective algorithm developed following the ideas that A. Chambolle first introduced for fidelity terms expressed with the ℓ_2-induced distance. We then succeed to prove its convergence to the solution at least for regularization parameters that do not exceed a given upper

bound. The resulting MAP-EM algorithm is both consistent and relatively fast and successfully competes experimentally with other algorithms proposed in the literature for the resolution of the same problem. Our results also tend to show that FISTA acceleration allows to further improve the convergence speed although we are aware that our setup is different from the one where FISTA was originally proposed.

Acknowledgements

The author acknowledges the financial support of the French National Research Agency through the ANR project 3DCLEAN (ANR-15-CE09-0009) and the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the ANR.

References

[1] Jeffrey A Fessler and W Leslie Rogers. Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs. *IEEE Transactions on Image processing*, 5(9):1346–1358, 1996.

[2] Johan Nuyts and Jeffrey A Fessler. A penalized-likelihood image reconstruction method for emission tomography, compared to postsmoothed maximum-likelihood with matched spatial resolution. *IEEE transactions on medical imaging*, 22(9):1042–1052, 2003.

[3] Amir Beck and Marc Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *Trans. Img. Proc.*, 18(11):2419–2434, November 2009.

[4] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39:1–38, 1977.

[5] K. Lange and R. Carson. EM reconstruction algorithms for emission and transmission tomography. *J Comput Assist Tomogr*, 8(2):306–316, 1984.

[6] Y. Vardi, L.A. Shepp, and L. Kaufman. A statistical model for positron emission tomography. *Journal of the American Statistical Association*, 80(389):8–20, 1985.

[7] Alfredo N. Iusem. A short convergence proof of the EM algorithm for a specific Poisson model. *Brazilian Journal of Probability and Statistics*, 6(1):57–67, 1992.

[8] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1-2):89–97, 2004.

[9] Sandrine Anthoine, Jean-François Aujol, Yannick Boursier, and Clothilde Melot. Some proximal methods for Poisson intensity CBCT and PET. *Inverse problems and Imaging*, 6(4):565–598, 2012.

[10] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[11] A. Sawatzky, C. Brune, F. Wubbeling, T. Kosters, K. Schafers, and M. Burger. Accurate EM-TV algorithm in PET with low SNR. In *Nuclear Science Symposium Conference Record, 2008. NSS'08. IEEE*, pages 5133–5137. IEEE, 2008.

[12] Ming Yan, Jianwen Chen, Luminita A. Vese, John Villasenor, Alex Bui, and Jason Cong. *EM+TV Based Reconstruction for Cone-Beam CT with Reduced Radiation*, pages 1–10. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[13] P. J. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Transactions on Medical Imaging*, 9(1):84–93, Mar 1990.

[14] V. Y. Panin, G. L. Zeng, and G. T. Gullberg. Total variation regulated EM algorithm. *IEEE Transactions on Nuclear Science*, 46(6):2202–2210, Dec 1999.

[15] M Persson, D Bone, and H Elmqvist. Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography. *Physics in Medicine & Biology*, 46(3):853, 2001.

[16] James M Ortega and Werner C Rheinboldt. *Iterative solution of nonlinear equations in several variables*, volume 30. Siam, 1970.

[17] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.