An alternating extragradient method for total variation based image restoration from Poisson data

S. Bonettini and V. Ruggiero

Dipartimento di Matematica, Università di Ferrara, Polo Scientifico Tecnologico, Blocco B, Via Saragat 1, I-44100 Ferrara, Italy

E-mail: `silvia.bonettini@unife.it`, `valeria.ruggiero@unife.it`

Abstract. Variational models are a valid tool for edge–preserving image restoration from data affected by Poisson noise. This paper deals with total variation and hypersurface regularization in combination with the Kullbach Leibler divergence as data fidelity function. We propose an iterative method, based on an alternating extragradient scheme, which is able to solve in a numerically stable way the primal–dual formulation of both total variation and hypersurface regularization problems. In this method, tailored for general smooth saddle–point problems, the stepsize parameter can be adaptively computed so that the convergence of the scheme is proved under mild assumptions. In the numerical experience, we focus the attention on the artificial smoothing parameter that makes different the total variation and hypersurface regularization. A set of experiments on image denoising and deblurring problems is performed in order to evaluate the influence of this smoothing parameter on the stability of the proposed method and on the features of the restored images.

1. Introduction

The image restoration problems arising in many applications, such as fluorescence microscopy, optical–infrared astronomy, digital radiography are characterized by data affected by Poisson noise. In particular, the image formation can be modeled as a linear process including deterministic and statistical aspects (see [1] for an overview). More precisely, denoting by $g \in \mathbb{R}^m$ the detected data, each measured value g_i is a realization of a Poisson random variable with expected value $(Hx + b)_i$, where $x \in \mathbb{R}^n$ is the original object, the matrix $H \in \mathbb{R}^{m \times n}$ represents the distortion due to the acquisition system and $b \in \mathbb{R}^m$ is a positive constant background term.

From this point of view, the image restoration consists in the inverse problem of finding an approximation of x, given g, H and b. When $H = I$ we have a denoising problem while, in the other cases, we deal with a deblurring problem.

In the Bayesian framework [2, 3], the restoration is obtained by solving the following optimization problem

$$\min_{x \geq \eta} f(x) \equiv f_0(x) + \beta f_1(x) \tag{1}$$

where $f_0(x)$ is a positive functional measuring the data discrepancy, $f_1(x)$ is a regularization term and β is a positive parameter balancing the relative weight of the two terms. The constraint $x \geq \eta$, where $\eta \in \mathbb{R}^n$, $\eta \geq 0$, is also imposed, since solutions with negative entries are not meaningful for the mentioned applications.

Under the hypothesis of Poisson noise, the following Kullback-Leibler (KL) divergence expresses the data discrepancy:

$$f_0(x) = \sum_k \left\{ g_k \ln \frac{g_k}{(Hx+b)_k} + (Hx+b)_k - g_k \right\} \tag{2}$$

with $g_k \ln g_k = 0$ if $g_k = 0$. Usual hypotheses on the imaging matrix H are the nonnegativity of its entries and a normalization condition $\sum_i H_{ij} = 1, \quad \forall j$.

The choice of the regularization functional $f_1(x)$ strongly affects some features of the solutions of (1), as for example the smoothness, the sparsity, the sharpness of the edges. In this paper we focus on the edge–preserving regularization via the Total Variation (TV) functional, which, in the discrete framework, is defined as

$$f_1(x) = \sum_{k=1}^n \|(\nabla x)_k\| \tag{3}$$

where $(\nabla x)_k$ denotes a discrete approximation of the gradient of x at the pixel k and $\|.\|$ is the usual l_2 norm of a vector ($\|a\| = \sqrt{\sum_i a_i^2}$).

The TV functional (3) introduces many difficulties from the analytical and numerical point of view, since it is not everywhere differentiable. In order to overcome these difficulties, several authors consider the following smoothed approximation

$$f_1(x) = \sum_{k=1}^n \left\| \begin{pmatrix} (\nabla x)_k \\ \delta \end{pmatrix} \right\| \tag{4}$$

where δ is a nonzero parameter [4]. This variant has been considered also in a more general edge–preserving regularization framework as Hypersurface (HS) potential [5].

Here and in the following, we will indicate the combination of (2) and (3) as KL-TV and its smoothed version (2)-(4) as KL-HS.

A more classical approach to the edge–preserving regularization is the combination of the quadratic term $f_0(x) = \frac{1}{2}\|Hx + b - g\|^2$ and (3), which, in the denoising case, is the well-known Rudin-Osher-Fatemi (ROF) model [6].

This model is extensively studied, and in the past years, a variety of algorithms have been proposed to solve it or its smoothed version in the nonnegatively constrained or in the unconstrained case: see for example [7, 8, 9, 10, 11, 12, 13, 14, 15, 16] and reference therein, while a survey of the available software can be found in [17].

For the smoothed ROF model, some authors discuss also the problem of estimating a correct value of the artificial parameter δ with respect to the numerical stability of the methods [7] and to its effects on the restored images [16].

At the best of our knowledge, the KL-TV model has been proposed first in [18] for SPECT image reconstruction and it has been recently motivated as the more suited for Poisson data for denoising and deblurring problems [19, 20]. Algorithms based

on the Expectation Maximization (EM) method [18, 21] and on Bregman iterations [22, 23, 24, 25, 26, 27] have been proposed. For the model KL-HS, we mention the projected quasi–Newton method proposed in [28] and the very efficient Scaled Gradient Projection (SGP) method, described in [29]. In [29, 30], the influence of the parameter δ on the restored images and its relation with the regularization parameter β are also investigated. However, the authors point out that for small values of δ the SGP method suffers from a very slow convergence and the restored image presents undesired oscillations.

The same degenerating behaviour has been observed also for others kinds of methods, specialized for smooth optimization problems [31, 32], when applied to the KL-HS model with small values of δ.

The aim of this paper is to propose a numerical method which is able to solve the nondifferentiable KL-TV problem as well as its smoothed version KL-HS in a numerically stable way, for all $\delta \geq 0$. Using this method we can compare the smoothed and unsmoothed KL-HS and KL-TV models, by investigating the actual effect of the artificial parameter $\delta \geq 0$ on denoising and deblurring problems. Furthermore we can use the proposed method to apply the discrepancy criterion proposed in [29, 30] for finding the regularization parameter.

The key point for introducing our method is the primal–dual (or saddle–point) equivalent formulation of problem (1) described in Section 2, which has the form

$$\min_{x \in X} \max_{y \in Y} F(x, y) \tag{5}$$

where X and Y are two appropriate constraint sets such that $\mathcal{D} = X \times Y$ is a closed and convex domain and F is a smooth convex-concave function.

We recall that a saddle–point problem is a monotone variational inequality problem (VIP) [33]. As a consequence, any numerical method proposed for monotone VIPs could be applied to solve (5).

One of the most effective approaches for the solution of a general monotone VIP is the class of extragradient–type methods [34, 35, 36], whose iterative schemes consist in projection steps with a suitable choice of the steplength parameter.

On the other side, a saddle–point problem has some specific features which could be exploited in order to design more effective numerical methods: in particular, the domain is a cartesian product of two sets.

With this in mind, in Section 3 we propose an extragradient scheme especially tailored for general smooth saddle–point problems, where the iterates are updated in an alternating, or Gauss-Seidel, way [37, 38] and the stepsize parameter can be adaptively computed. The convergence analysis for the proposed method is developed and a practical implementation of the method, suitable for KL-HS or KL-TV problems, is presented.

The results of a numerical experience on image deblurring and denoising problems with KL-HS and KL-TV models are reported in Section 4. In Section 5 we drawn some conclusions.

2. Primal–dual formulation

In this section we derive the primal–dual formulation of the KL-TV and KL-HS models. Any two dimensional image of $N \times N$ pixels can be identified with a vector in $\mathbb{R}^n$, $n = N^2$ obtained by stacking the pixels columnwise. For sake of simplicity we consider a two dimensional square image $x \in \mathbb{R}^{N \times N}$ with periodic boundary conditions, but the following remarks can be extended also to the 3D case and to rectangular images with Neumann or reflective boundary conditions (see [39] for details).

First, we observe that the discrete gradient of x at the pixel $k = (j-1)N + i$, for $i, j = 1, ..., N$ can be defined by the forward finite difference formula as follows

$$(\nabla x)_k = \begin{pmatrix} x_{(j-1)N+mod(i,N)+1} & - & x_k \\ x_{mod(j,N)N+i} & - & x_k \end{pmatrix} = A_k x \qquad (6)$$

where $A_k \in \mathbb{R}^{2 \times n}$ is a matrix with only two nonzero entries on each row, equal to -1 and 1.

When we consider the discrete primal–dual version of the smoothed and unsmoothed total variation functional [8], the KL-TV or KL-HS problem can be reformulated as

$$\min_{x \in X} \max_{y \in Y} F(x, y) \equiv \sum_k \left\{ g_k \ln \frac{g_k}{(Hx+b)_k} + (Hx+b)_k - g_k \right\} + \beta y^T z(x) \qquad (7)$$

where $X = \{x \in \mathbb{R}^n : x \geq \eta\}$ and $z(x)$ and Y in the case (3) are given by

$$z(x) = Ax, \quad A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix}, Y = \{y \in \mathbb{R}^{2n} : \sqrt{y_{2i-1}^2 + y_{2i}^2} \leq 1, i = 1, ..., n\} (8)$$

while for (4) we have

$$z(x) = \begin{pmatrix} Ax \\ \delta e_n \end{pmatrix}, \quad e_n = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n, \qquad (9)$$

$$Y = \{y \in \mathbb{R}^{3n} : \sqrt{y_{2i-1}^2 + y_{2i}^2 + y_{2n+i}^2} \leq 1, i = 1, ..., n\}.$$

The functional F in (7) is convex with respect to x and concave with respect to y. In addition, under the assumption that e_n does not belong to the null space of H, the objective function of (1) is coercive. Thus, we can restrict the variable x in a bounded subset of X and invoke the min–max theorem [40, p.397] to ensure the existence of a solution of (7). A similar analysis holds also for the least squares data fidelity term [13, 14, 41].

A solution $v^* = (x^*, y^*)$ of (5) satisfies also the following VI problem

$$(v - v^*)^T \Phi(v^*) \geq 0 \qquad \forall v \in \mathcal{D}, \qquad (10)$$

where $\mathcal{D} = X \times Y$, $\Phi(v) \equiv \Phi(x, y) = (\nabla_x F(x,y)^T, -\nabla_y F(x,y)^T)^T$ and the fixed point equation

$$v^* = P_{\mathcal{D}}(v^* - \alpha \Phi(v^*)) \quad \forall \alpha > 0,$$

where P denotes the orthogonal projection operator. We observe that, thanks to the separable structure of the domain $\mathcal{D}$, the last equality holds if and only if

$$
\begin{aligned}
x^* &= P_X(x^* - \alpha \nabla_x F(x^*, y^*)) \\
y^* &= P_Y(y^* + \alpha \nabla_y F(x^*, y^*))
\end{aligned}
\tag{11}
$$

(see [33] for an overview on the VIPs).

From the computational point of view, the projections onto the sets X and Y are both easy to compute: the first one is a simple thresholding, while the projection onto Y of a vector y is defined as $(P_Y(y))_k = s_k y_k$, with

$$
s_{2i-1} = s_{2i} = \frac{1}{\max(1, \sqrt{y_{2i-1}^2 + y_{2i}^2})}, \quad i = 1, ..., n
$$

when Y is defined as in (8) , while for the case (9) we have

$$
s_{2i-1} = s_{2i} = s_{2n+i} = \frac{1}{\max(1, \sqrt{y_{2i-1}^2 + y_{2i}^2 + y_{2n+i}^2})}, \quad i = 1, ..., n
$$

The simple structure of the constraints in (7) makes the projection methods attractive and many variants are available in the literature for general monotone VIPs.

However, the effectiveness of a projection method is strictly related to the choice of the steplength parameter. There are two main strategies to choose it: it can be held fixed and usually its value depends on an *a priori* estimate of the Lipschitz constant of the vector field Φ; otherwise it can be adaptively updated so that sufficient conditions for the convergence of the method are satisfied. The Lipschitz condition is often a sufficient condition for the convergence. In practice, the choice of the steplength from the Lipschitz constant could yield very short steps and, in addition, estimating the Lipschitz constant could be difficult. This motivates us to adopt the second strategy; in particular, we consider the extragradient method [34, 35, 36], which is defined by the following iteration

$$
\begin{aligned}
\bar{v}^{(k)} &= P_{\mathcal{D}}(v^{(k)} - \alpha_k \Phi(v^{(k)})) \\
v^{(k+1)} &= P_{\mathcal{D}}(v^{(k)} - \alpha_k \Phi(\bar{v}^{(k)}))
\end{aligned}
\tag{12}
$$

where the steplength α_k satisfies

$$
1 - \alpha_k^2 \frac{\|\Phi(v^{(k)}) - \Phi(\bar{v}^{(k)})\|^2}{\|v^{(k)} - \bar{v}^{(k)}\|^2} > 0.
$$

The previous condition is implementable and, in practice, the value of α_k is determined by a backtracking procedure, guaranteeing the convergence of the method under suitable monotonicity and Lipschitz conditions on Φ (see [36] and reference therein).

However, as remarked also in [13], the above general form of the extragradient method does not capture the special features of the problem (7). In the next section we propose a variant of (12) where the computation of the iterate $v^{(k)}$ is split in two successive updates of the component $x^{(k)}$ and $y^{(k)}$.

3. An alternating extragradient method

The method we present in this section is a generalization of [37, 38], allowing an adaptive choice of the steplength. Indeed we do not make explicit use of the Lipschitz constant of the gradient, but we compute only a local approximation of it, as in the Khobotov's method [35, 36].
The method is very simple and it is described by the following iteration formulae

$$\bar{y}^{(k)} \quad = P_Y(y^{(k)} + \alpha_k \nabla_y F(x^{(k)}, y^{(k)})) \tag{13}$$

$$x^{(k+1)} = P_X(x^{(k)} - \alpha_k \nabla_x F(x^{(k)}, \bar{y}^{(k)})) \tag{14}$$

$$y^{(k+1)} = P_Y(y^{(k)} + \alpha_k \nabla_y F(x^{(k+1)}, y^{(k)})) \tag{15}$$

In the following section we will prove the convergence of the previous scheme to a saddle point of F, when the steplength α_k is chosen in a bounded interval $[\alpha_{min}, \alpha_{max}]$ with $0 < \alpha_{min} < \alpha_{max}$ and

$$\begin{cases} 1 - 2\alpha_k A_k - 2\alpha_k^2 B_k^2 & \geq \quad \epsilon \\ 1 - 2\alpha_k C_k & \geq \quad \epsilon \end{cases} \tag{16}$$

where $0 < \epsilon < 1$ is constant and

$$\begin{aligned} A_k &= \frac{\|\nabla_x F(x^{(k+1)}, \bar{y}^{(k)}) - \nabla_x F(x^{(k)}, \bar{y}^{(k)})\|}{\|x^{(k+1)} - x^{(k)}\|} \\ B_k &= \frac{\|\nabla_y F(x^{(k+1)}, y^{(k)}) - \nabla_y F(x^{(k)}, y^{(k)})\|}{\|x^{(k+1)} - x^{(k)}\|} \\ C_k &= \frac{\|\nabla_y F(x^{(k+1)}, y^{(k)}) - \nabla_y F(x^{(k+1)}, \bar{y}^{(k)})\|}{\|y^{(k)} - \bar{y}^{(k)}\|} \end{aligned} \tag{17}$$

A practical implementation of the method (13)–(15) is given in Algorithm AEM. We remark that this scheme consists in successive gradient ascent (13) and descent (14) steps followed by an extragradient step (15). We note that the method can also be written by exchanging the role of the primal and dual variables x, y.

The saddle–point formulation of a TV regularization problem has been analyzed in the recent literature for the ROF model, and numerical methods able to solve it have been proposed for example in [13, 41, 42, 43]. At the best of our knowledge, this formulation is not yet studied for KL-TV or KL-HS problems. Recent papers deal with primal or dual formulation of these problems [18, 21, 22, 23, 24, 25, 26, 27, 29]. We believe that the scheme (13)–(15) has two main advantages making it attractive for the solution of the KL-TV and KL-HS problems. First, it is defined only by means of explicit steps. Indeed, many of the mentioned methods require the solution of a minimum problem at each iteration. Such solution has a closed form representation when the term $f_0(x)$ in (1) is the least squares functional, but this may not be true for the KL discrepancy. Secondly, the convergence of the whole scheme is guaranteed, without strict convexity assumptions, by means of an appropriate selection of the steplength α_k, which can be done adaptively. In particular, as we will see more in detail in the following section, it

(a)   (b)
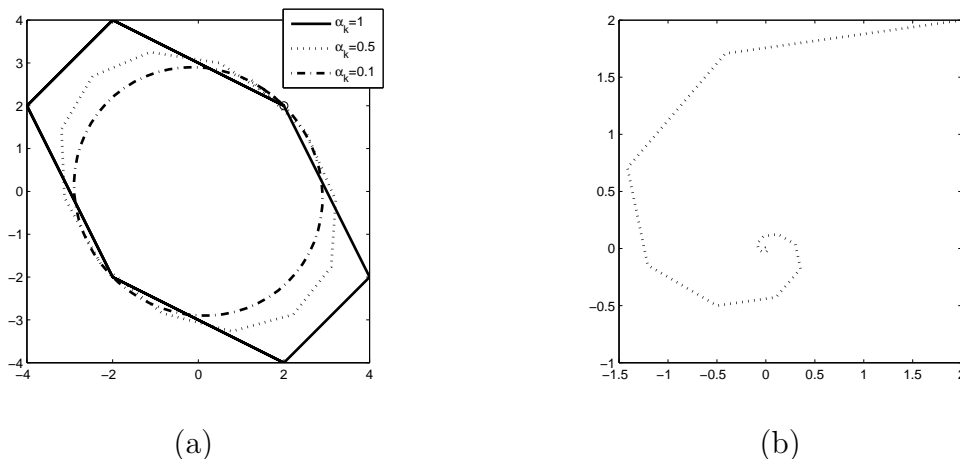
Figure 1. $F(x, y) = xy$. (a) Trajectory of the iterates generated by the alternating gradient ascent–descent method $y^{(k+1)} = P_Y(y^{(k)} + \alpha_k x^{(k)})$, $x^{(k+1)} = P_X(x^{(k)} - \alpha_k y^{(k+1)})$. (b) Iterates of the method (13)–(15) with $\alpha_k = (\sqrt{1-\epsilon})/2$ which satisfies (16) ($\epsilon = 10^{-4}$; $A_k = C_k = 0$, $B_k = 1$ for all k in this example).

is possible to implement a self-adjusting backtracking procedure to compute it.

Our approach is different from the one in [13, 41], where the steplengths in the ascent and descent directions are chosen *a priori* as suitable sequences. Furthermore, we observe that a strict convexity assumption is crucial for the convergence of several primal–dual methods [16, 42, 44]. In our method this assumption is not required, thanks to the extragradient step (15). Indeed, for general convex–concave functionals over bounded domains the successive alternation of projected gradient ascent–descent steps (13)–(14), with $y^{(k+1)} = \bar{y}^{(k)}$ is not convergent. This fact could be easily verified on a simple counterexample [37, 38] with $F(x, y) = xy$ and $\mathcal{D} = [-10, 10]^2 \subset \mathbb{R}^2$: projected gradient ascent–descent steps generate sequences which do not converge to the unique saddle point $(0, 0)$ and, for small values of α_k, the iterates approximate a circular trajectory around the origin (see Figure 1).

On the contrary, as explained in the following section, the method (13)–(15) equipped with a suitable choice of α_k has strong convergence properties for any convex–concave functional F whose gradient satisfies some Lipschitz conditions.

As we show in Section 3.2 this allows the application of the method to the KL-TV or KL-HS models in the formulation (7).

3.1. Convergence Analysis

In the subsequent analysis we will make use of the following two properties of the projection operator.

Lemma 3.1 *Let $\Omega \subseteq \mathbb{R}^n$ be a nonempty, closed, convex set, $w, z \in \mathbb{R}^n$, $u \in \Omega$.*

i) *[45] The projection operator is non-expansive*

$$\|P_\Omega(w) - P_\Omega(z)\| \le \|w - z\|$$

ii) *[36] The following inequality holds*

$$\|P_\Omega(z) - u\|^2 \le \|z - u\|^2 - \|P_\Omega(z) - z\|^2 \tag{18}$$

Now we give the complete proof of our main result for the scheme (13)–(14). Our proof generalizes the result in [38], and runs on less restrictive hypotheses on the parameter α_k, that in our approach can be adaptively chosen. Indeed we do not make explicit use of the Lipschitz constant of the gradient, but we compute only a local approximation of it, as in the Khobotov's method [36].

Theorem 1 *Assume that $F(x, y)$ is convex with respect to x and concave with respect to y in the domain $\mathcal{D} = X \times Y$, and that there exists a saddle point of F in $\mathcal{D}$. Let $\{(x^{(k)}, y^{(k)})\}$ be the sequence generated by the algorithm (13)–(15) where α_k is chosen in a bounded interval $[\alpha_{min}, \alpha_{max}]$ such that (16) holds. Then, $\{(x^{(k)}, y^{(k)})\}$ converges to a saddle point of F in $\mathcal{D}$.*

Proof. Let $(x^*, y^*) \in \mathcal{D}$ a saddle point of F. By applying Lemma 3.1 part ii) to (15), setting $z = y^{(k)} + \alpha_k \nabla_y F(x^{(k+1)}, y^{(k)})$ and $u = y^*$ we obtain

$$\begin{aligned}
\|y^{(k+1)} - y^*\|^2 &\le \\
&\le \|y^{(k)} + \alpha_k \nabla_y F(x^{(k+1)}, y^{(k)}) - y^*\|^2 - \|y^{(k+1)} - y^{(k)} - \alpha_k \nabla_y F(x^{(k+1)}, y^{(k)})\|^2 \\
&= \|y^{(k)} - y^*\|^2 + 2\alpha_k \langle \nabla_y F(x^{(k+1)}, y^{(k)}), y^{(k)} - y^* \rangle + \\
&\quad - \|y^{(k+1)} - y^{(k)}\|^2 + 2\alpha_k \langle \nabla_y F(x^{(k+1)}, y^{(k)}), y^{(k+1)} - y^{(k)} \rangle \\
&= \|y^{(k)} - y^*\|^2 - \|y^{(k+1)} - y^{(k)}\|^2 + 2\alpha_k \langle \nabla_y F(x^{(k+1)}, y^{(k)}), y^{(k+1)} - y^* \rangle
\end{aligned} \tag{19}$$

Similarly, from (13) and (18) with $z = y^{(k)} + \alpha_k \nabla_y F(x^{(k)}, y^{(k)})$ and $u = y^{(k+1)}$ we have

$$\begin{aligned}
\|\bar{y}^{(k)} - y^{(k+1)}\|^2 &\le \\
&\le \|y^{(k)} + \alpha_k \nabla_y F(x^{(k)}, y^{(k)}) - y^{(k+1)}\|^2 - \|\bar{y}^{(k)} - y^{(k)} - \alpha_k \nabla_y F(x^{(k)}, y^{(k)})\|^2 \\
&= \|y^{(k)} - y^{(k+1)}\|^2 + 2\alpha_k \langle \nabla_y F(x^{(k)}, y^{(k)}), y^{(k)} - y^{(k+1)} \rangle + \\
&\quad - \|\bar{y}^{(k)} - y^{(k)}\|^2 + 2\alpha_k \langle \nabla_y F(x^{(k)}, y^{(k)}), \bar{y}^{(k)} - y^{(k)} \rangle \\
&= \|y^{(k)} - y^{(k+1)}\|^2 - \|\bar{y}^{(k)} - y^{(k)}\|^2 + 2\alpha_k \langle \nabla_y F(x^{(k)}, y^{(k)}), \bar{y}^{(k)} - y^{(k+1)} \rangle
\end{aligned} \tag{20}$$

Adding and subtracting $2\alpha_k \langle \nabla_y F(x^{(k+1)}, y^{(k)}), \bar{y}^{(k)} - y^{(k+1)} \rangle$ to the right-hand side, one obtains

$$\begin{aligned}
\|\bar{y}^{(k)} - y^{(k+1)}\|^2 &\le \\
&\le \|y^{(k)} - y^{(k+1)}\|^2 - \|\bar{y}^{(k)} - y^{(k)}\|^2 + 2\alpha_k \langle \nabla_y F(x^{(k+1)}, y^{(k)}), \bar{y}^{(k)} - y^{(k+1)} \rangle \\
&\quad + 2\alpha_k \langle \nabla_y F(x^{(k)}, y^{(k)}) - \nabla_y F(x^{(k+1)}, y^{(k)}), \bar{y}^{(k)} - y^{(k+1)} \rangle
\end{aligned} \tag{21}$$

Summing (19) and (21) yields

$$
\begin{aligned}
\|y^{(k+1)} - y^*\|^2 \leq \\
\leq \|y^{(k)} - y^*\|^2 - \|\bar{y}^{(k)} - y^{(k+1)}\|^2 - \|\bar{y}^{(k)} - y^{(k)}\|^2 + \\
+ 2\alpha_k \langle \nabla_y F(x^{(k+1)}, y^{(k)}), \bar{y}^{(k)} - y^* \rangle + \\
+ 2\alpha_k \langle \nabla_y F(x^{(k)}, y^{(k)}) - \nabla_y F(x^{(k+1)}, y^{(k)}), \bar{y}^{(k)} - y^{(k+1)} \rangle
\end{aligned}
\tag{22}
$$

Since F is concave with respect to the variable y, we have

$$
\langle \nabla_y F(x^{(k+1)}, y^{(k)}), y^* - y^{(k)} \rangle \geq F(x^{(k+1)}, y^*) - F(x^{(k+1)}, y^{(k)})
$$

Then, adding and subtracting $2\alpha_k \langle \nabla_y F(x^{(k+1)}, y^{(k)}), y^{(k)} \rangle$, we have

$$
\begin{aligned}
\|y^{(k+1)} - y^*\|^2 \leq \\
\leq \|y^{(k)} - y^*\|^2 - \|\bar{y}^{(k)} - y^{(k+1)}\|^2 - \|\bar{y}^{(k)} - y^{(k)}\|^2 + \\
+ 2\alpha_k \langle \nabla_y F(x^{(k)}, y^{(k)}) - \nabla_y F(x^{(k+1)}, y^{(k)}), \bar{y}^{(k)} - y^{(k+1)} \rangle + \\
+ 2\alpha_k \langle \nabla_y F(x^{(k+1)}, y^{(k)}), \bar{y}^{(k)} - y^{(k)} \rangle + 2\alpha_k \{ F(x^{(k+1)}, y^{(k)}) - F(x^{(k+1)}, y^*) \}
\end{aligned}
\tag{23}
$$

Let now consider the definition of $x^{(k+1)}$ in (14). Invoking again (18) with $z = x^{(k)} + \alpha_k \nabla_y F(x^{(k)}, \bar{y}^{(k)})$ and $u = x^*$ we can write

$$
\begin{aligned}
\|x^{(k+1)} - x^*\|^2 \leq \\
\leq \|x^{(k)} - \alpha_k \nabla_x F(x^{(k)}, \bar{y}^{(k)}) - x^*\|^2 - \|x^{(k+1)} - x^{(k)} + \alpha_k \nabla_x F(x^{(k)}, \bar{y}^{(k)})\|^2 \\
= \|x^{(k)} - x^*\|^2 - \|x^{(k+1)} - x^{(k)}\|^2 + \\
+ 2\alpha_k \langle \nabla_x F(x^{(k)}, \bar{y}^{(k)}), x^* - x^{(k)} \rangle - 2\alpha_k \langle \nabla_x F(x^{(k)}, \bar{y}^{(k)}), x^{(k+1)} - x^{(k)} \rangle
\end{aligned}
\tag{24}
$$

By the convexity of F with respect to the variable x we have

$$
\langle \nabla_x F(x^{(k)}, \bar{y}^{(k)}), x^* - x^{(k)} \rangle \leq F(x^*, \bar{y}^{(k)}) - F(x^{(k)}, \bar{y}^{(k)})
$$

thus we can write

$$
\begin{aligned}
\|x^{(k+1)} - x^*\|^2 \leq \|x^{(k)} - x^*\|^2 - \|x^{(k+1)} - x^{(k)}\|^2 + \\
+ 2\alpha_k \{ F(x^*, \bar{y}^{(k)}) - F(x^{(k)}, \bar{y}^{(k)}) - \langle \nabla_x F(x^{(k)}, \bar{y}^{(k)}), x^{(k+1)} - x^{(k)} \rangle \}
\end{aligned}
\tag{25}
$$

Summing and subtracting $2\alpha_k F(x^{(k+1)}, \bar{y}^{(k)})$ to the right hand side and observing that the convexity implies also that

$$
F(x^{(k+1)}, \bar{y}^{(k)}) - F(x^{(k)}, \bar{y}^{(k)}) \leq \langle \nabla_x F(x^{(k+1)}, \bar{y}^{(k)}), x^{(k+1)} - x^{(k)} \rangle
$$

we obtain

$$
\begin{aligned}
\|x^{(k+1)} - x^*\|^2 \leq \|x^{(k)} - x^*\|^2 - \|x^{(k+1)} - x^{(k)}\|^2 + \\
+ 2\alpha_k \{ F(x^*, \bar{y}^{(k)}) - F(x^{(k+1)}, \bar{y}^{(k)}) + \\
\langle \nabla_x F(x^{(k+1)}, \bar{y}^{(k)}) - \nabla_x F(x^{(k)}, \bar{y}^{(k)}), x^{(k+1)} - x^{(k)} \rangle \}
\end{aligned}
\tag{26}
$$

Now we recall that the saddle point (x^*, y^*) satisfies

$$
F(x^*, \bar{y}^{(k)}) \leq F(x^*, y^*) \leq F(x^{(k+1)}, y^*)
$$

which leads to the following inequality

$$\|x^{(k+1)} - x^*\|^2 \leq$$
$$\leq \|x^{(k)} - x^*\|^2 - \|x^{(k+1)} - x^{(k)}\|^2 + 2\alpha_k\{F(x^{(k+1)}, y^*) - F(x^{(k+1)}, \bar{y}^{(k)})\}+ \quad (27)$$
$$+2\alpha_k\langle\nabla_x F(x^{(k+1)}, \bar{y}^{(k)}) - \nabla_x F(x^{(k)}, \bar{y}^{(k)}), x^{(k+1)} - x^{(k)}\rangle$$

Summing the inequalities (23) and (27) and adding and subtracting $2\alpha_k\langle\nabla_y F(x^{(k+1)}, \bar{y}^{(k)}), \bar{y}^{(k)} - y^{(k)}\rangle$ yields

$$\|x^{(k+1)} - x^*\|^2 + \|y^{(k+1)} - y^*\|^2 \leq$$
$$\leq \|x^{(k)} - x^*\|^2 + \|y^{(k)} - y^*\|^2 - \|x^{(k+1)} - x^{(k)}\|^2 - \|\bar{y}^{(k)} - y^{(k+1)}\|^2+$$
$$-\|\bar{y}^{(k)} - y^{(k)}\|^2+$$
$$+2\alpha_k\langle\nabla_x F(x^{(k+1)}, \bar{y}^{(k)}) - \nabla_x F(x^{(k)}, \bar{y}^{(k)}), x^{(k+1)} - x^{(k)}\rangle+$$
$$+2\alpha_k\langle\nabla_y F(x^{(k)}, y^{(k)}) - \nabla_y F(x^{(k+1)}, y^{(k)}), \bar{y}^{(k)} - y^{(k+1)}\rangle+$$
$$+2\alpha_k\{F(x^{(k+1)}, y^{(k)}) - F(x^{(k+1)}, \bar{y}^{(k)}) + \langle\nabla_y F(x^{(k+1)}, \bar{y}^{(k)}), \bar{y}^{(k)} - y^{(k)}\rangle\}+$$
$$+2\alpha_k\langle\nabla_y F(x^{(k+1)}, y^{(k)}) - \nabla_y F(x^{(k+1)}, \bar{y}^{(k)}), \bar{y}^{(k)} - y^{(k)}\rangle \quad (28)$$
$$\leq \|x^{(k)} - x^*\|^2 + \|y^{(k)} - y^*\|^2 - \|x^{(k+1)} - x^{(k)}\|^2 - \|\bar{y}^{(k)} - y^{(k+1)}\|^2+$$
$$-\|\bar{y}^{(k)} - y^{(k)}\|^2+$$
$$+2\alpha_k\langle\nabla_x F(x^{(k+1)}, \bar{y}^{(k)}) - \nabla_x F(x^{(k)}, \bar{y}^{(k)}), x^{(k+1)} - x^{(k)}\rangle+$$
$$+2\alpha_k\langle\nabla_y F(x^{(k)}, y^{(k)}) - \nabla_y F(x^{(k+1)}, y^{(k)}), \bar{y}^{(k)} - y^{(k+1)}\rangle+$$
$$+2\alpha_k\langle\nabla_y F(x^{(k+1)}, y^{(k)}) - \nabla_y F(x^{(k+1)}, \bar{y}^{(k)}), \bar{y}^{(k)} - y^{(k)}\rangle$$

where the last inequality follows from the concavity of F with respect to y. By the Cauchy-Schwartz inequality, we obtain

$$\|x^{(k+1)} - x^*\|^2 + \|y^{(k+1)} - y^*\|^2 \leq$$
$$\leq \|x^{(k)} - x^*\|^2 + \|y^{(k)} - y^*\|^2 - \|x^{(k+1)} - x^{(k)}\|^2 - \|\bar{y}^{(k)} - y^{(k+1)}\|^2+$$
$$-\|\bar{y}^{(k)} - y^{(k)}\|^2+$$
$$+2\alpha_k\|\nabla_x F(x^{(k+1)}, \bar{y}^{(k)}) - \nabla_x F(x^{(k)}, \bar{y}^{(k)})\|\|x^{(k+1)} - x^{(k)}\|+ \quad (29)$$
$$+2\alpha_k\|\nabla_y F(x^{(k)}, y^{(k)}) - \nabla_y F(x^{(k+1)}, y^{(k)})\|\|\bar{y}^{(k)} - y^{(k+1)}\|+$$
$$+2\alpha_k\|\nabla_y F(x^{(k+1)}, y^{(k)}) - \nabla_y F(x^{(k+1)}, \bar{y}^{(k)})\|\|\bar{y}^{(k)} - y^{(k)}\|$$

Now we recall that, since the projection operator is non-expansive, we can write

$$\|y^{(k+1)} - \bar{y}^{(k)}\| =$$
$$= \|P_Y(y^{(k)} + \alpha_k\nabla_y F(x^{(k+1)}, y^{(k)})) - P_Y(y^{(k)} + \alpha_k\nabla_y F(x^{(k)}, y^{(k)}))\| \quad (30)$$
$$\leq \alpha_k\|\nabla_y F(x^{(k+1)}, y^{(k)}) - \nabla_y F(x^{(k)}, y^{(k)})\|$$

Using the inequality (30) in (29) and recalling the definitions (17), we obtain

$$\|x^{(k+1)} - x^*\|^2 + \|y^{(k+1)} - y^*\|^2$$
$$\leq \|x^{(k)} - x^*\|^2 + \|y^{(k)} - y^*\|^2 - \|\bar{y}^{(k)} - y^{(k+1)}\|^2+$$
$$-(1 - 2\alpha_k C_k)\|\bar{y}^{(k)} - y^{(k)}\|^2+ \quad (31)$$
$$- \left(1 - 2\alpha_k A_k - 2\alpha_k^2 B_k^2\right)\|x^{(k+1)} - x^{(k)}\|^2$$

By the hypothesis (16), the coefficients $(1 - 2\alpha_k A_k - 2\alpha_k^2 B_k^2)$ and $(1 - 2\alpha_k C_k)$ are strictly positive and bounded away from zero. Thus, we must have that

$$\lim_k \|\bar{y}^{(k)} - y^{(k+1)}\| = 0$$

$$\lim_k \|\bar{y}^{(k)} - y^{(k)}\| = 0$$

$$\lim_k \|x^{(k+1)} - x^{(k)}\| = 0$$

Then there exists a point $(\bar{x}, \bar{y}) \in \mathcal{D}$ such that $x^{(k)}$ converges to $\bar{x}$ and $y^{(k)}$, $\bar{y}^{(k)}$ converge to $\bar{y}$. Consider now a subsequence $\{\alpha_{k_j}\}_j$ such that $\lim_j \alpha_{k_j} = \bar{\alpha} > 0$. Taking the limit for $j \to \infty$, by continuity of the projection operator and by the definition of the sequences $x^{(k)}$, $y^{(k)}$, $\bar{y}^{(k)}$, we have

$$\bar{y} = P_Y(\bar{y} + \bar{\alpha}\nabla_y F(\bar{x}, \bar{y}))$$
$$\bar{x} = P_X(\bar{x} - \bar{\alpha}\nabla_x F(\bar{x}, \bar{y}))$$

Thus, we can conclude that $(\bar{x}, \bar{y})$ is a saddle point of F. □

The previous theorem applies to every sequence defined as in (13)–(15), when condition (16) holds.

In other words, the scheme (13)–(15) is well defined and the previous theorem applies if a sequence $\{\alpha_k\}$ bounded away from zero satisfying (16) exists. We remark that the coefficients in the inequalities (16) can be regarded as a function of the steplength parameter

$$
\begin{aligned}
A_k(\alpha) &= \frac{\|\nabla_x F(x_k^+(\alpha), \bar{y}_k^+(\alpha)) - \nabla_x F(x^{(k)}, \bar{y}_k^+(\alpha))\|}{\|x_k^+(\alpha) - x^{(k)}\|} \\
B_k(\alpha) &= \frac{\|\nabla_y F(x_k^+(\alpha), y^{(k)}) - \nabla_y F(x^{(k)}, y^{(k)})\|}{\|x_k^+(\alpha) - x^{(k)}\|} \\
C_k(\alpha) &= \frac{\|\nabla_y F(x_k^+(\alpha), y^{(k)}) - \nabla_y F(x_k^+(\alpha), \bar{y}_k^+(\alpha))\|}{\|y^{(k)} - \bar{y}_k^+(\alpha)\|}
\end{aligned}
\tag{32}
$$

where

$$
\begin{aligned}
\bar{y}_k^+(\alpha) &= P_Y(y^{(k)} + \alpha\nabla_y F(x^{(k)}, y^{(k)})) \\
x_k^+(\alpha) &= P_X(x^{(k)} - \alpha\nabla_x F(x^{(k)}, y_k^+(\alpha)))
\end{aligned}
\tag{33}
$$

We devise the following sufficient conditions for the well posedness and convergence of the scheme (13)–(15).

Lemma 3.2 *Assume that there exists a saddle point $(x^*, y^*) \in \mathcal{D}$ of the convex-concave functional $F(x, y)$. If for every compact subset $K \subset \mathcal{D}$ there exists a positive constant L_K such that*

(A1) $\|\nabla_x F(x, y) - \nabla_x F(\bar{x}, y)\| \le L_K \|x - \bar{x}\|$
(A2) $\|\nabla_y F(x, y) - \nabla_y F(\bar{x}, y)\| \le L_K \|x - \bar{x}\|$
(A3) $\|\nabla_y F(x, y) - \nabla_y F(x, \bar{y})\| \le L_K \|y - \bar{y}\|$

for all $(x, y), (\bar{x}, y), (x, \bar{y}) \in K$, *then for any fixed* $\epsilon \in (0, 1)$ *there exists a positive number* $\bar{\alpha}_\epsilon > 0$ *such that*

$$
\begin{aligned}
1 - 2\alpha A_k(\alpha) - 2\alpha^2 B_k(\alpha)^2 &\geq \epsilon \\
1 - 2\alpha C_k(\alpha) &\geq \epsilon
\end{aligned}
\qquad \forall \alpha \in [0, \bar{\alpha}_\epsilon] \tag{34}
$$

for all k.

Proof. Let $(x^{(0)}, y^{(0)}) \in \mathcal{D}$ be the starting point of the algorithm. We define the following functionals for $(x, y) \in \mathcal{D}$ and $\alpha \in \mathbb{R}$

$$
\begin{aligned}
\phi(x, y, \alpha) &= P_Y(y + \alpha \nabla_y F(x, y)) \\
\psi(x, y, \alpha) &= P_X(x - \alpha \nabla_x F(x, \phi(x, y, \alpha)))
\end{aligned}
$$

Let $\rho_0^2 = \|x^* - x^{(0)}\|^2 + \|y^* - y^{(0)}\|^2$ and $B_{\rho_0}^*$ the ball with radius ρ_0 and center (x^*, y^*). Since ϕ and ψ are continuous with respect to their arguments, the following quantity is well defined

$$
\rho^2 = \max_{\substack{(x, y) \in B_{\rho_0}^* \cap \mathcal{D} \\ \alpha \in [0, \alpha_{max}]}} \|x^* - \psi(x, y, \alpha)\|^2 + \|y^* - \phi(x, y, \alpha)\|^2
$$

and $\rho \geq \rho_0$.

At the first iterate of the algorithm, we have that all the points $\bar{y}_0^+(\alpha)$ and $x_0^+(\alpha)$ defined in (33) for $\alpha \in [0, \alpha_{max}]$ belong to the compact set $B_\rho(x^*, y^*) \cap \mathcal{D}$.

By the assumption A1–A3, we can denote by $L \equiv L_K$ the Lipschitz constant in the set K.

Then, from (32) we have $A_0(\alpha) \leq L$, $B_0(\alpha) \leq L$ and $C_0(\alpha) \leq L$ for all $\alpha \in [0, \alpha_{max}]$. Since $1 - 2\alpha A_0(\alpha) - 2\alpha^2 B_0(\alpha)^2 \geq 1 - 2\alpha L - 2\alpha^2 L^2 \geq \epsilon$ and $1 - 2\alpha C_0(\alpha) \geq 1 - 2\alpha L > \epsilon$ when

$$
\alpha < \bar{\alpha}_\epsilon \equiv \min\left(\frac{\sqrt{3 - 2\epsilon} - 1}{2L}, \frac{1 - \epsilon}{2L} \right)
$$

the first step of the scheme (13)–(15) is well posed.

We observe now that, from (31) we have that $(x^{(1)}, y^{(1)}) \in B_{\rho_1}^* \cap \mathcal{D} \subset K$, with $\rho_1 < \rho_0$. Using the same arguments as before, by induction it follows that for all k conditions (16) are fulfilled for any $\alpha_k \in [0, \bar{\alpha}_\epsilon]$. $\qquad \square$

As a consequence of the previous lemma, there exists a sequence of parameters α_k satisfying (16) and such that $\alpha_k \geq \alpha_{min}$ (for example we may choose $\alpha_{min} = \frac{1}{2}\bar{\alpha}_\epsilon$ as lower bound). When α_k is bounded away from zero, Theorem 1 applies and the sequence generated by (13)–(15) converges to a saddle point of $F(x, y)$.

Algorithm AEM gives a practical realization of the alternating extragradient method (13)–(15).

Algorithm AEM Alternating Extragradient Method (AEM)

Choose the starting point $(x^{(0)}, y^{(0)}) \in \mathcal{D}$, set the parameters $\theta, \epsilon \in (0,1)$, $\alpha_{max} > 0$.

FOR $k = 0, 1, 2, ...$ DO THE FOLLOWING STEPS:

    STEP 1. Choose $\alpha \le \alpha_{max}$;

    STEP 2. Compute tentative points

$$\bar{y}^+ \leftarrow P_Y(y^{(k)} + \alpha \nabla_y F(x^{(k)}, y^{(k)}))$$

$$x^+ \leftarrow P_X(x^{(k)} - \alpha \nabla_x F(x^{(k)}, \bar{y}^+))$$

$$A \leftarrow \frac{\|\nabla_x F(x^+, \bar{y}^+) - \nabla_x F(x^{(k)}, \bar{y}^+)\|}{\|x^+ - x^{(k)}\|}$$

$$B \leftarrow \frac{\|\nabla_y F(x^+, y^{(k)}) - \nabla_y F(x^{(k)}, y^{(k)})\|}{\|x^+ - x^{(k)}\|}$$

$$C \leftarrow \frac{\|\nabla_y F(x^+, y^{(k)}) - \nabla_y F(x^+, \bar{y}^+)\|}{\|y^{(k)} - \bar{y}^+\|}$$

$$\bar{\alpha} \leftarrow \begin{cases} \min\{\frac{\sqrt{A^2 + 2B^2(1-\epsilon)} - A}{2B^2}, \frac{1-\epsilon}{2C}\} & \text{if } B > 0, C > 0 \\ \min\{\frac{1-\epsilon}{2A}, \frac{1-\epsilon}{2C}\} & \text{if } A > 0, C > 0, B = 0 \\ \frac{\sqrt{A^2 + 2B^2(1-\epsilon)} - A}{2B^2} & \text{if } B > 0, C = 0 \\ \frac{1-\epsilon}{2C} & \text{if } A = 0, C > 0, B = 0 \\ \frac{1-\epsilon}{2A} & \text{if } A > 0, C = 0, B = 0 \\ \alpha & \text{otherwise} \end{cases}$$

    STEP 3. Check convergence condition:

        IF $\alpha \le \bar{\alpha}$ THEN

            $\alpha_k = \alpha$;

            $\bar{y}^{(k)} = \bar{y}^+$;

            $x^{(k+1)} = x^+$;

        ELSE

            $\alpha \leftarrow \min(\bar{\alpha}, \theta\alpha)$;

            go to Step 2;

        ENDIF

    STEP 4. Set $y^{(k+1)} = P_Y(y^{(k)} + \alpha_k \nabla_y F(x^{(k+1)}, y^{(k)}))$.

END

The convergence condition $\alpha \le \bar{\alpha}$ at Step 3 is algorithmically equivalent to (16). The convergence properties of Algorithm AEM are stated in the following Corollary.

Corollary 1 *Assume that $F(x, y)$ is a convex-concave functional such that there exists a saddle point of F in $\mathcal{D}$. Furthermore, assume that the hypotheses (A1)–(A3) hold. Then, Algorithm AEM is well defined, and it generates a sequence converging to a saddle point of $F(x, y)$.*

Proof. In order to prove the well posedness of Algorithm AEM, we show that the loop between Step 2 and Step 3 terminates in a finite number of trials. Indeed, since α is reduced each time at least by a fixed amount θ, after a finite number of steps its value will be smaller than $\bar{\alpha}_\epsilon$ defined in Lemma 3.2 and, thus, the convergence condition is fulfilled.

This also yields that the sequence $\{\alpha_k\}$ generated by Algorithm AEM is bounded away from zero with $\alpha_{min} \geq \theta\bar{\alpha}_\epsilon$, thus the convergence result in Theorem 1 applies. $\square$

In our practical experience, we observed that the sequence $\{\alpha_k\}$ generated by Algorithm AEM tends to stick around a fixed value. For this reason, at the Step 1 we propose to adopt a quite conservative choice for the tentative value of α at the iterate $k \geq 1$, as the mean of the last M values

$$\alpha \leftarrow \min \left\{ \frac{1}{\min(k, M)} \sum_{j=1}^{\min(k,M)} \alpha_{k-j}, \quad \alpha_{max} \right\} \tag{35}$$

where M is a fixed integer. With this choice, we observed that a reduction of the steplength to fulfill the convergence condition is required a very few times.

In the next section we adapt Algorithm AEM to the special cases of the KL-TV or KL-HS models.

3.2. Application to the KL-TV or KL-HS models

As observed in subsection 2, for the functional $F(x, y)$ in (7) the hypotheses of Theorem 1 hold. The functional F is convex with respect to x and concave with respect to y. The domain Y is a compact set. Furthermore, under the assumption that e_n does not belong to the null space of H, the objective function of (1) is coercive and we can restrict the variable x in a bounded subset of X. Then, the min–max theorem in [40, p.397] ensures the existence of a solution of (7).

When $F(x, y)$ is defined as in (7), then we have

$$\nabla_x F(x, y) = e - H^T Z(x)^{-1} g + \beta A^T \tilde{y}$$
$$\nabla_y F(x, y) = \beta z(x)$$

where $\tilde{y} \equiv y$ in the TV model and $\tilde{y}$ represents the first $2n$ entries of the dual variable y in the HS model; furthermore, $Z(x)$ is a diagonal matrix whose diagonal entries are given by $(Hx + b)_i$.

Since $\nabla_y F(x, y) - \nabla_y F(x, \bar{y}) = 0$ for all $x \in X$, $y, \bar{y} \in Y$, we have $C_k = 0$ for all k. The Lipschitz constant of $\nabla_y F(\cdot, y)$ is the l_2 norm $\|A\|$ of matrix A, while Assumption A1 holds in every compact subset of the domain of the KL divergence. Then, Lemma 3.2 applies.

In order to define the set X, for the denoising case ($H = I$) we give a lower and an upper bound on some components of the solution, as stated in the following Lemma, which is a generalization of [32, Lemma 4.10].

Lemma 3.3 *Let x^* be a minimum point of $f(x) = f_0(x) + \beta f_1(x)$, $\beta > 0$, where $f_0(x)$ is the KL divergence with $H = I$ and $f_1(x)$ is defined in (3) or (4). Then, for all i such that $g_i > 0$ we have*

$$g_{min} \equiv \min_j \{g_j : g_j > 0\} \le x_i^* \le g_{max} \equiv \max_j g_j$$

Proof. Assume that there exists at least a component x_i^* such that $x_i^* > g_{max}$ or $x_i^* < g_{min}$ when $g_i > 0$ and define the vector $v \in \mathbb{R}^n$ such that

$$v_j = \begin{cases} g_{max} & \text{if } x_j^* > g_{max} \\ g_{min} & \text{if } x_j^* < g_{min} \text{ and } g_j > 0 \\ x_j^* & \text{otherwise} \end{cases}$$

Let us denote by $\mathcal{I}$ the set of indices such that $x_j^* = v_j$. Since the interval containing the components of v is included in that of x^*, then $f_1(v) \le f_1(x^*)$. Indeed, the functionals (3) and (4) can be written also as

$$f_1(x) = \sum_k \sqrt{\sum_{i \in \mathcal{N}_k} (x_i - x_k)^2 + \delta^2}, \ \delta \ge 0$$

where $\mathcal{N}_k$ contains the indices of the horizontal and vertical neighbours of the pixel x_k. By the assumption on x^*, there exists at least an index $k \notin \mathcal{I}$. Then, for $i \in \mathcal{N}_k$ we have the following possibilities,

$$\begin{array}{ll} x_k^* < g_{min} \le x_i^* \le g_{max} & x_k^* < g_{min} < g_{max} \le x_i^* \\ g_{min} \le x_i^* \le g_{max} < x_k^* & x_i^* \le g_{min}, x_k^* < g_{min} \\ x_i^* \le g_{min} < g_{max} < x_k^* & x_k^* > g_{max}, x_i^* \ge g_{max} \end{array}$$

which, by the definition of v, always imply $|v_i - v_k| \le |x_i^* - x_k^*|$ and, thus, $f_1(v) \le f_1(x^*)$. The last part of the proof consists in showing that $f_0(v) < f_0(x^*)$. When $H = I$, the KL divergence is separable and it can be written as $f_0(x) = \sum_i f_0^i(x_i)$, with $f_0^i(x_i) = g_i \log \frac{g_i}{x_i} + x_i - g_i$. For $i \in \mathcal{I}$ we have $f_0^i(x_i^*) = f_0^i(v_i)$. When $i \notin \mathcal{I}$, then we have two cases:

$$x_i^* < g_{min} = v_i \Rightarrow f_0^i(v_i) < f_0^i(x_i^*)$$
$$x_i^* > g_{max} = v_i \Rightarrow f_0^i(v_i) < f_0^i(x_i^*)$$

Then, we showed that $f(v) < f(x^*)$, which contradicts the minimum property of x^*. $\square$ The previous result can be employed to find the lower bound defining the constraint set X and the associated projection operator P_X in the denoising case:

$$\eta \in \mathbb{R}^n, \quad \eta_i = g_{min}, \text{ if } g_i > 0, \quad \eta_i = 0, \text{ if } g_i = 0, \quad i = 1, ..., n.$$

For deblurring problems $\eta = 0$. About the computational complexity of AEM, we observe that any iteration requires the following computations:

- evaluation of the gradient of the KL function;

- discrete approximation of the gradient operator $(Ax^{(k+1)})$;
- discrete approximation of the opposite of the divergence operator $(A^T \tilde{y}^{(k)})$;
- one projection on the set X;
- two projections on the set Y;

when a backtracking procedure is started, each step requires one additional evaluation of the gradient of the KL function, the computations of $Ax^{(k+1)}$ and $A^T \tilde{y}^{(k)}$ and two further projections on X and Y respectively.

Furthermore, we observe that for the saddle point (x^*, y^*), the following equalities hold

$$x^* = P_X(x^* - \alpha \nabla_x F(x^*, y^*))$$
$$y^* = P_X(y^* + \alpha \nabla_y F(x^*, y^*))$$

Thus, a significant stopping criterion for Algorithm AEM could be based on the relative difference between two successive iterates

$$\left\| \begin{pmatrix} x^{(k+1)} - x^{(k)} \\ y^{(k+1)} - y^{(k)} \end{pmatrix} \right\| / \left\| \begin{pmatrix} x^{(k+1)} \\ y^{(k+1)} \end{pmatrix} \right\| < \tau \tag{36}$$

where τ is a prefixed tolerance. Other stopping criteria could be defined upon the values of the primal objective function $f(x)$ and the primal–dual function $F(x, y)$. We adopt the criterion (36) since it can be also interpreted as a measure of the violation of the optimality conditions (11) at the current iteration.

4. Numerical Experience

4.1. Test Problems

This section is devoted to numerically evaluate the influence of the artificial parameter δ on the stability of some methods and on the features of the restored images. To this end we perform a set of numerical experiments in MATLAB environment, on a server with a dual Intel Xeon QuadCore E5620 processor at 2,40 GHz, 12 Mb cache and 18 Gb of RAM. In the experiments we consider a set of test-problems, where the Poisson noise has been simulated by the `imnoise` function in the Matlab Image Processing Toolbox. The considered test problems are described in the following.

Denoising problems

- LCR phantom: the original image is the phantom described in [20]; it is an array 256×256, consisting in concentric circles of intensities 70, 135 and 200, enclosed by a square frame of intensity 10, all on a background of intensity 5 (LCR-1). We can simulate a different noise level by multiplying the LCR phantom by a factor 10 (LCR-10) and 0.2 (LCR-0.2) and generating the corresponding noisy images. For each case, we generate 25 different realizations of noise: the mean of relative difference in l_2 norm between the noisy and the original images over the 25 realizations is 0.095318, 0.0300744, 0.212723 respectively.

- dental radiography (DR): the original image [29] is an array 512×512, with values in the range $[0, 255]$; for this test-problem, simulating a radiographic image obtained by a lower dose, the relative difference in l_2 norm between the noisy and the noise–free images is 0.179.

Deblurring problems

- *micro*: the original image is the confocal microscopy phantom of size 128×128 described in [46]; its values are in the range $[0, 70]$ and the total flux is $2.9461 \; 10^5$; the background term b in (2) is set to zero.
- *cameraman*: following [26], the simulated data are obtained by convolving the image 256×256 with a Gaussian psf with standard deviation $\sigma = 1.3$, then adding Poisson noise; the values of the original image are in the range $[0, 1000]$; the background term b in (2) is set to zero.

4.2. Stability of the numerical methods

In the first set of experiments we compare the numerical behavior of a method for smooth optimization, the SGP method in [29], with AEM on three denoising test problems LCR-1, LCR-10 and LCR-0.2, using the KL-HS model with values of δ of different order of magnitude.

To get an approximate solution $x^*_{\beta,\delta}$ of (7), we run AEM (with $M = 10$ in formula (35) and $\epsilon = 10^{-4}$) until the criterion (36) is satisfied with tolerance $\tau = 5 \; 10^{-7}$ for denoising and $5 \; 10^{-6}$ for deblurring problems or a maximum number of 5000 iterations is exceeded. The SGP parameters are $M = 1$ (monotone line–search), $\alpha_{min} = 10^{-5}$ and $\alpha_{max} = 10^5$ (see [29] for further implementation details). Following a strategy similar to that used in [29], the SGP iterations are stopped when the objective function value becomes smaller than the value $f(x^*_{\beta,\delta})$ obtained with AEM or after 5000 iterations. Both the algorithms have been initialized with $x^{(0)} = \max\{\eta, g\}$, where the maximum is intended componentwise and η is chosen as described in section 3.2. The initial guess for the dual variable $y^{(0)}$ in AEM has been set to zero.

We point out that, for $\delta = 10^{-8}$, the value of the objective function $f(x^*_{\beta,\delta})$ obtained by AEM with the stopping condition (36) is less than that related to the results shown for LCR–1 in Table 1 of [29]. This explains the difference between the value reported in our Table 1 and those in Table 1 of [29].

In Figures 2, 3, 4, we show the superpositions of the line–outs from row number 128 corresponding to the reconstructions related to the 25 different realizations of noise provided by SGP and AEM, using the values of (β, δ) employed in [29]. In Figure 5, we report the results obtained by applying AEM to the model KL-TV ($\delta = 0$).

In order to quantify the accuracy of the results and the effectiveness of the methods, in Table 1 we report the arithmetic mean over the 25 cases of some significant indices: the l_2 relative error (*mean-err*) defined as $\frac{\|\bar{x} - x\|}{\|x\|}$, where x is the original image and $\bar{x}$ is the reconstruction provided by the two methods; the mean elapsed time (*time*) in seconds

and the mean iterations number (*iter*) needed to satisfy the stopping criterion. Furthermore, in order to compare the convergence rate of SGP and AEM, we compute the ideal solution $\hat{x}$ of the considered KL-HS minimization problem, by running 100000 iterations of AEM. Then, we evaluate the progress toward the ideal solution at each iteration in terms of l_2 relative error $\frac{\|x^{(k)}-\hat{x}\|}{\|\hat{x}\|}$. It is noticed that computing the ideal solution with AEM makes a small bias in favour of AEM itself. However the obtained results reported in Figure 6 are sufficiently convincing to forget this bias. We observe that the behavior of the two methods is similar for a moderate value of δ, while for very small δ the iterates generated of the SGP method remain trapped near nonstationary points and the rate of convergence slows down. The value of the KL discrepancy at the computed solution for $(\beta, \delta) = (0.25, 10^{-1})$ is $f_0(x^*_{\beta,\delta}) = 2.7806 \cdot 10^4$ while $f(x^*_{\beta,\delta}) = 55566.4$ for both the methods SGP and AEM. For $(\beta, \delta) = (0.25, 10^{-8})$, we have $f_0(x^*_{\beta,\delta}) = 2.6426 \cdot 10^4$, $f(x^*_{\beta,\delta}) = 57581.1$ for SGP, while $f_0(x^*_{\beta,\delta}) = 2.8169 \cdot 10^4$, $f(x^*_{\beta,\delta}) = 54849.0$ for AEM.

The same experiment is performed also on the deblurring test problem micro. Figure 7 shows the superposition of line–outs of row number 64 corresponding to the reconstruction related to the 25 different realizations of noise provided by SGP and AEM, with $(\beta, \delta) = (10^{-8}, 0.56)$. Further results about this experiment are summarized in Table 1.

From this first set of experiments we can drawn the following considerations:

- the accuracy of the reconstructions provided by SGP and AEM applied to the KL-HS model are similar for moderately small value of δ;

- for very small values of δ, the reconstructions obtained by SGP exhibit a lot of slight oscillations, that however do not significantly affect the relative reconstruction error; this effect does not arise for AEM, that provides reconstructions of very similar accuracy for any value of δ, also for $\delta = 0$; the reconstruction errors of the results obtained by AEM are equal for $\delta = 10^{-8}$ and $\delta = 0$ on the first 13 significant digits;

- the AEM seems unaffected by the stagnation of the iterates that we can observe for the SGP method when we have to minimize KL-HS functional with δ close to 0.

4.3. Numerical evaluation of the model KL-HS

In the second set of experiments, we investigate the dependence of the l_2 relative reconstruction error on the pair (β, δ) for both denoising and deblurring problems. In particular, we evaluate the function

$$E(\beta, \delta) = \frac{\|x - x^*_{\beta,\delta}\|}{\|x\|}$$

where $x^*_{\beta,\delta}$ is the solution of the KL-HS model computed by AEM, on a grid of values in the (β, δ)–plane. We plot the level curves of $E(\beta, \delta)$ in Figure 8 and we report the pair (β, δ) corresponding to the minimum reconstruction error over the grid points in Table

2. For a comparison in Table 2 also the values $E(\beta, 0)$ are reported. In all cases, we observe that when δ is less than about 0.1% of the maximum value of the image, the l_2 relative reconstruction error does not significantly change. In Table 2 we show also the performance results of the AEM and SGP method in term of computational time (in seconds) and iterations number for the best tuning of (β, δ). The same information is reported in the case $(\beta, 0)$ for AEM. The stopping criteria are the same used for Table 1.

For smaller values of δ, the level curves become almost parallel to the δ axis, which means that the accuracy of the results depends only on the choice of β. These observations apply on both denoising and deblurring problems.

On the other hand, we can detect a difference about the effect of the regularization on the features of the solution. To focus this point, in Figure 9 we consider the deblurring test problem *micro* and we compare the plots of the row number 64 of the reconstructed images obtained by minimizing with AEM the KL-TV model with $\beta = 0.09$ and the KL-HS model with $(\beta, \delta) = (0.09, 0.1)$, which both give a reconstruction error of 9%. The optimal reconstruction is obtained with $(\beta, \delta) = (0.09, 0.56)$, with a reconstruction error of 8.8% (see Table 2). The complete images are shown in the lower panels of Figure 10. The TV regularization better emphasizes the edges, although a slight staircasing effect may occur, while the KL-HS model has the feature of being a smooth regularization without eroding too much the edges of the image.

When a smoothed edge–preserving regularization is convenient, the above analysis enables us to affirm that a value for δ slightly less than about the 1% of the maximum value of the image produces an high accuracy image; on the other hand, in this case, we can apply efficient methods for smooth optimization, as SGP, without undesired oscillating effects. Similar observations can be made for all the considered test problems. In particular, for the denoising test problem DR in Figure 11 we report the results obtained with $(\beta, \delta) = (0.27, 0.31)$, corresponding to the minimum reconstruction error, and $(\beta, \delta) = (0.27, 0)$.

For the LCR image, where different levels of Poisson noise are considered, from Figure 8 and Table 2 we point out that the values of β related to the minimum reconstruction error increase for growing levels of noise. On the contrary, the values of δ related to the minimum of $E(\beta, \delta)$ decrease for growing levels of noise, suggesting that, for very high noise, methods whose convergence behaviour is independent of δ must be used.

5. Conclusions

We have proposed an alternating extragradient method, which is able to solve in a numerically stable way the primal–dual formulation of the non differentiable image restoration problem involving the total variation (KL-TV) as well as its smoothed version (KL-HS) for data affected by Poisson noise. The method, tailored for general smooth saddle–point problems, consists in successive projected ascent and descent steps followed by an extragradient step with an adaptive choice of the stepsize parameter that
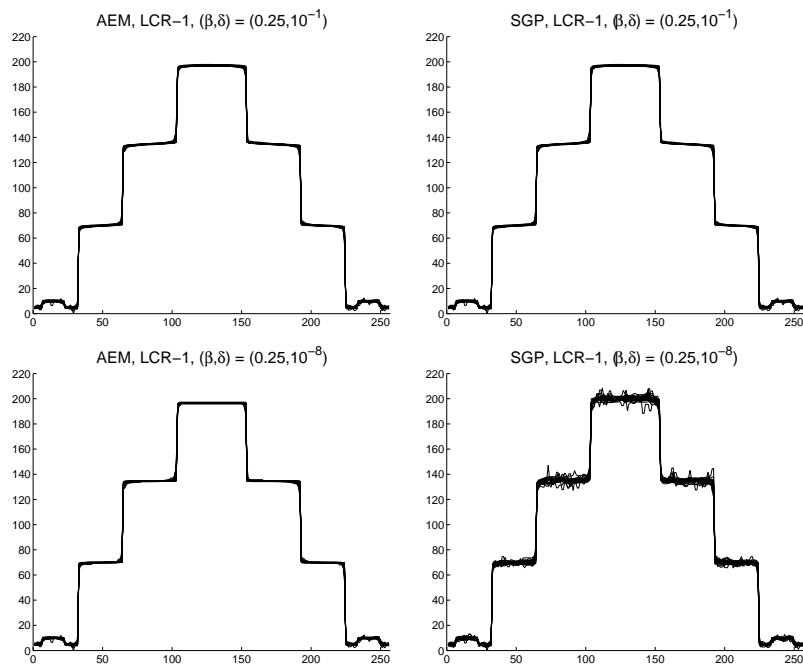
Figure 2. LCR-1: superposition of the line–outs from row number 128 for the reconstructions corresponding to 25 different realizations of noise. Left: AEM. Right: SGP. Top: $(\beta, \delta) = (0.25, 10^{-1})$. Bottom: $(\beta, \delta) = (0.25, 10^{-8})$.
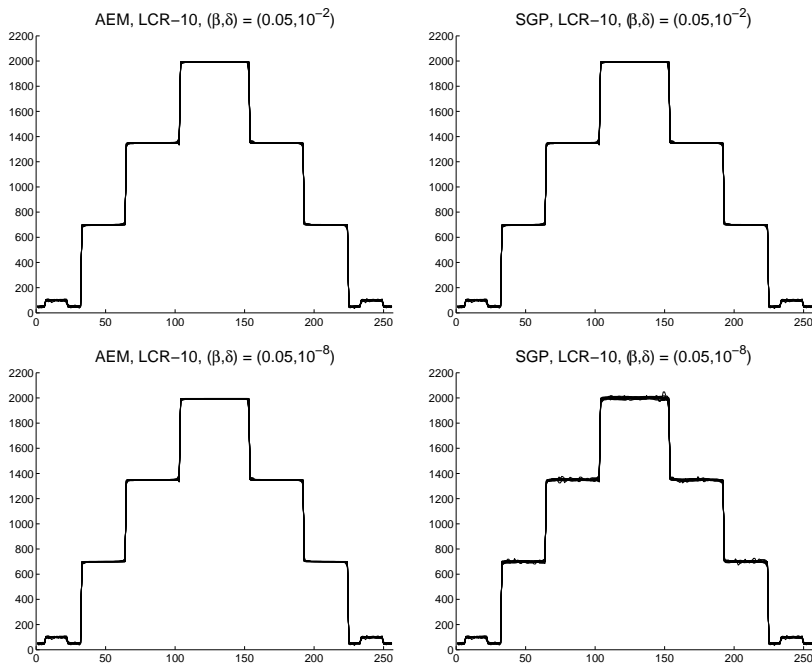
Figure 3. LCR-10: superposition of the line–outs from row number 128 for the reconstructions corresponding to 25 different realizations of noise. Left: AEM. Right: SGP. Top: $(\beta, \delta) = (0.05, 10^{-2})$. Bottom: $(\beta, \delta) = (0.05, 10^{-8})$.

Table 1. Mean values over 25 different noise realization of ℓ_2 reconstruction error, computational time and iterations number. The asterisk indicates that SGP fails to satisfy the stopping criterion within 5000 iterations, for all the 25 cases.

Problem	β	δ	AEM			SGP		
			mean err.	time	iter	mean err.	time	iter
	0.25	10^{-1}	0.0254501	34.2	631.8	0.0254598	14.0	323.7
	0.2	10^{-2}	0.0236812	41.1	785.2	0.0236811	71.8	1716.5
LCR-1	0.25	10^{-8}	0.0246004	50.8	972.2	0.027646*	195.4*	5000*
	0.25	0	0.0246004	50.9	972.2			
	0.05	10^{-1}	0.0084788	49.0	802.8	0.0084879	24.4	508.4
	0.05	10^{-2}	0.0084521	57.4	833.5	0.0084697	130.8	3030.4
LCR-10	0.05	10^{-8}	0.0084537	50.2	871.0	0.0095699*	204.0*	5000*
	0.05	0	0.0084537	53.4	871.0			
	0.575	10^{-1}	0.0500644	24.9	400.0	0.0500754	38.8	740.2
	0.625	10^{-2}	0.0455512	79.8	1171.3	0.0455511	100.8	2381.5
LCR-0.2	0.575	10^{-8}	0.0446964	280.4	4293.3	0.0543569*	302.9*	5000*
	0.575	0	0.0446964	272.3	4293.3			
	0.09	0.56	0.0873963	73.0	2249.2	0.0877281	5.2	202.4
	0.09	10^{-2}	0.089910	90.0	2815.4	0.0907422	12.2	548.0
micro	0.09	10^{-8}	0.090006	91.2	2831.2	0.913585*	142.1*	5000*
	0.09	0	0.090006	93.4	2831.2			

Table 2. Position of the minimum of $E(\beta, \delta)$ for KL-HS model over the grid points and $E(\beta, 0)$ for KL-TV and performance of the methods AEM and SGP. The asterisk indicates that the method fails to satisfy the stopping criterion within 5000 iterations.

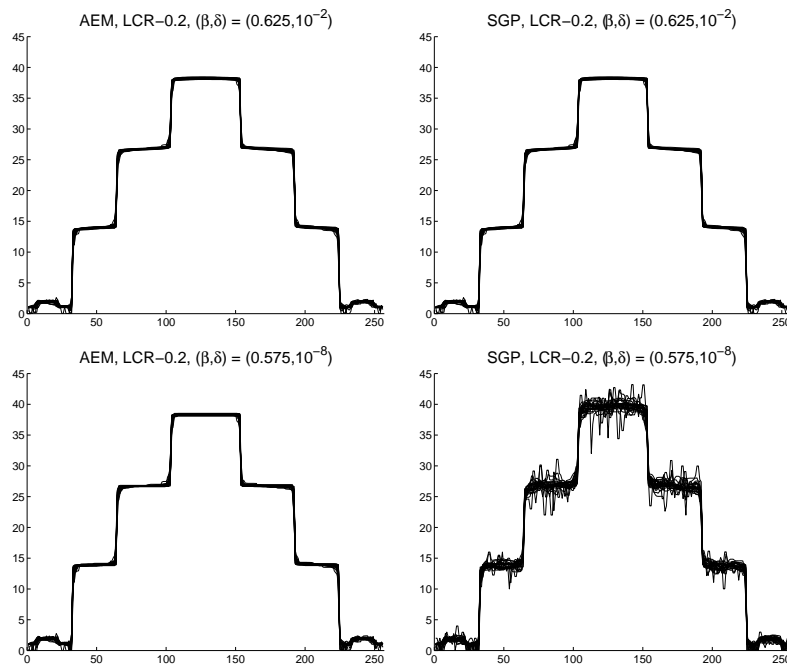|	β	δ	err	AEM		SGP	
				time	iter	time	iter
LCR-1	0.172	$10^{-3.75}$	0.0238555	113.28	1970	255.68	5000*
	0.172	0	0.0238557	49.81	889		
LCR-10	0.041	10^{-2}	0.00824262	39.99	712	133.24	2935
	0.041	0	0.00824303	41.06	741		
LCR-0.2	0.55	$10^{-4.5}$	0.0447176	283.08	5000*	258.7	5000*
	0.55	0	0.0447178	228.27	4148		
DR	0.27	$10^{-0.5}$	0.0276958	50.92	210	20.7	112
	0.27	0	0.0296606	516.8	2206		
micro	0.09	$10^{-0.25}$	0.0880993	58.61	1846	5.99	209
	0.09	0	0.0912806	78.78	2492		
cameraman	0.0045	$10^{0.25}$	0.0869088	132.65	1495	6.1	81
	0.0045	0	0.087225	141.57	1587		

Figure 4. LCR-0.2: superposition of the row number 128 for the reconstructions corresponding to 25 different realizations of noise. Left: AEM. Right: SGP. Top: $(\beta, \delta) = (0.625, 10^{-2})$. Bottom: $(\beta, \delta) = (0.575, 10^{-8})$.
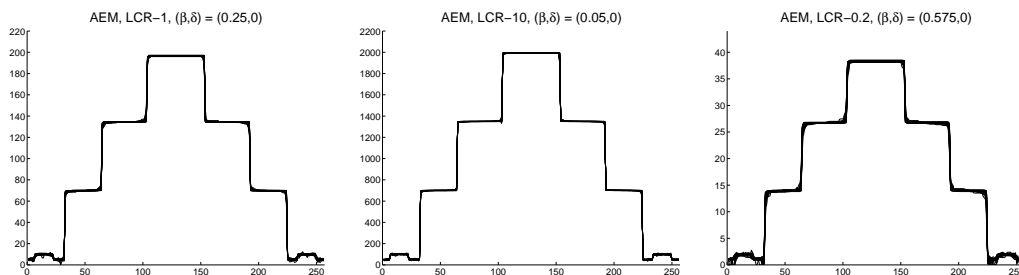
Figure 5. AEM reconstructions on the KL-TV model: superposition of row number 128 of the reconstructions corresponding to 25 different realizations of noise. Left: LCR-1, $\beta = 0.25$. Middle: LCR-10, $\beta = 0.05$. Right: LCR-0.2, $\beta = 0.575$.

can be computed by a self–adjusting backtracking procedure. The convergence of the scheme is proved under mild assumptions and a practical implementation is given. Then, the application of the method is discussed for the KL-TV and KL-HS models, where the simple structure of the constraints allows an easy implementation the projection steps. Furthermore, for denoising problems, we give a lower and upper bound on the components of the solution. By a set of numerical experiments concerning denoising and deblurring problems, we compare the stability of the proposed method with a method for smooth optimization as SGP and the features of the restored images, with respect to different choices of the smoothing parameter δ. We observe that the AEM seems
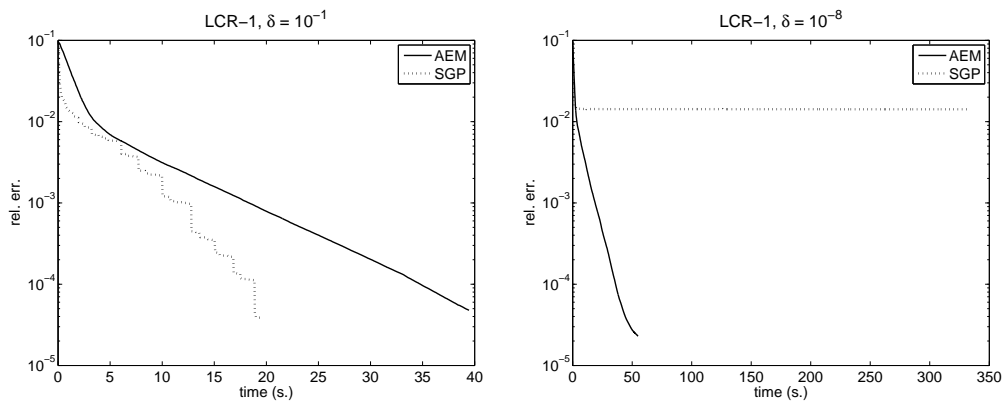
Figure 6. LCR-1: l_2 relative distance of each iteration from to the ideal solution on the KL-HS model with $(\beta, \delta) = (0.25, 0.1)$ (left panel) and $(\beta, \delta) = (0.25, 10^{-8})$ (right panel).
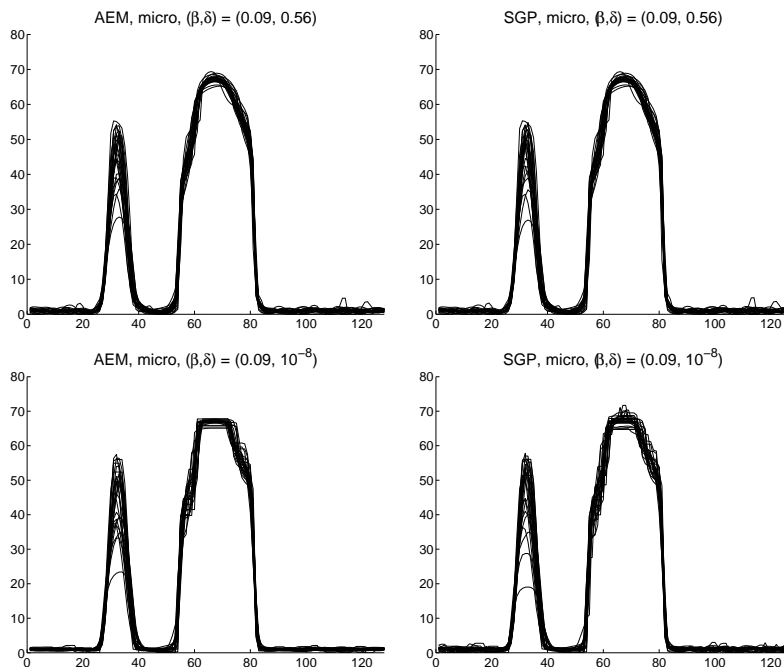

Figure 7. micro: superposition of the row number 64 for the reconstructions corresponding to 25 different realizations of noise. Left: AEM. Right: SGP. Top: $(\beta, \delta) = (0.09, 0.56)$. Bottom: $(\beta, \delta) = (0.09, 10^{-8})$.

unaffected by the stagnation of the iterates arising when we apply SGP to the KL-HS functional with δ close to 0. Then, we employ AEM to evaluate the accuracy of the restored images, in terms of relative reconstruction error, with respect to the pair of parameters (β, δ). When δ is less than 0.1% of the maximum image value, the l_2 relative reconstruction error does not significantly change and the accuracy depends only from the choice of β. Nevertheless, the aspect of the restored images is different, since the TV
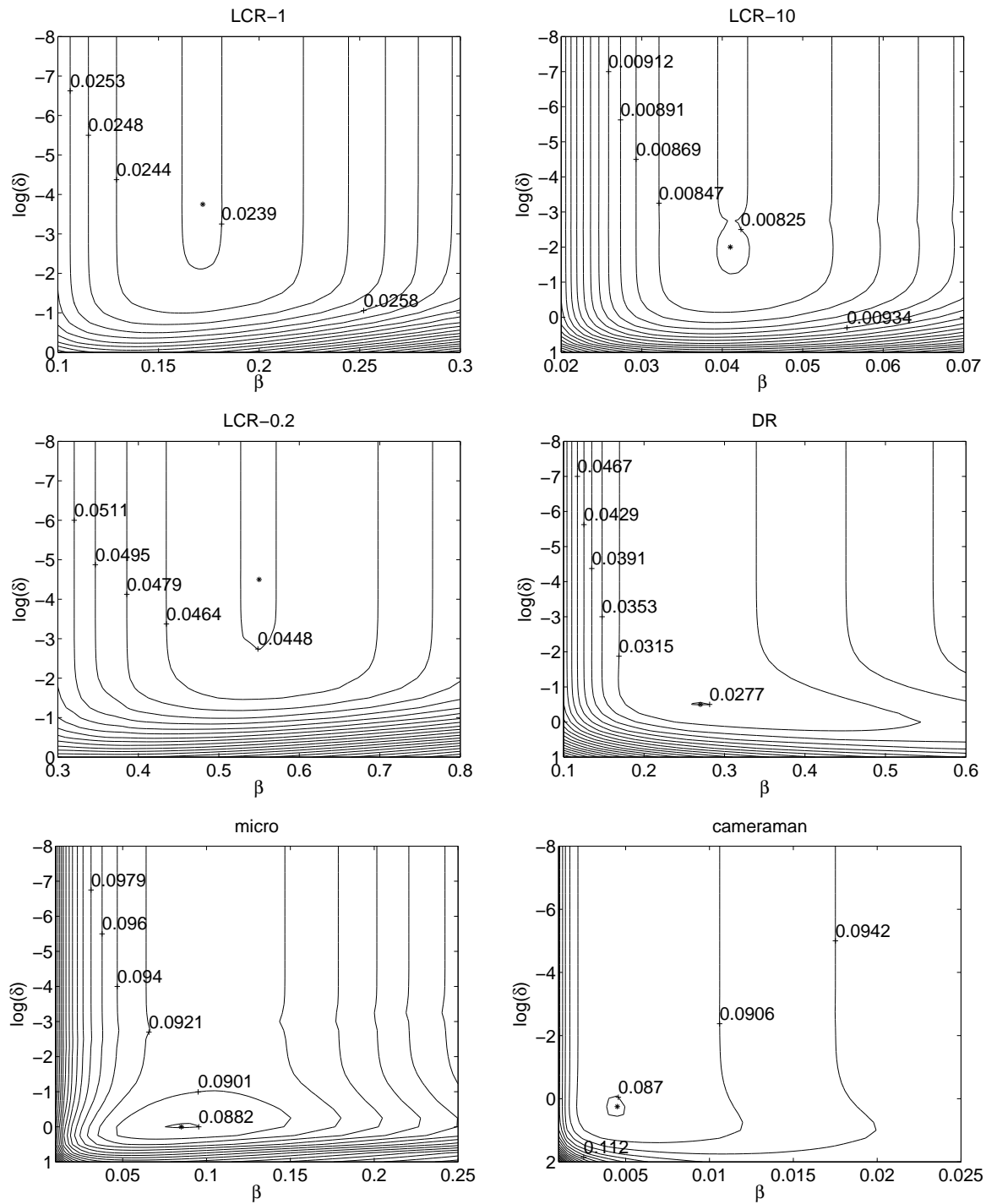
Figure 8. Level curves of $E(\beta, \delta)$, obtained using AEM to compute $x^*_{\beta,\delta}$. Top left: LCR-1. Top right: LCR-10. Middle left: LCR-0.2. Middle right: DR. Bottom left: *micro*. Bottom right: *cameraman*. The asterisk indicates the position where $E(\beta, \delta)$ takes a minimum over the grid points.
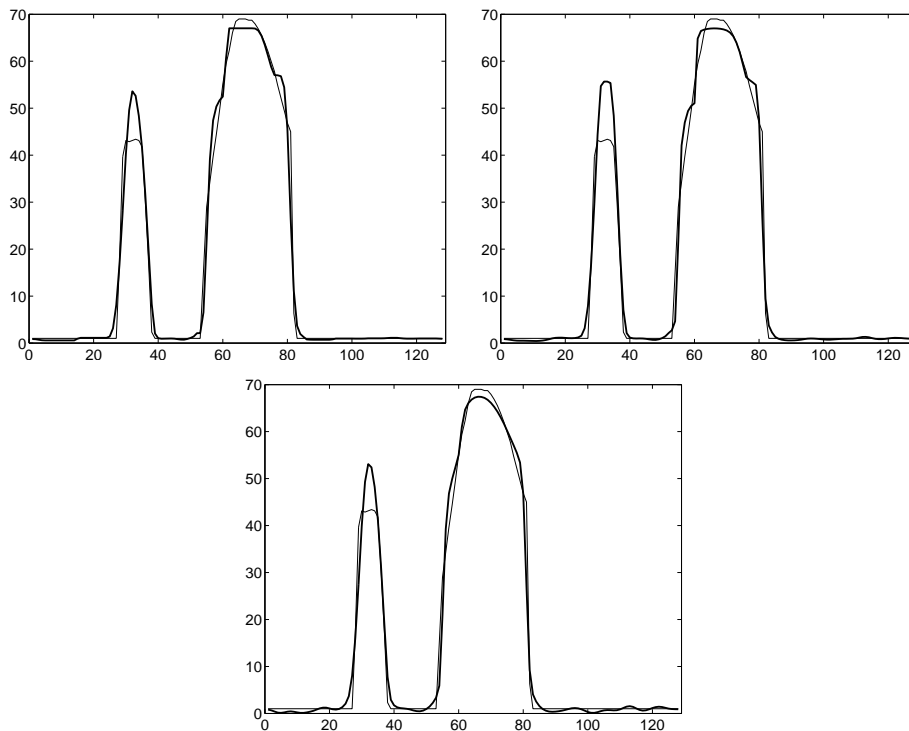
Figure 9. *micro*; the solid black lines plot the row number 64 of the solution of the KL-TV model (top left) with $\beta = 0.09$ and of the KL-HS model with $(\beta, \delta) = (0.09, 10^{-1})$ (top right) and $(\beta, \delta) = (0.09, 0.56)$ (bottom). The thin black line is the profile of row number 64 of the original image.
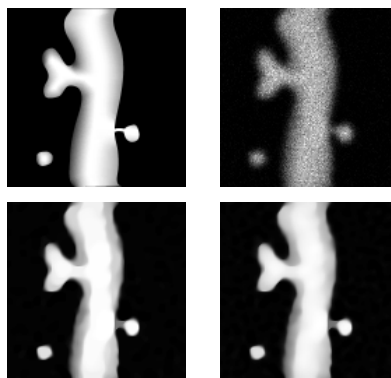

Figure 10. Test-problem *micro*. Top left: original image. Top right: blurred noisy image. Bottom left: solution of the KL-TV problem with $\beta = 0.09$. Bottom right: solution of the KL-HS model with $(\beta, \delta) = (0.09, 0.56)$.
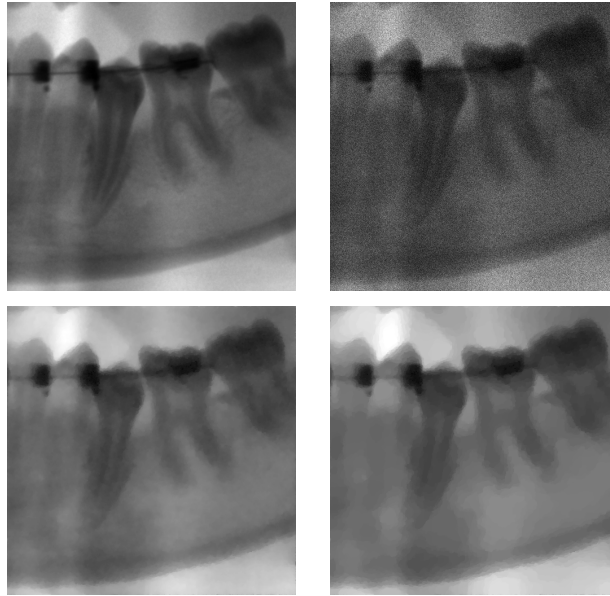
Figure 11. Test-problem DR. Top left: original image. Top right: blurred noisy image. Bottom left: solution of the KL-TV problem with $\beta = 0.28$. Bottom right: solution of the KL-HS model with $(\beta, \delta) = (0.28, 0.31)$ (minimum reconstruction error).

regularization realizes cartoon reconstructions with sharp edges where some staircasing effect can occur, while the KL–HS model has the feature of being a smooth regularization without eroding too much the edges. When a smoothed edge–preserving regularization is convenient, efficient methods for smooth optimization, as SGP, can be successfully applied, using values of δ that produce accurate results without encountering oscillating effects due to the stagnation of the iterations.

Since the practical convergence of AEM has been shown on both smooth and nonsmooth models, our future work will concern with the comparison in terms of accuracy and computational efficiency between AEM and other methods, especially tailored for each model.

Acknowledgments

References

[1] Bertero M, Boccacci P, Desiderà G, and Vicidomini G. Image deblurring with Poisson data: from cells to galaxies. *Inverse Problems*, 25:123006, 2009.
[2] Shepp L A and Vardi Y. Maximum likelihood reconstruction for emission tomography. *Trans. Med. Imaging*, MI–1:113–122, 1982.

[3] Geman S and Geman D. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[4] Vogel C. *Computational methods for inverse problems*. SIAM, Philadelphia, 2002.

[5] Charbonnier P, Blanc-Féraud L, Aubert G, and Barlaud A. Deterministic edge–preserving regularization in computed imaging. *IEEE Trans. Image Processing*, 6, 1997.

[6] Rudin L, Osher S, and Fatemi E. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[7] Chan T F, Golub G H, and Mulet P. A nonlinear primal–dual method for Total Variation based image restoration. *SIAM J. Sci. Comput.*, 20:1964–1977, 1999.

[8] Chambolle A. An algorithm for Total Variation minimization and applications. *J. Math. Imag. Vis.*, 20:89–97, 2004.

[9] Chambolle A. Total variation minimization and a class of binary mrf models. *EMMCVPR 05. Lecture Notes in Computer Sciences*, 3757:136–152, 2005.

[10] Goldfarb D and Win W. Second–order cone programming methods for total variation based image restoration. *SIAM J. Sci. Comput.*, 27(2):622–645, 2005.

[11] Osher S, Burger M, Goldfarb D, Xu J, and Yin W. An iterative regularization method for total variation-based image restoration. *SIAM Journal on Multiscale Modeling and Simulation*, 4(2):460–489, 2005.

[12] Goldstein T and Osher S. The split Bregman algorithm for L1 regularized problems. *UCLA CAM Report*, April 2008.

[13] Zhu M and Chan T F. An efficient primal–dual hybrid gradient algorithm for Total Variation image restoration. CAM Report 08-34, UCLA, 2008.

[14] Zhu M, Wright S J, and Chan T F. Duality-based algorithms for total-variation-regularized image restoration. *Computational Optimization and Applications*, 47:377–400, 2008.

[15] Beck A and Teboulle M. Fast gradient–based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Processing*, 18:2419–2434, 2009.

[16] Aujol J F. Some first-order algorithms for total variation based image restoration. *J. Math. Imaging Vis.*, 34(3):307–327, 2009.

[17] Dahl J, Hansen P C, Holdt S, Tobias J, and Jensen L. Algorithms and software for total variation image reconstruction via first-order methods. *Numerical Algorithms*, 53:67–92, 2010.

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

[19] Chan R H and Chen K. Multilevel algorithms for a Poisson noise removal with Total–Variation regularization. *Int. J. Comput. Math.*, 84:1183–1198, 2007.

[20] Le T, Chartrand R, and Asaki T J. A variational approach to reconstructing images corrupted by Poisson noise. *J. Math. Imaging Vis.*, 27:257–263, 2007.

[21] Dey N, Blanc-Féraud L, Zimmer C, Roux P, Kam Z, Olivo-Marin J C, and Zerubia J. Richardson–Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution. *Microscopy Research and Technique*, 69(4):260–266, 2006.

[22] Brune C, Sawatzky A, and Burger M. *Scale Space and Variationale Methods in Computer Vision*, volume 5567 of *LNCS*, chapter Bregman–TV–EM methods with application to optical nanoscopy, pages 235–246. Springer, 2009.

[23] Figueiredo M A T and Bioucas-Dias J M. Deconvolution of poissonian images using variable augmented lagrangian optimization. *IEEE Workshop on Statistical Signal Processing, Cardiff*, 2009.

[24] Brune C, Sawatzky A, and Burger M. Primal and dual Bregman methods with application to optical nanoscopy. *Int. J. Comput. Vis.*, in press:1–19, 2010.

[25] Mohler G O, Bertozzi A L, Goldstein T A, and Osher S J. Fast TV regularization for 2D Maximum penalized likelihood estimation. *J. Comp. Graph. Stat*, 2010.

[26] Setzer S, Steidl G, and Teuber T. Deblurring Poissonian images by split Bregman techniques. *J. Vis. Commun. Image R.*, 21:193–199, 2010.

[27] Setzer S. Operator Splitting, Bregman Methods and Frame Shrinkage in Image Processing. *Int. J. Comput. Vis.*, DOI: 10.1007/s11263-010-0357-3, 2010.

[28] Bardsley J M and Luttman A. Total Variation-Penalized Poisson Likelihood Estimation for Ill-Posed Problems. *Adv Comput Math*, 31(1–3):35–59, 2009.

[29] Zanella R, Boccacci P, Zanni L, and Bertero M. Efficient gradient projection methods for edge-preserving removal of Poisson noise. *Inverse Problems*, 25, 2009.

[30] Bertero M, Boccacci P, Talenti G, Zanella R, and Zanni L. A discrepancy principle for Poisson data. *Inverse Problems*, 26:10500, 2010.

[31] Bonettini S and Ruggiero V. Analysis of interior point methods for edge–preserving removal of Poisson noise. submitted, 2011.

[32] Brune C, Sawatzky A, Wübbeling F, Kösters T, and Burger M. An analytical view on EM–TV based methods for inverse problems with Poisson noise. 2009.

[33] Patriksson M. *Nonlinear programming and variational inequality problems: a unified approach.* Kluwer Academic Publisher, 1999.

[34] Korpelevich G. The extragradient method for finding saddle points and other problems. *Ekonomika i Matematicheskie Metody*, 12:747–756, 1976.

[35] Khobotov E N. Modification of the extragradient method for solving variational inequalities and certain optimization problems. *USSR Computat. Math. Phys.*, 27:120–127, 1987.

[36] Marcotte P. Application of Khobotov's algorithm to variational inequalities and network equilibrium problems. *INFOR.*, 29:258–270, 1991.

[37] Antipin A S. From optima to equilibria. In *Dynamics of non-homogeneous systems. Proceedings of ISA-RAS*, 2000.

[38] Antipin A S. Feedback-controlled saddle gradient processes. *Automation and Remote Control*, 55(3):311–320, 2003.

[39] Bonettini S and Ruggiero V. On the uniqueness of the solution of image reconstruction problems with Poisson data. In T.E. Simos et al., editor, *Proceedings of ICNAAM 2010*, volume 1281 of *AIP conference proceedings*, pages 1803–1806. AIP, 2010.

[40] Rockafellar R T. *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.

[41] Esser E, Zhang X, and Chan T. A general framework for a class of first order primal–dual algorithms for TV minimization. CAM Reports 09-67, UCLA Center for Applied Math., 2009.

[42] Chambolle A and Pock T. A first–order primal–dual algorithm for convex problems with applications to imaging. *R.I. 685*, May 2010.

[43] Zhang X, Burger M, and Osher S. A unified primal–dual algorithm framework based on bregman iteration. *J. Sci. Comput.*, 46:20–46, 2011.

[44] Arrow K J, Hurwicz L, and Uzawa H. *Studies in Linear and Non-Linear Programming.* Stanford University Press, Stanford, 1958.

[45] Bertsekas D and Tsitsiklis J. *Parallel and Distributed Computation: Numerical Methods.* Prentice-Hall, 1988.

[46] Willett R M and Nowak R D. Platelets: A multiscale approach for recovering edges and surfaces in photon limited medical imaging. *IEEE Transactions on Medical Imaging*, 22:332–350, 2003.