Inverse Problems **25** (2009) 045010 (24pp) doi:10.1088/0266-5611/25/4/045010

Efficient gradient projection methods for edge-preserving removal of Poisson noise

R Zanella[1], P Boccacci[2], L Zanni[1] and M Bertero[2]

[1] Dipartimento di Matematica Pura e Applicata, Università di Modena e Reggio Emilia, Via G. Campi 213/b, I-41100 Modena, Italy
[2] Dipartimento di Informatica e Scienze dell'Informazione, Università di Genova, Via Dodecaneso 35, I-16146, Genova, Italy

Received 11 November 2008, in final form 26 January 2009
Published 17 February 2009
Online at stacks.iop.org/IP/25/045010

Abstract
Several methods based on different image models have been proposed and developed for image denoising. Some of them, such as total variation (TV) and wavelet thresholding, are based on the assumption of additive Gaussian noise. Recently the TV approach has been extended to the case of Poisson noise, a model describing the effect of photon counting in applications such as emission tomography, microscopy and astronomy. For the removal of this kind of noise we consider an approach based on a constrained optimization problem, with an objective function describing TV and other edge-preserving regularizations of the Kullback–Leibler divergence. We introduce a new discrepancy principle for the choice of the regularization parameter, which is justified by the statistical properties of the Poisson noise. For solving the optimization problem we propose a particular form of a general scaled gradient projection (SGP) method, recently introduced for image deblurring. We derive the form of the scaling from a decomposition of the gradient of the regularization functional into a positive and a negative part. The beneficial effect of the scaling is proved by means of numerical simulations, showing that the performance of the proposed form of SGP is superior to that of the most efficient gradient projection methods. An extended numerical analysis of the dependence of the solution on the regularization parameter is also performed to test the effectiveness of the proposed discrepancy principle.

1. Introduction

Image denoising is an important topic in imaging science. We mention a few of the most important approaches that have been proposed: anisotropic filtering based on nonlinear diffusion equations [1, 28], total variation (TV) [27, 30], other edge-preserving regularizations

 © 2009 IOP Publishing Ltd Printed in the UK

[12, 21], wavelet thresholding [13, 17], etc. A recent review of these and other approaches is given in [7] (see also [10, chapter 4]).

Wavelet thresholding and TV regularization are essentially based on the assumption of additive white Gaussian noise. However, in applications such as emission tomography, microscopy and astronomy, the main source of noise corrupting the images is photon counting. In such a case, the statistics of the noisy images is described by a multi-valued Poisson process and the amount of noise is controlled by the total number of photons: the higher the number of photons, the lower the (relative) noise corrupting the image. Even if, for a sufficiently large number of photons (for instance, greater than 30), a Poisson process can be approximated by a suitable Gaussian process, the expected value and variance of the Gaussian depend on the pixel value of the unknown object.

In the case of astronomical images it has been proposed to use a variance stabilizing transform (VST), such as the Anscombe transform [2] or its extension to the case of low counting [32], followed by a thresholding based on a wavelet, ridgelet or curvelet representation of the detected images.

Another approach that has been recently proposed [9, 26] consists of extending TV regularization to the case of Poisson noise: the functional to be minimized is obtained by penalizing the Kullback–Leibler (KL) divergence of the (unknown) noise-free image from the noisy one, by means of the TV functional. The minimization is obtained by solving the related diffusion equation in [26] and by a multilevel method in [9].

In this paper we investigate, from the point of view of optimization theory, the discrete version of TV regularization and of similar problems related to other edge-preserving regularizations [12]. We propose for its solution a fast iterative algorithm derived from a scaled gradient projection (SGP) method recently introduced for image deblurring [6]. The efficiency of this method depends on the step-length updating rules and on the selection of the scaling matrix, which is related to the specific problem one considers. In this paper, the scaling used in designing the algorithm is obtained from the split-gradient method (SGM) described in [25]. The SGM is based on a decomposition (splitting) of the gradient of the objective function into a positive and a negative part. Thus, the use of this scaling does not require additional computations. We remark that the SGM is a scaled gradient method with a positive scaling and, from this point of view, is superior to the one-step late (OSL) algorithm proposed in [23] (see appendix B), which is also a scaled gradient method but with a scaling that may take negative values.

By investigating the statistical properties of Poisson noise, we also propose a modification of the so-called discrepancy principle (also known as the Morozov discrepancy principle [15]), which is used for the choice of the regularization parameter. Since the standard regularization theory is tacitly based on the assumption of additive Gaussian noise, which is object independent, the Morozov discrepancy principle in its more general form requires knowledge of the expected value and of the covariance matrix of the noise. Therefore, it cannot be applied to the case of Poisson noise, whose statistics is object dependent (however, for the approximation of the negative log of the Poisson likelihood with a weighted least-square problem, see [4]). The new discrepancy principle depends only on the detected image and is based on the KL divergence of the computed images from the measured one. It provides sensible results in our numerical experiments but is not based on a complete theory; therefore, it deserves further investigation.

The paper is organized as follows. In section 2, we formulate the optimization problem and we propose the new discrepancy principle. In section 3 we describe the SGP algorithm, including the rules for step length and scaling selection. In section 4 we report about our numerical experiments and, in section 5, we give conclusions.

2. Formulation of the problem

We consider a formulation of the denoising problem based on a Bayesian approach. Then the computation of the so-called *maximum a posteriori* (MAP) estimate can be reduced to the minimization of the negative log of the posterior probability density. In this section, we derive this functional and we describe its main features. Moreover, we introduce the arguments that justify the new discrepancy principle.

2.1. The likelihood

We denote by y the detected image, an array of integer numbers (the detected photons), with elements $y_{i,j}$, where i, j are indices ranging from 1 to n. Then the basic assumption is that $y_{i,j}$ is the realization of a Poisson random variable (r.v.) $Y_{i,j}$ with the expected value $\bar{x}_{i,j}$. We denote by Y the multi-valued r.v. formed by $Y_{i,j}$. Sometimes, in the following of this paper, we denote $\bar{x}$ as the 'true object' or 'noise-free object'.

In general (see the discussion at the beginning of section 3) the expected values $\bar{x}_{i,j}$ are strictly positive, while some of the detected values $y_{i,j}$ can be zero, if the corresponding $\bar{x}_{i,j}$ are not sufficiently large. On the other hand, $y_{i,j}$ are strictly positive in the case of the so-called 'background-dominated' images, which can be relevant in some astronomical applications. In conclusion, the denoising problem is just to find a positive estimate x^* of the unknown expected value $\bar{x}$, it being given the detected image y.

If we assume that the Poisson r.v.s associated with different pixels are statistically independent, then the likelihood function of the problem is given by

$$L_y^Y(x) = \prod_{i,j=1}^{n} \frac{\mathrm{e}^{-x_{i,j}} x_{i,j}^{y_{i,j}}}{y_{i,j}!}, \tag{1}$$

where y is the detected image, a realization of Y. In the maximum likelihood (ML) approach, the solution of the denoising problem is provided by any array x^* that maximizes the likelihood function.

By taking the negative log of the likelihood and adding a suitable constant, the maximization of the likelihood is reduced to the minimization of the KL divergence (also called Csiszár I divergence [14]) of x from y:

$$J_y^{(0)}(x) = \sum_{i,j=1}^{n} \left\{ y_{i,j} \ln \frac{y_{i,j}}{x_{i,j}} + x_{i,j} - y_{i,j} \right\}, \tag{2}$$

where $y_{i,j} \ln y_{i,j} = 0$ if $y_{i,j} = 0$. It is well known that this function is convex, strictly convex if and only if $y_{i,j} > 0$ for any i, j. In all cases, it has a unique minimum given by $x^* = y$. Therefore the solution of the ML problem exists, is unique, but *is trivial*. For this reason, it is necessary to 'regularize' the ML approach by a suitable prior describing the statistical properties of the unknown solution.

2.2. Edge-preserving priors

In a Bayesian approach, if we consider Gibbs priors, then the MAP estimate of the denoising problem is obtained by minimizing a functional of the following form:

$$J_y(x) = J_y^{(0)}(x) + \beta J^{(R)}(x), \tag{3}$$

where β is a *regularization parameter*. A method for estimating an 'optimal' value of this parameter is discussed in the following subsection. In the following, we denote by x_β^* a

minimum point of this functional. If $J^{(R)}(x)$ is convex and y is strictly positive, then functional (3) has a unique minimum point.

2.2.1. Total variation and hypersurface regularization. The standard 'regularization' of the denoising problem in the case of additive Gaussian noise is given in terms of the *total variation* functional [30] that, in a continuous model, is given by

$$F(x) = \int_{\Omega} |\nabla x|, \tag{4}$$

where Ω is the image domain. For an overview see, for instance, [10]. As mentioned in section 1, this approach has been already used for the denoising of images corrupted by Poisson noise [9, 26]. For a statistical interpretation of the Gibbs prior associated with this functional see, for instance [3, 8].

Since our approach is completely discrete, for discretizing equation (4) we assume, for simplicity, that one row and one column have been added to the $n \times n$ array x of the sampling values $x_{i,j}$ by taking $x_{n+1,j} = x_{1,j}$ and $x_{i,n+1} = x_{i,1}$. Then, with a notation used by Vogel [31], we write the discrete version of the TV functional in the following form:

$$J^{(R)}(x) = \frac{1}{2} \sum_{i,j=1}^{n} \psi_{\delta}\left(D_{i,j}^2\right), \tag{5}$$

where

$$D_{i,j}^2 = (x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2, \tag{6}$$

and $\psi_{\delta}(t)$ is given by

$$\psi_{\delta}(t) = 2\sqrt{t + \delta^2}, \qquad \psi'_{\delta}(t) = \frac{1}{\sqrt{t + \delta^2}}, \tag{7}$$

δ being a small quantity introduced for removing the singularity of the gradient in pixels where $D_{i,j} = 0$ (in our simulations, we take $\delta = 10^{-8}$). Another example of a function approximating the discrete TV functional is provided by the Huber potential [31].

However, the same function ψ_{δ} has been considered in a more general approach to edge-preserving regularization [12]. In such a case, it is called *hypersurface potential* [11] and δ is considered as a thresholding parameter which tunes the value of the gradient above which a discontinuity is detected. For instance in [12], the authors take $\delta = 7$ for a specific numerical simulation. We remark that in domains where $D_{i,j}$ is smaller (much smaller) than δ, the regularization provided by this functional approximates Tikhonov regularization in terms of the ℓ_2 norm of the modulus of the discrete gradient, computed over these domains. Indeed, global Tikhonov regularization is obtained by replacing the function ψ_{δ} in equation (5) with $\psi(t) = t$. For these reasons, we also consider the case where δ can be greater (much greater) than the default value 10^{-8}. Moreover, as we remark in section 4.5, our approach can also be easily applied to the case of Tikhonov regularization.

2.2.2. Markov random field (MRF) regularization. The use of Markov random fields in image restoration and processing was first introduced by Geman and Geman [20] and subsequently applied to edge-preserving image reconstruction [21]. In these papers the focus is on tomography but, of course, the same approach can be applied to image deconvolution, denoising, etc.

In the most simple case, the regularization functional can be written as a sum of potentials involving only the 'nearest neighbor' interactions between the components of x. Therefore, given a pixel i, j, we consider the set of the indices of all its first neighbors, denoted as $\mathcal{N}_{i,j}$.

In order to get symmetric sets also for the boundary pixels, we add another row and column to the x-array by setting $x_{0,j} = x_{n,j}$ and $x_{i,0} = x_{i,n}$.

One example of these regularization functionals is given by

$$J^{(R)}(x) = \frac{1}{4}\sum_{i,j=1}^{n}\sum_{k,l\in\mathcal{N}_{i,j}}\psi_\delta\left[\left(\frac{x_{i,j}-x_{k,l}}{w_{k,l}}\right)^2\right], \tag{8}$$

with ψ_δ being defined in equation (7), $w_{k,l} = 1$ for horizontal and vertical neighbors and $w_{k,l} = \sqrt{2}$ for diagonal neighbors, which are included to provide an isotropic edge-preserving effect, as discussed in [22].

A more general class of edge-preserving functionals can be obtained by replacing ψ_δ with a continuously differentiable, nondecreasing function $\psi(t)$, satisfying conditions discussed in [12]. We remark that the potential function $\phi(t)$, introduced in that paper, is related to our function $\psi(t)$ by $\phi(t) = \psi(t^2)$. The conditions ensuring the edge-preserving property are satisfied by the function of equation (7) and by the potential function proposed in [21]:

$$\psi_\delta(t) = \frac{\delta t}{t+\delta^2}, \qquad \psi'_\delta(t) = \frac{\delta^3}{(t+\delta^2)^2}, \tag{9}$$

where δ is again a thresholding parameter tuning the discontinuities. The corresponding functional, called Geman & McClure functional, is not convex; moreover, for large values of t the function tends to a constant value and therefore the functional does not introduce regularization when the jumps are much greater than δ. On the other hand, if the modulus of the gradient is much smaller than δ, the functional provides essentially a Tikhonov regularization as that corresponding to the function of equation (7).

In this paper we do not investigate the effect of the Geman & McClure functional, even if the algorithm we propose can also be easily applied to this case. For simplicity and comparison with TV regularization, we restrict our numerical experiments to the case of function (7). We call it the MRF regularization.

2.3. A discrepancy principle

In all regularization approaches, a difficult problem is the choice of a suitable value of the regularization parameter β. In the case of the TV regularization applied to the case of additive white Gaussian noise, this choice is performed by means of the *discrepancy principle* [15] which states that the reconstruction should have a mean-squared difference from the noisy data that is equal to the variance of the noise. This principle is not applicable to the case of Poisson noise since the variance is pixel and object dependent. For this reason, in [26] the authors state that, in their simulation, the value of β is obtained by means of a 'generalized discrepancy principle' that, in our notations, can be formulated as follows: if x^*_β is the minimizer of the functional $J_y(x)$, then the value of β is obtained by solving the equation

$$J_y^{(0)}(x^*_\beta) = J_y^{(0)}(\bar{x}), \tag{10}$$

where $\bar{x}$ is the true object.

This approach is not applicable to the case of real data because it requires knowledge of the true object. However, let us consider the following lemma, whose proof is elementary even if we did not find it in the refereed literature. A sketch of the proof is given in appendix A.

Lemma 1. *Let Y_λ be a Poisson r.v. with the expected value λ and consider the following function of Y_λ:*

$$F(Y_\lambda) = 2\left\{Y_\lambda \ln\left(\frac{Y_\lambda}{\lambda}\right) + \lambda - Y_\lambda\right\}. \tag{11}$$

Then, for large λ*, the following asymptotic estimate of the expected value of* $F(Y_\lambda)$ *holds true:*

$$E\{F(Y_\lambda)\} = 1 + O\left(\frac{1}{\lambda}\right). \tag{12}$$

As a consequence of this lemma, we introduce the following definition.

Definition 1. *For a given detected image* y*, the Poisson discrepancy of* x *from* y *is the following normalized KL divergence:*

$$D_y(x) = \frac{2}{n^2} J_y^{(0)}(x) = \frac{2}{n^2} \sum_{i,j=1}^{n} \left\{ y_{i,j} \ln \frac{y_{i,j}}{x_{i,j}} + x_{i,j} - y_{i,j} \right\}. \tag{13}$$

If we compute this function at $x = \bar{x}$, the true expected value, then the terms in the sum fluctuate around $1/2$, i.e. each term takes a value a bit greater or a bit smaller than $1/2$. Therefore, thanks to compensation between fluctuations over and below this value, the value of $D_y(\bar{x})$ is approximately 1. The approximation becomes better and better for increasing values of the object intensity. For instance, in the case of the three phantoms considered in section 4, we find $D_y(\bar{x}) = 1.0874$, 1.0318 and 1.0127 for increasing intensity of $\bar{x}$.

The previous observations suggest the following principle.

Discrepancy principle for Poisson data: for a given y, an 'optimal' value of β is a solution to the equation

$$D_y(x_\beta^*) = 1. \tag{14}$$

Work is in progress for proving the existence of at least one value of β solving this equation. The difficult point is to show that if the noise tends to zero, then β also tends to zero as well as the reconstruction error. Indeed, in the Poisson case, noise tending to zero means photon counts tending to infinity.

However we have strong numerical evidence of the utility of this principle, as shown in the section devoted to the numerical experiments. Since $D_y(y) = 0$ and x_β^* tends to y for β tending to zero, we have that the Poisson discrepancy is zero for $\beta = 0$; then, in all our numerical experiments, we find that, for increasing β, the function grows from zero and crosses 1, thus providing a unique value of β, indicating where to search for a value that is 'optimal' with respect to figures of merit suggested by the specific application one considers. We also find that when the noise decreases, i.e. the intensity of $\bar{x}$ increases, the value of β provided by this principle decreases and also the relative reconstruction error decreases, thus suggesting that if the noise tends to zero, then the solution provided by this principle converges, in some sense, to the 'true' solution.

As a final remark, we point out the main difference between additive Gaussian noise and Poisson noise, explaining why the previous principle is reasonable. Indeed, in the case of additive Gaussian noise, the expected value and variance of the noise are independent of the intensity of the object while, in the case of Poisson noise, the intensity of the object in a given pixel dictates both the expected value and the variance of the noise in that pixel. Therefore, a discrepancy principle in the Poisson case does not require additional information on the noise beyond that already contained in the detected image.

3. The scaled gradient projection method

The functional $J_y^{(0)}(x)$ is singular on the boundary of the non-negative orthant. However, images produced by emission processes are, in general, also corrupted by some background emission. In the case of image deblurring it is important to insert the expected value of

the background in the inversion algorithm [29], while this additional information is not so useful in denoising problems. The important consequence of background emission is that the unknown object (including background) is strictly positive so that we can consider the following constrained minimization problem:

$$\begin{aligned} &\min \quad J_y(x) = J_y^{(0)}(x) + \beta J^{(R)}(x) \\ &\text{sub. to} \qquad x \geqslant \eta, \end{aligned} \tag{15}$$

where η is a constant, which must be smaller than the background. In such a way, the functional $J_y^{(0)}(x)$ is continuously differentiable on the feasible domain and numerical solvers suited for differentiable optimization can be applied. The choice of η does not influence the quality of the reconstruction, so that one can take some small value such as $\eta = 10^{-5}$, as we do in our numerical experiments.

3.1. The algorithm

We propose to solve the minimization problem (15) by a special version of the scaled gradient projection method proposed in [6]. To describe this solver, we need the following notation. First, we denote by $P_\eta(\cdot)$ the projection operator onto the feasible region of (15):

$$P_\eta(x) \equiv \arg\min_{z \geqslant \eta} \|z - x\|_2, \tag{16}$$

where $\|\cdot\|_2$ denotes the usual ℓ_2 norm of arrays. Moreover, just because we consider arrays and not vectors, diagonal scalings of the gradient are obtained by multiplying, in the Hadamard sense (i.e. pixel by pixel), the gradient by a suitable array S with entries $s_{i,j}$. Then, in order to satisfy conditions required by the convergence theorem proved in [6], we assume that, for a given positive scalar $L > 1$, the scaling S belongs to the set $\mathcal{S}_L$ consisting of all the arrays such that

$$\frac{1}{L} \leqslant s_{i,j} \leqslant L, \qquad i, j = 1, 2, \ldots, n. \tag{17}$$

The minimization method that we consider is the scaled gradient projection-positivity constraints (SGP-POS) algorithm described below.

Algorithm SGP-POS (scaled gradient projection method-positivity constraints)

Choose the starting point $x^{(0)} \geqslant \eta$, set the parameters $\gamma, \theta \in (0, 1)$, $0 < \alpha_{\min} < \alpha_{\max}$ and fix a positive integer M. Set $k \leftarrow 0$.

REPEAT

STEP 1. Choose the parameter $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$ and the scaling array $S_k \in \mathcal{S}_L$.

STEP 2. Projection: $z^{(k)} = P_\eta(x^{(k)} - \alpha_k S_k \nabla J_y(x^{(k)}))$.

STEP 3. Descent direction: $d^{(k)} = z^{(k)} - x^{(k)}$.

STEP 4. Set $\lambda_k = 1$ and $J_{\max} = \max_{0 \leqslant j \leqslant \min(k, M-1)} J_y(x^{(k-j)})$.

STEP 5. Backtracking loop:

IF $J_y(x^{(k)} + \lambda_k d^{(k)}) \leqslant J_{\max} + \gamma \lambda_k \nabla J_y(x^{(k)})^T d^{(k)}$ THEN

go to Step 6;

ELSE

set $\lambda_k = \theta \lambda_k$ and go to Step 5;

ENDIF

STEP 6. Set $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$ and $k \leftarrow k + 1$.

UNTIL stopping criterion is satisfied.

Compared with the general SGP method [6], it differs in the projection step, which is appropriately rewritten by exploiting the special constraint (15) and the assumption on the scaling S_k. By proceeding as in [6], it is possible to show the following convergence result for SGP-POS: if the level set $\Omega_0 = \{x \geqslant \eta : J_y(x) \leqslant J_y(x^{(0)})\}$ is bounded, then every accumulation point of the sequence $\{x^{(k)}\}$ is a stationary point of (15), hence a minimum point when (15) is convex. The SGP scheme has been evaluated in [6] on image deconvolution problems, in the case of Poisson noise and without regularization: due to special choices of the scaling and of the step-length parameter, it has shown a very good convergence rate and remarkable performance improvements in comparison to other iterative regularization methods such as the expectation maximization (EM) algorithm and the modified residual norm steepest descent algorithm [4]. These promising results also motivate the application of the scheme to the denoising problem (15).

In order to effectively solve (15) by means of SGP-POS, we must provide suited strategies for the step-length selection and the scaling updating.

We refer to subsection 4.1 for details on the possible choices of the starting point $x^{(0)}$ and of the stopping condition.

3.2. Step-length selection

We suggest to update the step length α_k by adapting the Barzilai–Borwein (BB) rules [5] to the context of the scaling gradient methods, widely used in standard non-scaled gradient methods. When the scaled direction $S_k \nabla J_y(x^{(k)})$ is exploited within a step of the form $(x^{(k)} - \alpha_k S_k \nabla J_y(x^{(k)}))$, the BB step-length rules become

$$\alpha_k^{(1)} = \frac{\left\| S_k^{-1} r^{(k-1)} \right\|_2^2}{\left\langle r^{(k-1)}, S_k^{-1} t^{(k-1)} \right\rangle_2}, \qquad \alpha_k^{(2)} = \frac{\langle r^{(k-1)}, S_k t^{(k-1)} \rangle_2}{\| S_k t^{(k-1)} \|_2^2}, \tag{18}$$

where $r^{(k-1)} = x^{(k)} - x^{(k-1)}$, $t^{(k-1)} = \nabla J_y(x^{(k)}) - \nabla J_y(x^{(k-1)})$, and $\langle \cdot, \cdot \rangle_2$ denotes the scalar product of arrays. Moreover, S_k^{-1} is the array with entries $1/(S_k)_{i,j}$. It may be observed that these formulae reduce to the standard BB rules in the case of non-scaled gradient methods, which is when S_k is the identity array 1 (all entries equal to 1).

The recent literature on the step-length selection in gradient methods suggests to design step-length updating strategies able to benefit from some kind of alternation of the BB rules. The effective way to do this is, for example, given in [16, 19, 33]. In the SGP-POS algorithm, by assuming that an initial value α_0 is given, we update the step length α_k, $k = 1, 2, \ldots$, as proposed in [6], which is by the following adaptive alternation criterion:

IF $\alpha_k^{(2)} / \alpha_k^{(1)} \leqslant \tau_k$ THEN

$\alpha_k = \min \left\{ \alpha_j^{(2)}, j = \max\{1, k - M_\alpha\}, \ldots, k \right\}$; $\tau_{k+1} = \tau_k * 0.9$;

ELSE

$\alpha_k = \alpha_k^{(1)}$; $\tau_{k+1} = \tau_k * 1.1$;

ENDIF

where M_α is a prefixed non-negative integer and $\tau_1 \in (0, 1)$ ($M_\alpha = 2$ and $\tau_1 = 0.5$ in this study). We refer to [6] for more details on this adaptive selection and for comparison with other standard step-length rules.

3.3. Scaling selection

The scaling of the gradient is obtained from a scaled gradient method, known as a split-gradient method and proposed in [25]. The SGM as well as its relationship with the OSL algorithm [23] is briefly described in appendix B.

We first observe that the gradient of $J_y^{(0)}(x)$ is given by

$$\nabla J_y^{(0)}(x) = 1 - \frac{y}{x}, \tag{19}$$

where 1 is the identity array and the quotient of two arrays is also in the Hadamard sense, i.e. pixel by pixel.

Next, we consider a decomposition (splitting) of the gradient $\nabla J^{(R)}(x)$ into a positive and a negative part:

$$-\nabla J^{(R)}(x) = U^{(R)}(x) - V^{(R)}(x), \tag{20}$$

with $U^{(R)}$ and $V^{(R)}$ non-negative arrays. It is obvious that such a decomposition is arbitrary to a large extent; explicit expressions for the (U, V)-functions for different regularization functionals are given in [25].

If we observe that the gradient of the complete functional can be written in the following form:

$$-\nabla J_y(x) = \left(\frac{y}{x} + \beta U^{(R)}(x)\right) - (1 + \beta V^{(R)}(x)), \tag{21}$$

then the scaling array S is defined as the ratio of the current iterate x to the positive part of the gradient. A justification of this choice is given in appendix B. More precisely, for ensuring that $S \in \mathcal{S}_L$, the scaling is the array given by

$$s_{i,j} = \min\left[L, \max\left\{\frac{1}{L}, \frac{x_{i,j}}{1 + \beta V_{i,j}^{(R)}(x)}\right\}\right]. \tag{22}$$

3.3.1. TV and HS regularizations. In the functional of equation (5), for a given $x_{i,j}$, only the terms corresponding to $D_{i,j}$, $D_{i,j-1}$ and $D_{i-1,j}$ depend on it; therefore, the (i, j)-component of the gradient is given by

$$\begin{aligned}\nabla_{i,j} J^{(R)}(x) = {} & \psi_\delta'\big(D_{i,j}^2\big)(2x_{i,j} - x_{i+1,j} - x_{i,j+1}) \\ & + \psi_\delta'\big(D_{i,j-1}^2\big)(x_{i,j} - x_{i,j-1}) + \psi_\delta'\big(D_{i-1,j}^2\big)(x_{i,j} - x_{i-1,j}),\end{aligned} \tag{23}$$

and, on the non-negative orthant, we have the following (U, V)-splitting:

$$U_{i,j}^{(R)} = \psi_\delta'\big(D_{i,j}^2\big)(x_{i+1,j} + x_{i,j+1}) + \psi_\delta'\big(D_{i,j-1}^2\big)x_{i,j-1} + \psi_\delta'\big(D_{i-1,j}^2\big)x_{i-1,j} \tag{24}$$

$$V_{i,j}^{(R)} = \big\{2\psi_\delta'\big(D_{i,j}^2\big) + \psi_\delta'\big(D_{i,j-1}^2\big) + \psi_\delta'\big(D_{i-1,j}^2\big)\big\}x_{i,j}. \tag{25}$$

3.3.2. MRF regularization. In the computation of the (i, j)-component of the gradient of the functional of equation (8), by taking into account the symmetry of $\mathcal{N}_{i,j}$, it is easy to see that the contribution of the terms depending on a given $x_{i,j}$ is given by

$$\frac{1}{2}\sum_{k,l\in\mathcal{N}_{i,j}} \psi_\delta\left[\left(\frac{x_{i,j} - x_{k,l}}{w_{k,l}}\right)^2\right]. \tag{26}$$

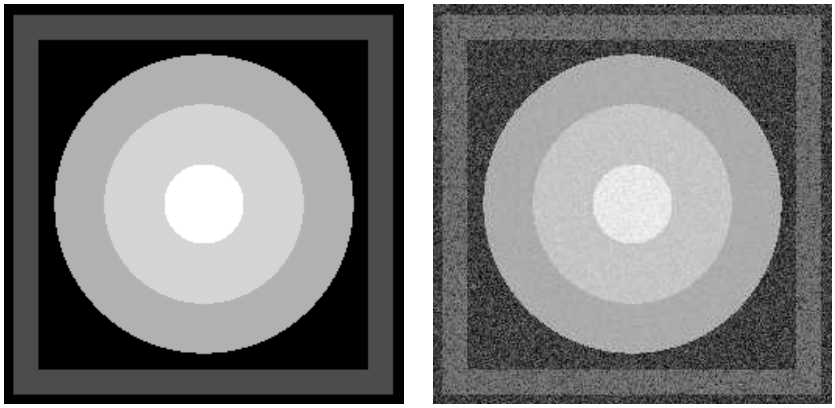

Figure 1. Left panel: the LCR phantom. Right panel: the corresponding noisy version. Image display intensity has been adjusted to make the frame visible.

Therefore, we obtain

$$\nabla_{i,j} J^{(R)}(x) = \sum_{k,l \in \mathcal{N}_{i,j}} \psi'_\delta \left[\left(\frac{x_{i,j} - x_{k,l}}{w_{k,l}} \right)^2 \right] \frac{x_{i,j} - x_{k,l}}{w_{k,l}}, \tag{27}$$

and, on the non-negative orthant, we have the following (U, V)-splitting:

$$U^{(R)}_{i,j} = \sum_{k,l \in \mathcal{N}_{i,j}} \psi'_\delta \left[\left(\frac{x_{i,j} - x_{k,l}}{w_{k,l}} \right)^2 \right] \frac{x_{k,l}}{w_{k,l}}, \tag{28}$$

$$V^{(R)}_{i,j} = \sum_{k,l \in \mathcal{N}_{i,j}} \psi'_\delta \left[\left(\frac{x_{i,j} - x_{k,l}}{w_{k,l}} \right)^2 \right] \frac{x_{i,j}}{w_{k,l}}. \tag{29}$$

4. Numerical results

For testing the proposed algorithm we use the phantom described in [26], which we call the LCR phantom. It is shown in figure 1 together with a corresponding noisy version. It is a 256×256 array, consisting of circles of intensities 70, 135 and 200, enclosed by a square frame of intensity 10, all on a background of intensity 5. There is no additional parameter for the noisy image since Poisson noise depends only on object intensity.

Obviously, in different experiments, one can measure different realizations of the same noise. In numerical simulations, this effect can be obtained by changing the seed of the random sequence.

4.1. Efficiency of the algorithm: TV and HS regularizations

In this subsection, we compare the efficiency of the proposed SGP with that of two more traditional algorithms. The first is a gradient projection (GP) method that is obtained from SGP by replacing the scaling array with the identity array 1. This allows us to demonstrate the utility of introducing a suitable scaling within a gradient projection scheme. The second gradient projection algorithm is obtained from GP by substituting the alternating step-length

selection with the first BB rule (i.e. the step length $\alpha_k^{(1)}$ in (18) with $S_k = 1$). This algorithm will be denoted by GP-BB; it is very similar to the GPSR-BB method proposed in [18] for image reconstruction problems with sparse priors but GP-BB exploits Armijo-like line searches instead of the limited minimization rule used by GPSR-BB. For the three algorithms, we consider both the monotone and the nonmonotone versions. The monotone scheme uses $M = 1$ in order to ensure that $J_y(x^{(k+1)})$ is lower than $J_y(x^k)$ at each iteration; the nonmonotone version is obtained by setting $M = 10$, which forces $J_y(x^{(k+1)})$ to be lower than the maximum of the values of the objective function on the last ten iterations. The other parameters of the methods are set as follows: $L = 10^{10}$, $\alpha_{\min} = 10^{-30}$, $\alpha_{\max} = 10^{30}$, $\alpha_0 = 1.3$, $\theta = 0.4$, $\gamma = 10^{-4}$.

We apply these algorithms to two test problems consisting in the denoising of the LCR phantom with two different regularizations:

- TV regularization with $\delta = 10^{-8}$ in (7),
- HS regularization with $\delta = 10^{-1}$ in (7).

In both cases we take $\beta = 0.25$, i.e. the value given in [26] and also used in [9]. The discussion on the choice of the value of this parameter is deferred to the following subsection.

All the considered algorithms are initialized with $x^{(0)} = P_\eta(y)$ and since we are interested in their behavior when they are pushed to convergence, we use the following convergence criterion: we stop the monotone SGP when

$$|J_y(x^{(k+1)}) - J_y(x^{(k)})| \leqslant \text{tol} J_y(x^{(k+1)}), \qquad \text{tol} = 10^{-7}; \tag{30}$$

then we compute the corresponding value of the objective function J_y and stop the other methods when they reach the same value of the objective function (allowing a maximum number of 2000 iterations). In all cases, we denote as x_β^* the result of these computations.

For measuring the quality of a given x_β^*, we consider the following figures of merit:

- relative error in the ℓ_1 norm

$$\rho_1^* = \frac{\|x_\beta^* - \bar{x}\|_1}{\|\bar{x}\|_1}, \tag{31}$$

- relative error in the ℓ_2 norm (relative r.m.s. error)

$$\rho_2^* = \frac{\|x_\beta^* - \bar{x}\|_2}{\|\bar{x}\|_2}. \tag{32}$$

The first is proportional to the mean absolute error used in [9] while the square of the second is proportional to the mean-squared difference between the reconstructed and the original images used in [26].

The results of our numerical experiments are given in table 1 where the values of δ corresponding to different regularizations are shown in the first column, the tested algorithms are indicated in the second column and the performance measures in terms of the number of iterations, ℓ_2 and ℓ_1 relative errors and computational time in seconds are reported in the last four columns. The asterisk after the iteration count means that the stopping criteria have not been satisfied within the maximum number of iterations. For simplicity, in this table as well as in the following ones SGP-POS is replaced by SGP.

Looking at table 1, we may observe that the SGP versions largely outperform the other algorithms, confirming the importance of both the scaling strategy and the alternating step-length selection. According to these results and to the computational study reported in [6], the two SGP versions (monotone and nonmonotone) show very similar performance, which depends on both the total number of iterations and the total number of function evaluations involved by the two different line-search strategies. However, in general, the monotone SGP

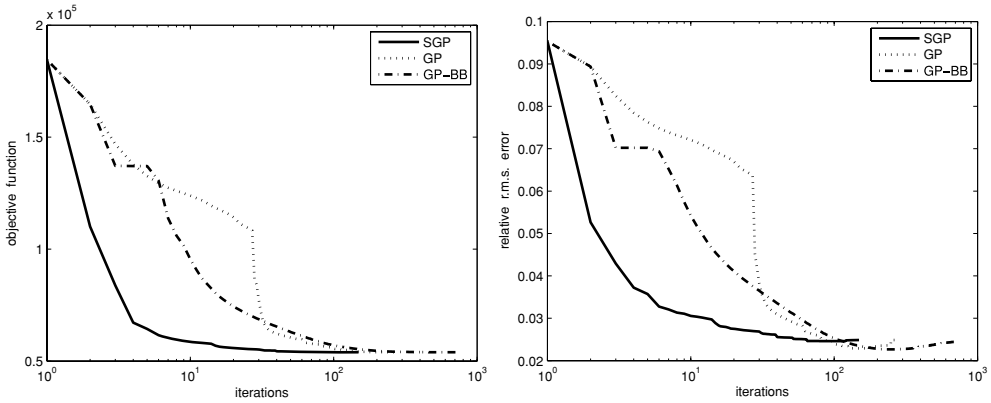

Figure 2. The objective function (left) and the ℓ_2 relative error (right) as a function of the number of iterations for the monotone algorithms in the case $\delta = 10^{-1}$.

Table 1. Performance of the algorithms on two denoising problems.

δ	Algorithm	Iterations	ℓ_2 rel. err.	ℓ_1 rel. err.	Time[a]
	SGP monotone	148	0.024 87	0.016 81	14.30
	SGP nonmonotone	167	0.024 91	0.016 79	13.53
10^{-1}	GP monotone	280	0.024 90	0.016 87	23.23
(HS)	GP nonmonotone	305	0.024 97	0.016 96	26.65
	GP-BB monotone	735	0.024 59	0.016 66	70.62
	GP-BB nonmonotone	600	0.024 66	0.016 72	51.63
	SGP monotone	369	0.027 28	0.020 96	35.14
	SGP nonmonotone	318	0.026 44	0.019 38	29.26
10^{-8}	GP monotone	2000*	0.031 57	0.025 89	159.13
(TV)	GP nonmonotone	2000*	0.031 51	0.025 80	157.52
	GP-BB monotone	2000*	0.031 88	0.026 13	156.35
	GP-BB nonmonotone	2000*	0.036 93	0.026 25	156.90

[a] Test environment: Matlab 7.5.0 on an AMD Opteron Dual Core 2.4 GHz processor.

exhibits a smoother behavior with respect to changes of the parameters β and δ. For this reason, the simulations in subsections 4.2 and 4.3 are performed by using the monotone SGP.

In figure 2, we plot the objective function and the relative r.m.s. error as a function of the iteration count for the three monotone methods in the case $\delta = 10^{-1}$. We observe a minimum in the error plot, implying that the value $\beta = 0.25$ is not the best value from the point of view of the ℓ_2 error. We also observe that, in the case of SGP, the reconstruction error does not change significantly for a large number of iterations. This suggests that it may be possible to reduce the tolerance in the stopping criterion without degrading the error. The behavior of the monotone algorithms for different values of the tolerance is shown in table 2 for the test problem corresponding to $\delta = 10^{-1}$. The results in this table confirm that a suitable setting of the tolerance in the stopping rule can imply significant time savings with satisfactory relative errors.

Finally we observe that if we multiply the relative ℓ_2 error by the ℓ_2 norm of the object, take the square of the result and divide the result by the number of pixels, then we obtain the

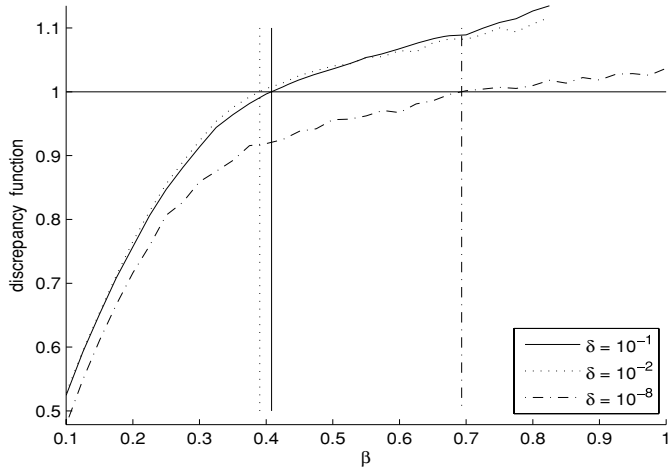

Figure 3. Plot of the discrepancy function (13) for HS regularization ($\delta = 10^{-1}$ and 10^{-2}) and TV regularization ($\delta = 10^{-8}$).

Table 2. Performance of the monotone algorithms for different tolerances in the stopping criterion. HS regularization with $\delta = 10^{-1}$.

	tol $= 10^{-3}$			tol $= 10^{-5}$			tol $= 10^{-7}$		
Algorithm	It.	ℓ_2 rel. err.	Time	It.	ℓ_2 rel. err.	Time	It.	ℓ_2 rel. err.	Time
SGP	18	0.027 57	1.73	72	0.024 63	6.35	148	0.024 87	14.30
GP	96	0.024 46	7.27	199	0.023 45	17.32	280	0.024 90	23.23
GP-BB	124	0.024 16	9.68	383	0.023 18	32.70	735	0.024 59	70.62

mean-squared difference used in [26] for measuring the accuracy of the reconstruction. From the values reported in table 1 for SGP in the case of TV regularization, we obtain the value 3.45 for this quantity, which is a bit lower than the value 4.40 reported in [26].

4.2. Choice of the regularization parameter

The monotone SGP algorithm is used for verifying the discrepancy principle introduced in section 2.3. For different values of β, we push the algorithm to convergence and we compute $D_y(x^*_\beta)$. We plot the graph of this function in figure 3 for three values of the parameter δ in equation (7): $\delta = 10^{-8}$ (TV regularization), $\delta = 10^{-2}$ and $\delta = 10^{-1}$ (HS regularization with two different thresholding values). The values of β corresponding to the crossing points are 0.693, 0.390 and 0.408, respectively. For the three pairs (δ, β), the relative ℓ_2 errors are 0.0345, 0.0266 and 0.0305 (the corresponding relative ℓ_1 errors are a bit smaller: 0.0216, 0.0139 and 0.0196). We remark that for $\delta = 10^{-1}$ and $\delta = 10^{-8}$, these values are a bit larger than those reported in table 1, even if the differences are not very significant. Therefore, we can conclude that the values of β obtained by means of the discrepancy principle provide reasonable results even if they are not optimal from the point of view of the ℓ_2 or ℓ_1 relative error. We also remark that $\delta = 10^{-2}$ provides smaller errors than the other two choices of δ.

In all cases the crossing occurs at a value greater than 0.25, the value quoted in [26]. The previous results imply that this value of β corresponds to a smaller value of the r.m.s. error in

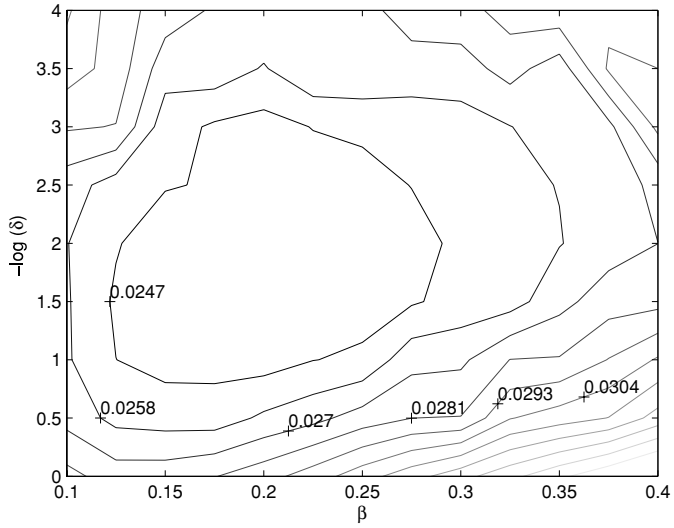

Figure 4. The relative ℓ_2 error as a function of β and δ. The behavior is represented in terms of level curves.

the case of TV regularization and presumably for this reason it has been chosen in [26]. We conclude that, similar to the Morozov discrepancy principle for Tikhonov regularization, the proposed discrepancy principle for Poisson data seems to overestimate the suited value of the regularization parameter, still providing sensible reconstructions.

Since ℓ_1 and ℓ_2 errors depend on β and δ, we investigate this dependence in the case of the LCR phantom. For simplicity, we consider only the ℓ_2 error and we compute this quantity on a grid of the (β, δ)-plane.

We plot the result in figure 4 in terms of level curves. We find a minimum around $\beta = 0.2, \delta = 10^{-2}$, corresponding to HS regularization with a moderate thresholding. The corresponding relative ℓ_2 error is 0.0235 (for $\delta = 10^{-2}$, the discrepancy principle provides $\beta = 0.39$ with a relative ℓ_2 error of 0.027; not so bad!).

Since the previous choice of (β, δ) is the best one from the point of view of the ℓ_2 error, we compare the result provided by this pair with that provided by TV regularization with $\beta = 0.25$, the value corresponding to the minimum relative ℓ_2 error (0.0272) for this kind of regularization (this is just the value of β given in [26]). To this purpose, we consider the normalized residual defined by

$$R^*_{i,j} = \frac{(x^*_\beta)_{i,j} - y_{i,j}}{\sqrt{(x^*_\beta)_{i,j}}}. \tag{33}$$

If $x^*_\beta = \bar{x}$, then this residual is a realization of white Gaussian noise with the expected value 0 and variance 1. We show the results provided by the two regularizations in the upper panels of figure 5. The two residuals are approximately equivalent, showing small artifacts in correspondence with the edges. The expected value and variance are -0.0389 and 0.7601 for HS, and -0.0425 and 0.7984 for TV, respectively.

Moreover, we compute the reconstructions provided by the two methods using 25 noisy images corresponding to 25 different realizations of noise. For each reconstruction, we consider the line-out from the middle of the image and we superimpose the line-outs corresponding to the 25 reconstructions for the two different methods. The results are shown

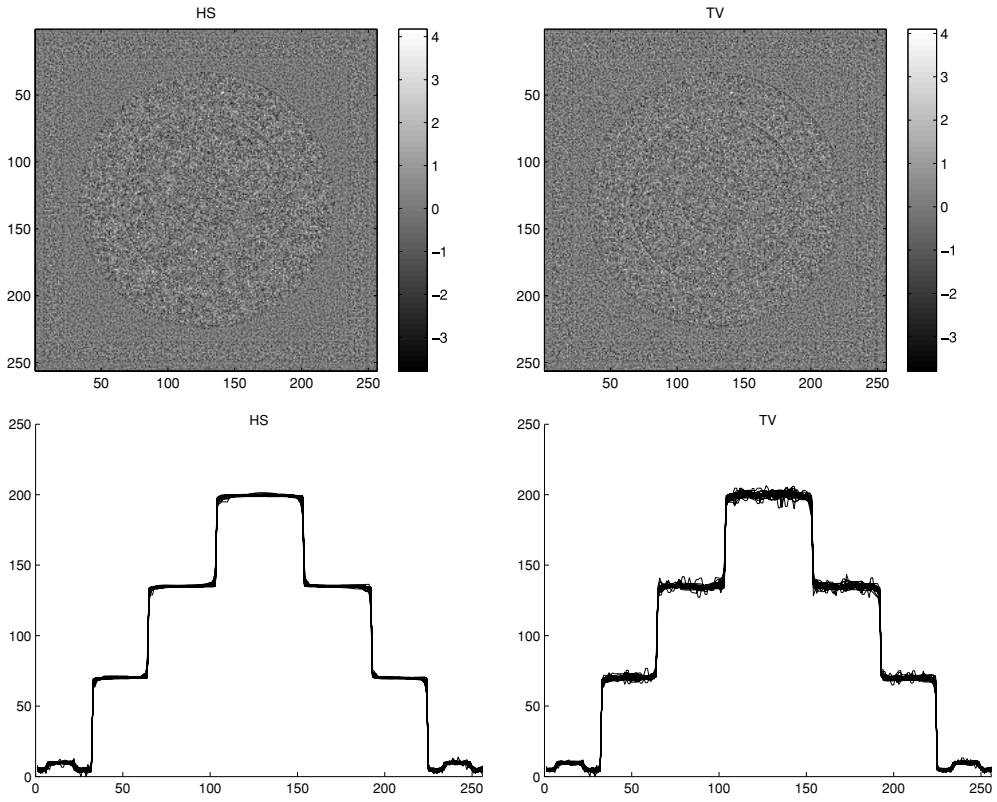

Figure 5. Upper panels: the normalized residuals defined in (33) for the reconstruction provided by HS regularization (left) with $\delta = 10^{-2}$, $\beta = 0.2$ and TV regularization (right) with $\delta = 10^{-8}$, $\beta = 0.25$. Lower panels: superposition of the line-outs from row number 128 for the reconstructions corresponding to 25 different realizations of noise, provided by HS regularization (left) and TV regularization (right). The values of the parameters are the same used for obtaining the upper panels.

in the lower panels of figure 5. HS regularization seems to provide better noise removal, at least in this specific example.

4.3. Noise dependence

As is well known, in the case of photon counting the noise level depends on the total number of photons: the higher is the number of photons, the lower is the relative noise level. More precisely, if we have a Poisson process with the expected value λ, then the variance is also λ and the standard deviation is $\sqrt{\lambda}$, so that the relative standard deviation is $1/\sqrt{\lambda}$.

In order to simulate this situation, we have generated two test phantoms: the first obtained by multiplying the LCR phantom by a factor of 10 (LCR_10; lower noise) and the second by multiplying by a factor of 0.2 (LCR_0.2; higher noise). In the first case, the maximum value is 2000 while in the second one it is 40. Then we repeat the analysis of the previous section. For the two problems, in figure 6 we show the discrepancy functions $D_y(x_\beta^*)$ corresponding to the three values of δ previously considered. By comparing the plots in figures 3 and 6 we observe that, for each δ, the values of β suggested by the discrepancy principle increase for growing levels of noise, a behavior similar to that of the regularization parameter in the case of Tikhonov regularization.

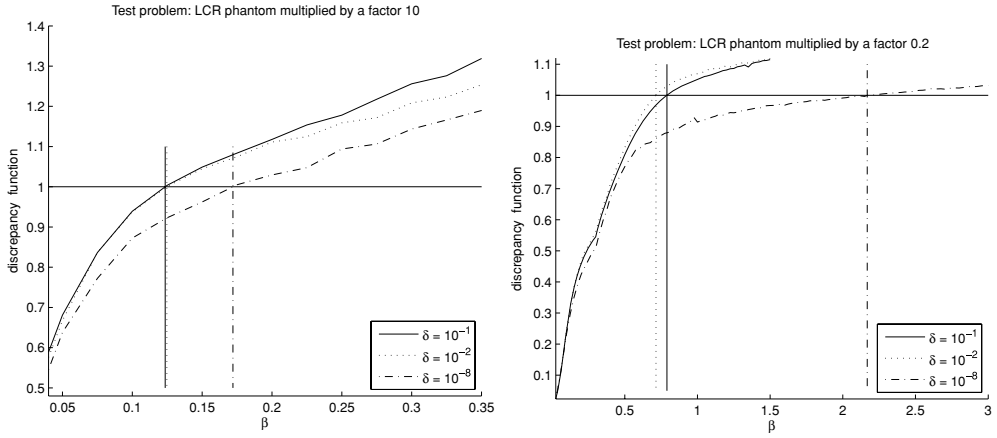

Figure 6. Plots of the discrepancy function (13) for HS regularization ($\delta = 10^{-1}$ and 10^{-2}) and TV regularization ($\delta = 10^{-8}$).

For space saving, in the case of these test problems we do not report the level curves of the ℓ_2 error as a function of the parameters β and δ. We simply remark that in the case of the test problem LCR_10, the minimum ℓ_2 error (about 0.008) is obtained for $\beta = 0.05$ and $\delta = 0.03$; furthermore, for $\beta = 0.05$, similar errors are obtained for a wide range of δ, since this small β makes the error slightly dependent on δ. In the case of the test problem LCR_0.2, the minimum error (about 0.042) is obtained for $\beta = 0.675$ and $\delta = 0.003$.

We complete this analysis by showing the superpositions of the line-outs from row 128 corresponding to 25 reconstructions with different realizations of noise. In figure 7, we report, in the upper panels, the superpositions obtained for the test problem LCR_10 by assuming $\delta = 10^{-2}$, $\beta = 0.05$ in the case of HS regularization (left) and $\delta = 10^{-8}$, $\beta = 0.05$ in the case of TV regularization (right); in the lower panels, we show the superpositions obtained for the test problem LCR_0.2 by assuming $\delta = 10^{-2}$, $\beta = 0.625$ in the case of HS regularization and $\delta = 10^{-8}$, $\beta = 0.575$ in the case of TV regularization.

4.4. Efficiency of the algorithm: MRF regularization

We again consider the original LCR phantom and we apply to the denoising problem the MRF regularization as defined by equation (8), with the function ψ defined in equation (7).

The first point is to estimate a value of β for a given δ. To this purpose, we again use the Poisson discrepancy principle. In figure 8, we plot the discrepancy functions $D_y(x^*_\beta)$ corresponding to different values of δ.

For $\delta = 10^{-1}$, $\delta = 10^{-2}$ and $\delta = 10^{-8}$, the values of β suggested by the principle are 0.175, 0.174 and 0.436, respectively. Starting from these suggestions, we compute the ℓ_2 error on a grid in the (β, δ)-plane and we obtain the minimum error (about 0.017, a bit smaller than that provided by HS regularization) for $\beta = 0.125$ and $\delta = 0.03$.

Anyway, for $\beta = 0.1$ and the values of δ considered in subsection 4.1 ($\delta = 10^{-1}$ and $\delta = 10^{-8}$), we apply the gradient projection methods by using the stopping criterion introduced in subsection 4.1. The results are described in table 3. Also in these experiments, the SGP methods provide the best performance. In comparison with the results reported in table 1 concerning HS and TV regularizations, it is interesting to remark the increased cost per iteration exhibited by the algorithms in the case of MRF regularization: this is due to the

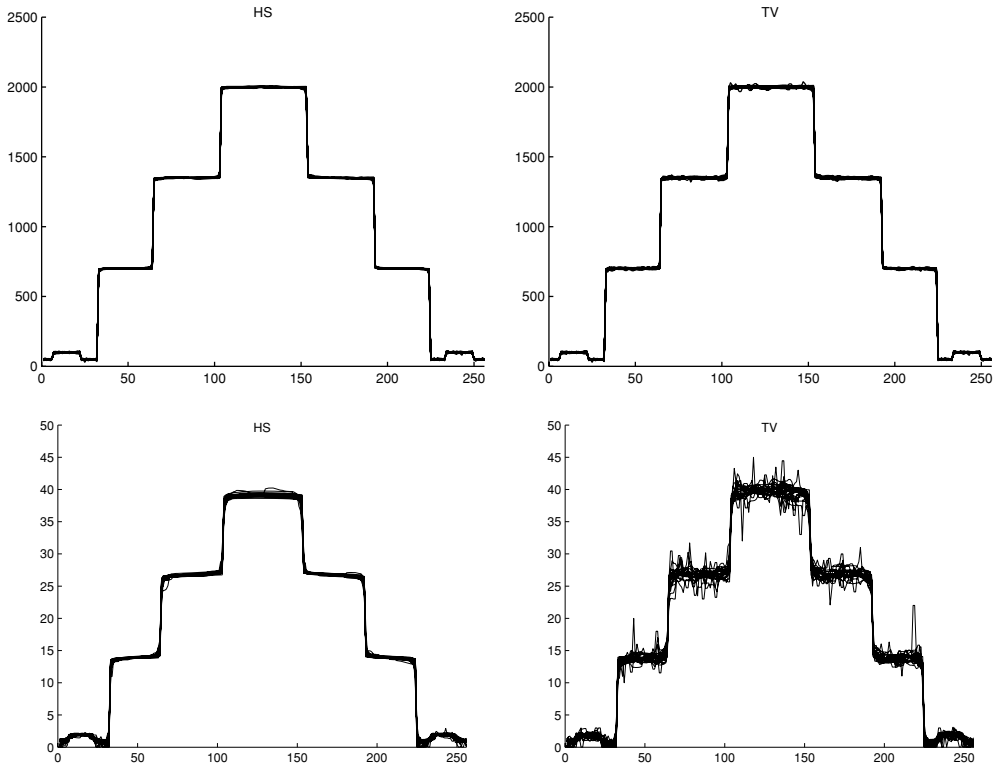

Figure 7. Superposition of the line-outs from row number 128 for the reconstructions corresponding to 25 different realizations of noise. Test problems: LCR_10 (upper panels) and LCR_0.2 (lower panels).

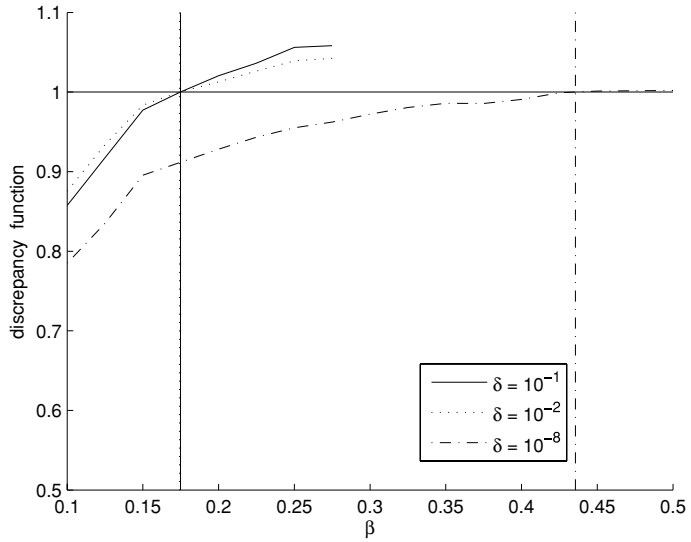

Figure 8. Plot of the discrepancy function (13) in the case of MRF regularization.

Table 3. Performance of the algorithms on the denoising problems with MRF regularization.

δ	Algorithm	Iterations	ℓ_2 rel. err.	ℓ_1 rel. err.	Time
10^{-1}	SGP monotone	184	0.019 33	0.016 62	33.85
	SGP nonmonotone	194	0.020 14	0.017 23	39.80
	GP monotone	289	0.019 25	0.016 57	56.67
	GP nonmonotone	227	0.019 16	0.016 49	44.73
	GP-BB monotone	491	0.019 24	0.016 54	116.26
	GP-BB nonmonotone	491	0.019 86	0.017 00	104.13
10^{-8}	SGP monotone	415	0.026 05	0.022 66	129.78
	SGP nonmonotone	482	0.021 00	0.018 43	185.86
	GP monotone	2000*	0.030 21	0.024 06	380.55
	GP nonmonotone	2000*	0.023 85	0.019 36	338.34
	GP-BB monotone	2000*	0.030 51	0.024 33	380.51
	GP-BB nonmonotone	2000*	0.033 92	0.024 57	379.18

more expensive computations involved by the regularization functional (8) in comparison with the regularization functional (5). On the other hand, it is also interesting to remark that MRF regularization provides a bit smaller reconstruction error.

4.5. Example of applications

In this section, we investigate potential applications of the denoising method based on HS regularization.

The first example is taken from medical imaging. We consider a good quality dental radiography, sized 512×512 and shown in the upper-left panel of figure 9. By reducing the total intensity and perturbing the result with Poisson noise, we simulate a noisy version corresponding to a lower dose. The result is shown in the upper-right panel of the same figure. The relative difference in the ℓ_2 norm between the noisy and the noise-free images is 0.179, i.e. 17.9%.

For estimating the values of the parameters β, δ, we first compute the map of the relative ℓ_2 error in the domain $\delta \in [10^{-3}, 10], \beta \in [0.1, 1.2]$. We find a minimum point at $\delta = 0.1, \beta = 0.25$, with a minimum value of about 2.79%, hence an improvement with respect to the noisy image of about a factor of 6. A map of the relative ℓ_1 error provides the same minimum point. The reconstructed image is shown in the lower-left panel of figure 9. In this reconstruction, the parts without sharp edges are more homogeneous than in the original one. Therefore, after a visual inspection of several reconstructed images, we believe that a better reconstruction is obtained with $\delta = 1, \beta = 0.3$. The result is shown in the lower-right panel of the same figure. The corresponding relative ℓ_2 error is 2.97%, not significantly greater than the minimum value. In other words, the minimum of the ℓ_2 error is considerably flat.

To check the proposed discrepancy principle, we compute the discrepancy function for $\delta = 0.1$ and $\delta = 1$. In the first case, the crossing occurs at $\beta = 0.556$ and in the second case at $\beta = 1.021$; these values are greater than the best values but still reasonable.

Finally, we check if it is possible to get a better reconstruction with a different regularization, namely with the ℓ_2 norm of the modulus of the gradient (Tikhonov regularization). In such a case, the functional is given by equation (5) with ψ_δ replaced by $\psi(t) = t$. The corresponding V-array is given by $V^{(R)}(x) = 4x$. The discrepancy principle provides $\beta = 0.69$, while the minimum of the reconstruction error occurs at $\beta = 0.2$

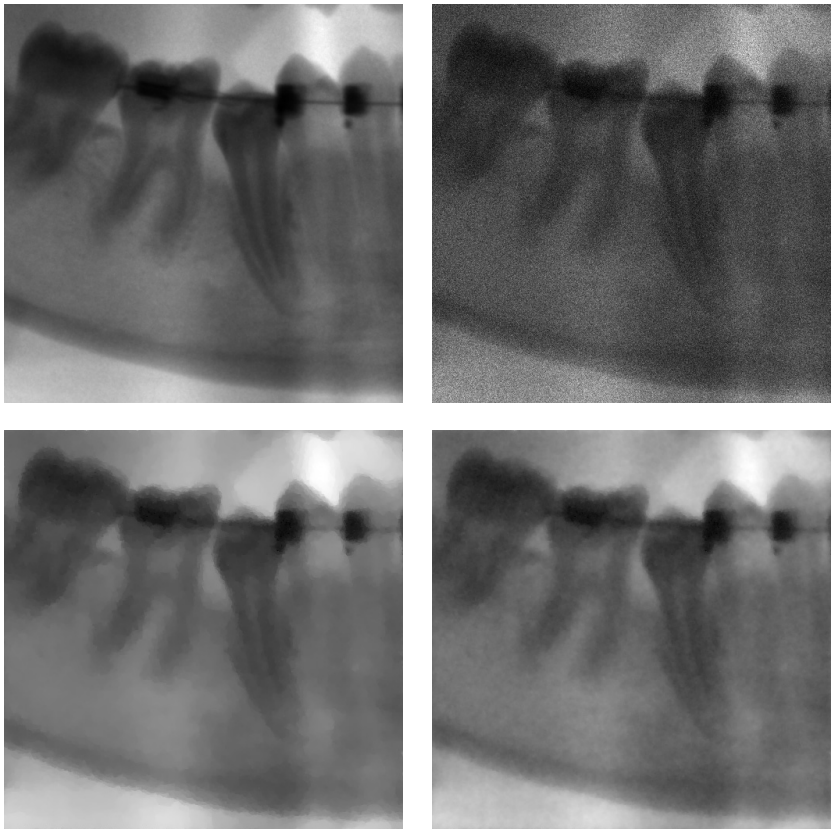

Figure 9. Upper-left panel: the dental radiography; upper-right panel: the lower dose version. Lower-left panel: the reconstruction corresponding to the minimum of the relative ℓ_2 error; lower-right panel: the reconstruction corresponding to $\delta = 1$, $\beta = 0.3$.

and is about 3.69%. The corresponding reconstruction does not seem better than that provided by the edge-preserving approach.

The second example is the astronomical image of the nebula NGC5979, sized 256×256 and shown in the upper-left panel of figure 10. In such a case, we simulate the result that should be obtained with a shorter integration time. It is shown in the upper-right panel of the same figure. The relative difference in the ℓ_2 norm is 1.003, i.e. about 100%.

We find a minimum of the reconstruction error at $\delta = 0.1$, $\beta = 0.02$ with a minimum value of 10.4%, hence an improvement of about a factor of 10 with respect to the noisy image. However, the reconstruction, shown in the lower-left panel of figure 10, is not satisfactory because the regions corresponding to small values of the gradient are exceedingly uniform. A better reconstruction is obtained with $\delta = 10$, $\beta = 0.02$, and is shown in the lower-right panel of the same figure. The large value of δ implies that we are looking for a compromise between edge-preserving and Tikhonov regularizations.

Also in this case the discrepancy principle provides reasonable values since for $\delta = 0.1$, the crossing occurs for $\beta = 0.039$ while for $\delta = 10$, it occurs for $\beta = 0.071$. However, it is obvious that the main problem is the choice of the correct value of δ. We conclude by observing that Tikhonov regularization does not provide a better result, so that the compromise provided

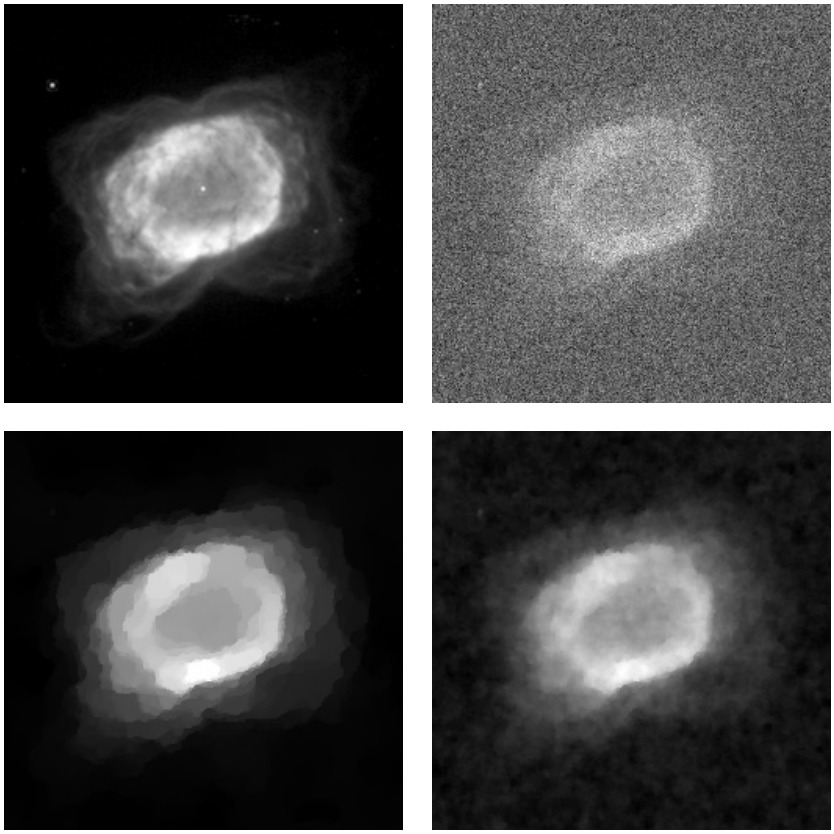

Figure 10. Upper-left panel: the image of the nebula NGC5979; upper-right panel: the image of the same nebula obtained with a lower integration time. Lower-left panel: the reconstruction corresponding to the minimum of the relative ℓ_2 error; lower-right panel: the reconstruction corresponding to $\delta = 10$, $\beta = 0.02$.

by HS regularization with a sufficiently high value of δ is presumably the most convenient one.

5. Concluding remarks

In this paper, we propose an efficient algorithm for edge-preserving removal of Poisson noise, formulated as an optimization problem in the framework of a Bayesian approach. The algorithm is a scaled gradient projection method obtained from the scheme recently proposed in [6]. It combines the scaled gradient directions used by the split-gradient method [25], with effective step-length updating rules introduced in [6, 19]. The computational study reported in this paper shows that the proposed algorithm is much more efficient than other widely used gradient projection methods.

In addition, we propose a new discrepancy principle for the choice of the regularization parameter and we show, by means of numerical simulations, that, in general, it provides a sensible overestimate of the value corresponding to the minimum of the reconstruction error.

Finally, our numerical simulations emphasize the relevance of the thresholding parameter introduced in the regularization functional for obtaining satisfactory reconstructions; at the present stage of this research, it is not clear how this parameter could be selected in the case of real data. A possible explication of its relevance in the case of Poisson data can be provided by the following remark. A Gaussian noise generates fluctuations in the data that have a uniform amplitude over the image domain; on the other hand, the amplitudes of the fluctuations due to Poisson noise are higher in the subdomains where the number of counts is higher. Therefore, the thresholding parameter could be related to these fluctuations and help to recognize that the jumps they produce are not due to some kind of edges but to noise.

Acknowledgments

This work was partially supported by MUR (Italian Ministry for University and Research), grant 2006018748, and by INAF (Italian Institute for Astronomy and Astrophysics).

Appendix A

Proof of Lemma 1. For any non-negative integer n, let us write

$$F(n) = 2\left\{n \ln\left(1 + \frac{n-\lambda}{\lambda}\right) + \lambda - n\right\}. \tag{A.1}$$

From the fourth-order Taylor formula for $\ln(1+\xi)$, we get

$$\ln(1+\xi) = \xi - \tfrac{1}{2}\xi^2 + \tfrac{1}{3}\xi^3 - \tfrac{1}{4}\xi^4 + e_4(\xi), \tag{A.2}$$

with

$$e_4(\xi) = \int_0^\xi \frac{(\xi - t)^4}{(1+t)^5}\,\mathrm{d}t. \tag{A.3}$$

By inserting in (A.1), after some algebra, we get

$$F(n) = \frac{(n-\lambda)^2}{\lambda} - \frac{1}{3}\frac{(n-\lambda)^3}{\lambda^2} + \frac{1}{6}\frac{(n-\lambda)^4}{\lambda^3} - \frac{1}{2}\frac{(n-\lambda)^5}{\lambda^4} + 2ne_4\left(\frac{n-\lambda}{\lambda}\right), \tag{A.4}$$

so that the expected value of the random variable is given by

$$E\{F(Y_\lambda)\} = \frac{\mu_2}{\lambda} - \frac{1}{3}\frac{\mu_3}{\lambda^2} + \frac{1}{6}\frac{\mu_4}{\lambda^3} - \frac{1}{2}\frac{\mu_5}{\lambda^4} + 2E\left\{Y_\lambda e_4\left(\frac{Y_\lambda - \lambda}{\lambda}\right)\right\}, \tag{A.5}$$

where μ_k are the central moments:

$$\mu_k = E\{(Y_\lambda - \lambda)^k\}. \tag{A.6}$$

In the case of a Poisson random variable, we have [24]

$$\mu_2 = \lambda, \qquad \mu_3 = \lambda, \qquad \mu_4 = 3\lambda^2 + \lambda, \qquad \mu_5 = 10\lambda^2 + \lambda, \tag{A.7}$$

so that

$$E\{F(Y_\lambda)\} = 1 + \frac{1}{6\lambda} - \frac{29}{6\lambda^2} - \frac{1}{2\lambda^3} + 2E\left\{Y_\lambda e_4\left(\frac{Y_\lambda - \lambda}{\lambda}\right)\right\}, \tag{A.8}$$

with the corresponding inequality

$$|E\{F(Y_\lambda)\} - 1| \leqslant \frac{1}{6\lambda} + \frac{29}{6\lambda^2} + \frac{1}{2\lambda^3} + 2\left|E\left\{Y_\lambda e_4\left(\frac{Y_\lambda - \lambda}{\lambda}\right)\right\}\right|. \tag{A.9}$$

In order to estimate the last term in (A.9), we remark that

$$e_4(\xi) \leqslant \xi^4 \int_0^\infty \frac{\mathrm{d}t}{(1+t)^5} = \frac{1}{4}\xi^4 \tag{A.10}$$

so that

$$\left| E\left\{ Y_\lambda e_4\left(\frac{Y_\lambda - \lambda}{\lambda}\right)\right\}\right| \leqslant \frac{1}{4} E\left\{ Y_\lambda \left(\frac{Y_\lambda - \lambda}{\lambda}\right)^4\right\} = \frac{\mu_4}{4\lambda^3} + \frac{\mu_5}{4\lambda^4} = O\left(\frac{1}{\lambda}\right). \tag{A.11}$$

and the lemma is proved. □

Appendix B

In this appendix, we derive the SGM iterative algorithm that motivates the choice of the scaling factor proposed in section 3.3.

The splitting of the gradient of the complete functional is given in (21). Then, the first KKT condition for the non-negative minimum points x_β^* of the functional (3), $x_\beta^* \nabla J_y(x_\beta^*) = 0$, can be written as the following fixed point equation:

$$x_\beta^* = \frac{x_\beta^*}{1 + \beta V^{(R)}(x_\beta^*)} \left\{ \frac{y}{x_\beta^*} + \beta U^{(R)}(x_\beta^*)\right\}. \tag{B.1}$$

By formally applying the method of successive approximations, we get the iterative algorithm

$$x^{(k+1)} = \frac{x^{(k)}}{1 + \beta V^{(R)}(x^{(k)})} \left\{ \frac{y}{x^{(k)}} + \beta U^{(R)}(x^{(k)})\right\}, \tag{B.2}$$

which can be written in the following form:

$$x^{(k+1)} = x^{(k)} - \frac{x^{(k)}}{1 + \beta V^{(R)}(x^{(k)})} \nabla J_y(x^{(k)}). \tag{B.3}$$

This iteration describes the SGM algorithm that can be considered a scaled gradient method with step length 1. We also remark that, in such a way, the non-negativity of the iterates is automatically satisfied by (B.2). The convergence of this algorithm is not proved; in [25], convergence is obtained by introducing a step length and applying the Armijo rule inside the interval of values that ensure the non-negativity of the iterates.

Finally, we point out the main difference between this algorithm and the OSL algorithm proposed in [23]. Indeed, in the case of denoising, OSL can be formally derived by writing the first KKT condition as the following fixed point equation:

$$x_\beta^* = \frac{x_\beta^*}{1 + \beta \nabla J^{(R)}(x_\beta^*)} \frac{y}{x_\beta^*}, \tag{B.4}$$

so that, by again applying the successive approximation method, we get

$$x^{(k+1)} = \frac{x^{(k)}}{1 + \beta \nabla J^{(R)}(x^{(k)})} \frac{y}{x^{(k)}}, \tag{B.5}$$

which can be written in the following form:

$$x^{(k+1)} = x^{(k)} - \frac{x^{(k)}}{1 + \beta \nabla J^{(R)}(x^{(k)})} \nabla J_y(x^{(k)}). \tag{B.6}$$

Also in this case, it is possible to get convergence by introducing a suitable and sufficiently small step length.

A comparison of (B.2) and (B.6) shows the main difference between the two methods. In SGM, at the denominator of the scaling, we have only a positive part of the gradient of the regularization functional, while, in the OSL algorithm, we have the complete gradient, which can also take, of course, negative values. Therefore, the scaling of OSL may not be positive everywhere if β is not sufficiently small.

References

[1] Alvarez L, Lions P L and Morel J M 1992 Image selective smoothing and edge detection by nonlinear diffusion II *SIAM J. Numer. Anal.* **29** 845–66

[2] Anscombe F J 1948 The transformation of Poisson, binomial and negative-binomial data *Biometrika* **35** 246–54

[3] Bardsley J M 2008 An efficient computational method for total variation-penalized Poisson likelihood estimation *Inverse Problems Imaging* **2** 167–85

[4] Bardsley J and Nagy J 2006 Covariance-preconditioned iterative methods for non-negatively constrained astronomical imaging *SIAM J. Matrix Anal. Appl.* **27** 1184–98

[5] Barzilai J and Borwein J M 1988 Two point step size gradient methods *IMA J. Numer. Anal.* **8** 141–48

[6] Bonettini S, Zanella R and Zanni L 2009 A scaled gradient projection method for constrained image deblurring *Inverse Problems* **25** 015002

[7] Buades A, Coll B and Morel J M 2005 A review of image denoising algorithms with a new one *Multiscale Model. Simul.* **4** 490–530

[8] Calvetti D and Somersalo E 2007 A Gaussian hypermodel for recovering blocky objects *Inverse Problems* **23** 733–54

[9] Chan R H and Chen K 2007 Multilevel algorithms for a Poisson noise removal model with total-variation regularization *Int. J. Comput. Math.* **84** 1183–98

[10] Chan T F and Shen J 2005 *Image Processing and Analysis* (Philadelphia, PA: SIAM)

[11] Charbonnier P 1994 Reconstruction d'image: Régularization avec prise en compte des discontinuités *PhD dissertation* Univ. Nce-Sophia Antipolis, available at http://picabia.u-strasbg.fr/lsiit/perso/Charbonnier.htm

[12] Charbonnier P, Blanc-Féraud L, Aubert G and Barlaud A 1997 Deterministic edge-preserving regularization in computed imaging *IEEE Trans. Image Process.* **6** 298–311

[13] Coifman R R and Donoho D 1995 Translation-invariant denoising *Lecture Notes in Statistics* vol 103 (Berlin: Springer) pp 125–50

[14] Csiszár I 1991 Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems *Ann. Stat.* **19** 2032–66

[15] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)

[16] Dai Y H, Hager W W, Schittkowski K and Zhang H 2006 The cyclic Barzilai–Borwein method for unconstrained optimization *IMA J. Numer. Anal.* **26** 604–27

[17] Donoho D 1995 De-noising by soft thresholding *IEEE Trans. Inf. Theory* **41** 613–27

[18] Figueiredo M A T, Nowak R D and Wright S J 2007 Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems *IEEE J. Sel. Top. Signal Process.* **1** 586–97

[19] Frassoldati G, Zanghirati G and Zanni L 2008 New adaptive stepsize selections in gradient methods *J. Ind. Manag. Optim.* **4** 299–312

[20] Geman S and Geman D 1984 Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–41

[21] Geman S and McClure D E 1985 Bayesian image analysis: an application to single photon emission tomography *Proc. Statistical Computation Section* (Washington, DC: American Statistical Association) pp 12–8

[22] Geman S, Manbeck K and McClure D E 1993 A comprehensive statistical model for single photon emission tomography *Markov Random Fields: Theory and Applications* ed R Chellappa and A Jain (Boston: Academic) pp 93–130

[23] Green J P 1990 Bayesian reconstructions from emission tomography data using a modified EM algorithm *IEEE Trans. Med. Imaging* **9** 84–93

[24] Johnson N L, Katz S and Kemps A W 1993 *Univariate Discrete Distributions* (New York: Wiley) p 157

[25] Lanteri H, Roche M and Aime C 2002 Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms *Inverse Problems* **18** 1397–419

[26] Le T, Chartran R and Asaki T J 2007 A variational approach to reconstructing images corrupted by Poisson noise *J. Math. Imaging Vis.* **27** 257–63

[27] Osher S, Burger M, Goldfarb D, Xu J and Yin W 2005 An iterative regularization method for total-variation based image restoration *Multiscale Model. Simul.* **4** 460–89

[28] Perona P and Malik J 1990 Scale space and edge detection using anisotropic diffusion *IEEE Trans. Pattern Anal. Mach. Intell.* **12** 629–39

[29] Politte D G and Snyder D L 1991 Correction for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron-emission tomography *IEEE Trans. Med. Imaging* **10** 82–9

[30] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica* D **60** 259–68

[31] Vogel C 2002 *Computational Methods for Inverse Problems* (Philadelphia, PA: SIAM)

[32] Zhang B, Fadili J M and Starck J-L 2008 Wavelets, ridgelets and curvelets for Poisson noise removal available at http://jstarck.free.fr/IEEE_Poisson08.pdf

[33] Zhou B, Gao L and Dai Y H 2006 Gradient methods with adaptive step-sizes *Comput. Optim. Appl.* **35** 69–86