PAPER

Analysis of an approximate model for Poisson data reconstruction and a related discrepancy principle

Inverse Problems **27** (2011) 125003 (20pp) doi:10.1088/0266-5611/27/12/125003

Analysis of an approximate model for Poisson data reconstruction and a related discrepancy principle

A Staglianò, P Boccacci and M Bertero

Dipartimento di Informatica e Scienze dell'Informazione, Università di Genova, Via Dodecaneso 35, I 16146 Genova, Italy

E-mail: bertero@disi.unige.it

Received 16 February 2011, in final form 12 July 2011
Published 4 November 2011
Online at stacks.iop.org/IP/27/125003

Abstract

In this paper, we investigate an approximate model for Poisson data reconstruction inspired by a discrepancy principle for the selection of the regularization parameter, recently proposed by Bardsley and Goldes. The model can be obtained by approximating the generalized Kullback–Leibler (KL) divergence in terms of a weighted least-squares function, with weights depending on the object to be reconstructed. We show that it is possible to develop a complete theory, based on this approximation, including results of existence and uniqueness of regularized solutions and simple gradient-based reconstruction algorithms for their computation. Moreover, in this context, the criterion of Bardsley and Goldes is a natural one and it is possible to prove that, in several important cases, it provides a unique value of the regularization parameter. We describe a few numerical tests for comparing the approximate approach with the exact one based on the generalized KL divergence. In the case of a moderate or large number of photons, they provide essentially the same results and therefore the approximate model can be considered as a possible alternative to the exact one.

1. Introduction

The maximum likelihood approach to Poisson data reconstruction [31] leads to the minimization of a data-fidelity function defined in terms of the generalized Kullback–Leibler (KL) divergence as follows [7]:

$$f_0^{(P)}(x;y) = \sum_{i\in S}\left\{y_i \ln\frac{y_i}{(Hx+b)_i} + (Hx+b)_i - y_i\right\}, \qquad x \geqslant 0, \tag{1}$$

where y is the detected image, with components labeled by the multi-index $i \in S$, H is the imaging matrix, x is the unknown object to be reconstructed and b is an array/cube describing the expected value of the background emission. Denoising corresponds to $H = I$, the unit matrix; moreover, background emission can be included in the unknown object x [36]

(i.e. $b = 0$ in (1)). The approach can also be used when, in addition to Poisson noise due to photon counting, an additive Gaussian noise with zero mean and given variance σ^2 is present [32], the so-called read-out noise (RON). Indeed, as shown in [33], it is possible to approximate the RON by a Poisson process with the expected value σ^2 if σ^2 is added to each component of the image y and of the background b.

Since as a consequence of noise and ill-posedness of the inverse problem, the minimizers of the function $f_0^{(P)}(x; y)$ are sparse, the so-called *night-sky* [5] or *checkerboard* [27] objects, in general they are not reliable solutions of the image reconstruction problem, except if the object itself is sparse, for instance a star cluster in astronomical imaging [7]. Therefore, the minimization of the function $f_0^{(P)}(x; y)$ is replaced by the minimization of a penalized function with the following structure:

$$f_\beta^{(P)}(x; y) = f_0^{(P)}(x; y) + \beta f_1(x), \qquad x \geqslant 0, \tag{2}$$

where β plays the role of a regularization parameter and $f_1(x)$ is a suitable penalty function enforcing properties of the solution such as smoothness, edges etc. This approach can be justified in a Bayesian formulation of the problem (see, for instance, [20, 21], or [7] for a review) and is now frequently used in image reconstruction with Poisson data.

The standard algorithm for the minimization of $f_0^{(P)}(x; y)$ is the so-called Richardson–Lucy (RL) algorithm [26, 29] (also called *expectation maximization* in emission tomography [31]), defined by the following iteration:

$$x^{k+1} = \frac{x^k}{h} H^T \frac{y}{Hx^k + b}, \tag{3}$$

where products and quotients of arrays/cubes are intended pixel by pixel or voxel by voxel, and $h = H^T e$, with e being the array/cube with all entries being equal to 1. In the case $b = 0$, convergence of these algorithms has been proved by several authors (see, for instance, [27]); the convergence of the continuous version is investigated in [28]. The algorithm has several advantages such as non-negativity of the iterates, flux conservation (in the case $b = 0$) and semi-convergence, but it is slow.

In view of obtaining more efficient algorithms, several authors have considered an approach which can be used in the case of large photon counts and consists in approximating the KL divergence with a weighted least-squares function, the weights being the inverse of the measured data. This approach is used both in medical imaging (see, for instance, [18, 22]) and image deconvolution [1, 2]. It allows the use of efficient methods introduced for solving least-squares problems as well as the application of related selection rules for the choice of the regularization parameter.

In a recent paper [4], Bardsley and Goldes used this quadratic approximation to the generalized KL divergence for introducing selection rules of the regularization parameter β in (2); in particular, the well-known Morozov discrepancy principle [17] has been used. However, the authors claim that, on the basis of numerical simulations, more accurate results are obtained if, in the weights of the quadratic approximation, the measured data are replaced by the computed data which depend on x. A theoretical difficulty of this modified criterion is that it seems difficult to prove that it is able to provide a value, possibly a unique value, of the regularization parameter. Another discrepancy principle, with good theoretical properties, is proposed in [8] and is based directly on the generalized KL divergence rather than on its quadratic approximation. Preliminary results on the comparison of the two discrepancy principles in the deconvolution of 3D confocal data are discussed in [12, 13].

In this paper, we show that the discrepancy principle proposed in [4] has good theoretical properties in the context of a model for Poisson data reconstruction based on a suitable

approximation of the KL divergence which can be introduced in a simple way as follows. We consider the function

$$\phi_y(x) = y \ln \frac{y}{x} + x - y, \qquad x, y > 0 \tag{4}$$

and represent the logarithm as a power series of $(x-y)/x$; then, after some manipulation, we obtain

$$\phi_y(x) = \frac{1}{2}\frac{(x-y)^2}{x}\left\{1 + \sum_{n=1}^{\infty} \frac{2}{(n+1)(n+2)}\left(\frac{x-y}{x}\right)^n\right\}, \tag{5}$$

with the series being convergent for $x > y/2$. It is easy to check that, for moderate and large values of y (let us say $y \geqslant 10$), the leading term of this series

$$\psi_y(x) = \frac{1}{2}\frac{(x-y)^2}{x} \tag{6}$$

provides a sufficiently accurate approximation of $\phi_y(x)$ for $|x-y| \leqslant \sqrt{y}$. If we apply this approximation to (1) we obtain the new data-fidelity function

$$f_0(x; y) = \frac{1}{2}\sum_{i \in S} \frac{(Hx + b - y)_i^2}{(Hx + b)_i}, \qquad x \geqslant 0, \tag{7}$$

and the penalized function (2) is replaced by the following one:

$$f_\beta(x; y) = f_0(x; y) + \beta f_1(x), \qquad x \geqslant 0. \tag{8}$$

The function $f_0(x; y)$ is just the basis of the discrepancy principle proposed in [4] in the case of deblurring and in [25] in the case of denoising for the choice of β in (2). Since by extending results proved in [8, 10], we show that it is possible to develop an approach to Poisson data reconstruction based on the data-fidelity function (7), this context is the natural one for discussing the criterion proposed in [4, 25] for the selection of β in (8); indeed, we can prove existence and uniqueness of the value provided by this criterion, at least for some relevant penalty functions.

In section 2, we derive the properties of the data-fidelity function (7). In section 3, we discuss the minimization of the penalized function (8) in the case of some relevant penalty functions and prove results of existence and uniqueness of the minimizer. In section 4, we discuss the criterion proposed in [4, 25] as a criterion for the selection of β in (8) and give conditions for existence and uniqueness of the selected value. Section 5 is devoted to the discussion of reconstruction algorithms, and we show that gradient-based algorithms for the minimization of (7) or (8) can be obtained by a simple modification of gradient-based algorithms for the minimization of (1) or (2), respectively. Finally, in section 6, by means of a few numerical tests, we compare the results provided by the approximate approach with those provided by the exact one, i.e. that based on the KL divergence (1).

2. Properties of the data-fidelity function

We assume that the components of the object x are labeled by a multi-index j with a range R so that, in the case of deblurring, we write

$$(Hx)_i = \sum_{j \in R} H_{i,j} x_j, \qquad i \in S. \tag{9}$$

In addition, we assume that the imaging matrix H satisfies the conditions

$$H_{i,j} \geqslant 0; \qquad \sum_{i \in S} H_{i,j} > 0, \forall j \in R; \qquad \sum_{j \in R} H_{i,j} > 0, \forall i \in S. \tag{10}$$

In other words, for each fixed value of the multi-index i or j, there exists at least one non-zero entry. We denote by h and $\tilde{h}$ the arrays/cubes with components

$$h_j = \sum_{i \in S} H_{i,j},\ j \in R; \qquad \tilde{h}_i = \sum_{j \in R} H_{i,j},\ i \in S. \tag{11}$$

We denote by M the cardinality of S and by N the cardinality of R. In the case of denoising, $H = I$, we have $i = j$ and $N = M$.

In this section and in the following, products, quotients, powers, square roots etc of arrays/cubes are intended pixel by pixel or voxel by voxel.

Theorem 1. *The data-fidelity function (7) is non-negative, convex and coercive on the non-negative orthant; its gradient and Hessian are given by*

$$\begin{aligned} \nabla_x f_0(x; y) &= \frac{1}{2} h - \frac{1}{2} H^T \left(\frac{y}{Hx + b} \right)^2, \\ \nabla_x^2 f_0(x; y) &= H^T \operatorname{diag} \left(\frac{y_i^2}{(Hx + b)_i^3} \right) H. \end{aligned} \tag{12}$$

Proof. The non-negativity of the function is obvious. The expressions of the gradient and Hessian can be easily obtained by elementary computations. The Hessian is positive semi-definite, and its null space is given by

$$\mathcal{N}[\nabla^2 f_0(x; y)] = \{u | (Hu)_i = 0, \text{ if } y_i > 0\}. \tag{13}$$

Therefore, in the case of denoising ($H = I, b = 0$), the null space is not trivial if some components of y are zero while, in the case of deblurring, it contains the null space of H. Incidentally, we remark that it coincides with the null space of the Hessian of the function (1) [8]. It follows that the function (7) is convex and strictly convex iff $y > 0$ and the null space of H is trivial.

Finally, the coercivity of $f_0(x; y)$ is obvious in the case of denoising. In the case of deblurring, we first prove that $||Hx||_2$ is coercive. This result is also obvious if the null space of H is trivial. If it is not trivial, the assumed properties of H imply that all the elements of its null space must have at least one negative component. It follows that $||Hx||_2$ cannot be zero on the intersection of the non-negative orthant with the unit sphere, so that it has a positive minimum value α. Since it is homogeneous of order 1, we obtain $||Hx||_2 \geqslant \alpha ||x||_2$, $x \geqslant 0$, and therefore coercivity holds. If we write

$$\frac{(Hx + b - y)_i^2}{(Hx + b)_i} = (Hx + b)_i - 2y_i + \frac{y_i^2}{(Hx + b)_i}, \tag{14}$$

from the inequality $||Hx||_1 \geqslant ||Hx||_2 \geqslant \alpha ||x||_2$, the coercivity of $f_0(x; y)$ follows. □

The previous results imply that $f_0(x; y)$ *has absolute minimizers on the non-negative orthant*, i.e. there exist solutions of the following optimization problem:

$$\begin{aligned} &\text{minimize } f_0(x; y) = \frac{1}{2} \left\| \frac{Hx + b - y}{\sqrt{Hx + b}} \right\|_2^2 \\ &\text{subject to } x \geqslant 0. \end{aligned} \tag{15}$$

We denote as x^* any solution of this problem. In the case of denoising ($H = I$ and $b = 0$), the solution is unique and is trivial: $x^* = y$. In the case of deblurring, the noise and the ill-posedness of the problem imply that the solutions (or the solution if $f_0(x; y)$ is strictly

convex) are sparse, as follows from a general analysis of the non-negative minimizers of non-regularized data-fidelity functions [5]. Therefore, we do not obtain reliable reconstructions, except for particular objects such as star clusters in astronomical imaging.

We conclude with the following result which will be used in section 4.

Lemma 1. *In the case of deblurring, $x = 0$ is not a minimizer of $f_0(x; y)$ if*

$$\sum_{i\in S} \tilde{h}_i \left\{ \left(\frac{y_i}{b_i} \right)^2 - 1 \right\} > 0. \tag{16}$$

In the particular case $\tilde{h}_i = 1$ and $b_i = b$ (constant background), the condition becomes

$$y_{\text{rms}} = \sqrt{\frac{1}{M} \sum_{i\in S} y_i^2} > b, \tag{17}$$

i.e. the root-mean-square (rms) value of the data must be greater than the constant background.

Proof. Let us consider the one-dimensional subspace of the constant vectors, $x = ce_N$, where e_N is the N-dimensional array/cube with all entries being equal to 1, and the restriction of $f_0(x; y)$ to this subspace, i.e. the function

$$\psi(c) = \frac{1}{2} \sum_{i\in S} \frac{(\tilde{h}_i c + b_i - y_i)^2}{\tilde{h}_i c + b_i}. \tag{18}$$

The null element is not a minimizer if $\psi'(0) < 0$. With some manipulation, we obtain

$$\psi'(0) = \frac{1}{2} \sum_{i\in S} \tilde{h}_i \left\{ 1 - \left(\frac{y_i}{b_i} \right)^2 \right\}, \tag{19}$$

and the result follows. □

Remark 1. The conditions $\tilde{h}_i = 1$ are satisfied, as well as the conditions $h_j = 1$, if the imaging matrix H is the cyclic convolution with a point-spread function (PSF) K normalized to 1 (in this case $S = R$)

$$Hx = K * x, \qquad \sum_{i\in S} K_i = 1. \tag{20}$$

This model is frequently used in the applications to astronomy and microscopy.

3. The regularized problem

In the regularized approach, problem (15) is replaced by the following one:

$$\begin{aligned} &\text{minimize } f_\beta(x; y) = f_0(x; y) + \beta f_1(x) \\ &\text{subject to } x \geqslant 0, \end{aligned} \tag{21}$$

and in this section we consider the following regularization functions given, for simplicity, in the 2D case, i.e. the multi-index j is a pair of indices $j = \{j_1, j_2\}$ taking values from 1 to n, so that $N = n^2$.

- Tikhonov regularization

$$f_1(x) = \frac{1}{2} \sum_{j_1, j_2 = 1}^{n} x_{j_1, j_2}^2. \tag{22}$$

- Tikhonov gradient regularization

$$f_1(x) = \frac{1}{2} \sum_{j_1, j_2=1}^{n} D^2_{j_1, j_2}. \tag{23}$$

- Hyper-surface regularization [14]

$$f_1(x) = \sum_{j_1, j_2=1}^{n} \psi_\delta\left(D^2_{j_1, j_2}\right), \qquad \delta \neq 0. \tag{24}$$

- Markov random field regularization [20]

$$f_1(x) = \frac{1}{2} \sum_{j_1, j_2=1}^{n} \sum_{k_1, k_2 \in \mathcal{N}_{j_1, j_2}} \psi_\delta \left[\left(\frac{x_{j_1, j_2} - x_{k_1, k_2}}{w_{k_1, k_2}} \right)^2 \right], \tag{25}$$

  with $\mathcal{N}_{j_1, j_2}$ being the set of the first neighbors of j_1, j_2 and $w_{k_1, k_2} = 1$ for horizontal and vertical neighbors, and $w_{k_1, k_2} = \sqrt{2}$ for diagonal neighbors; these are included for providing an isotropic edge-preserving effect, as discussed in [20].

In these equations

$$\psi_\delta(t) = \sqrt{t + \delta^2}, \quad D^2_{j_1, j_2} = \left(x_{j_1+1, j_2} - x_{j_1, j_2}\right)^2 + \left(x_{j_1, j_2+1} - x_{j_1, j_2}\right)^2. \tag{26}$$

For δ being small, the regularization (24) is used as a smoothed approximation to total variation (TV) (see for instance [3, 15, 35, 36]; for TV regularization, see [11, 16, 24, 30]).

All the functions above are convex and therefore the results of section 2 imply that the function $f_\beta(x; y)$ is convex and coercive. In the case of Tikhonov regularization (22), the function is strictly convex and this property also holds true in the other cases, with some additional specification in the definition of the penalty functions. Indeed, the following result is proved in [10].

Lemma 2. *If the functions $f_1(x)$, defined in equations (23)–(25), are extended by periodic or reflexive boundary conditions, then the null space of the Hessian matrix $\nabla^2_x f_1(x)$ is given by the set of minimum points of $f_1(x)$.*

In all the cases indicated above, the set of the minimizers of $f_1(x)$ has dimension 1 and contains only the constant arrays. Since we have $\nabla^2_x f_\beta(x; y) = \nabla^2_x f_0(x; y) + \beta \nabla^2_x f_1(x)$ and all the Hessians are positive semi-definite, the null space of $\nabla^2_x f_\beta(x; y)$ is the intersection of the null spaces of $\nabla^2_x f_0(x; y)$ and $\nabla^2_x f_1(x)$, and this intersection contains only the null element. In conclusion, we have the following result.

Theorem 2. *The function $f_\beta(x; y)$ is strictly convex for all the regularizers introduced above and therefore the solution of the optimization problem (21) is unique. It will be denoted as x^*_β.*

From the analysis performed in [10], one can easily obtain the following result which will be used in the following section.

Lemma 3. *If the functions $f_1(x)$ defined in equations (23)–(25) are extended up to the boundary by periodic or reflexive boundary conditions, then the following identity holds true:*

$$\sum_{j \in R} [\nabla_x f_1(x)]_j = 0. \tag{27}$$

4. Selection of the regularization parameter

In this section, we first show that the discrepancy principle, empirically introduced by Bardsley and Goldes [4] for the selection of β in (2), is a natural one in the framework outlined in the previous sections and prove that it can allow the selection of a unique value of β in (8).

As is well known, the expected value and the variance of a Poisson random variable coincide, i.e. if Y_λ is Poisson with the expected value λ, then

$$E\left\{\frac{1}{\lambda}(Y_\lambda - \lambda)^2\right\} = 1. \tag{28}$$

Therefore, if $\tilde{x}$ is the unknown object which generates the data y (i.e. the y_i are realizations of $Y_i = \text{Poisson}\{H\tilde{x} + b\}_i$), then each term of $f_0(\tilde{x}; y)$ is the realization of a random variable with the expected value 1 and the values of these terms fluctuate around 1. Since the positive fluctuations in general compensate the negative ones, we expect that the value of $f_0(\tilde{x}; y)$ is approximately $M/2$.

This argument motivates the following recipe, which is the application to (8) of the selection rule given in [4, 25]: if x^*_β is the (unique) solution of (21), then search for a value of β such that the function of β defined by

$$D_y(\beta) = \frac{2}{M} f_0(x^*_\beta : y) = \frac{1}{M} \sum_{i \in S} \frac{(Hx^*_\beta + b - y)_i^2}{(Hx^*_\beta + b)_i} \tag{29}$$

obeys

$$D_y(\beta) = 1. \tag{30}$$

We call $D_y(\beta)$ the *discrepancy function* and equation (30) the *discrepancy equation*.

We stress that this criterion is proposed in [4, 25] for problems where the data-fidelity function is given by (1) while, in this paper, is considered for problems where the data-fidelity function is given by (7). As a general comment, we point out that it is possible to prove results of uniqueness and existence of the solution of the discrepancy equation if the data-fidelity function used in the definition of the discrepancy function coincides with the data-fidelity function used in the optimization problem.

Since x^*_β is unique for the optimization problems discussed in the previous section, for these problems the discrepancy function is well defined. Moreover, an important property of this function derives from this lemma proved in [8].

Lemma 4. *If the function $f_\beta(x; y)$ is strictly convex, coercive and differentiable on the non-negative orthant and x^*_β is the unique minimizer, then $f_0(x^*_\beta; y)$ and $f_1(x^*_\beta)$ are respectively an increasing and a decreasing function of β.*

The assumptions of the lemma are satisfied by the problems of the previous section so that, for these problems, the discrepancy function is an increasing function of β and the solution of the discrepancy equation, if it exists, is unique. As concerns existence, we must investigate the behavior of x^*_β for $\beta \to 0$ and $\beta \to \infty$. Moreover, for generality, we consider a slight modification of equation (30)

$$D_y(\beta) = d, \tag{31}$$

where d is a number close to 1.

The case of Tikhonov regularization is discussed in appendix A. Here, we give the results concerning the three other regularizers introduced in section 3. The proofs are given in appendix B.

Lemma 5. *Let x^*_β be the unique minimizer of $f_\beta(x; y)$ when $f_1(x)$ is one of the penalties defined in (23)–(25); then*

$$\lim_{\beta\to 0} x^*_\beta = x^*, \tag{32}$$

where x^* is a minimizer of $f_0(x; y)$; in the case of denoising $x^* = y$. Moreover,

$$\lim_{\beta\to\infty} x^*_\beta = c_\infty e_N. \tag{33}$$

In the case of denoising

$$c_\infty = y_{\text{rms}}, \tag{34}$$

while in the case of deblurring, if condition (16) is satisfied, then c_∞ is the unique solution of the transcendental equation

$$\sum_{i\in S} \tilde{h}_i \left\{1 - \left(\frac{y_i}{\tilde{h}_i c_\infty + b_i}\right)\right\} = 0. \tag{35}$$

In the particular case $\tilde{h}_i = 1$, $b_i = b$ the solution is

$$c_\infty = y_{\text{rms}} - b, \tag{36}$$

a positive constant thanks to (17).

Theorem 3. *There exists a unique solution of equation (31) if d satisfies the conditions*

$$D_y(0) < d < D_y(\infty). \tag{37}$$

In the case of denoising

$$D_y(0) = 0, D_y(\infty) = 2(y_{\text{rms}} - y_{\text{am}}), \tag{38}$$

where y_{am} is the arithmetic mean of the values of y. In the case of deblurring

$$D_y(0) = \frac{1}{M}\left\|\frac{Hx^* + b - y}{\sqrt{Hx^* + b}}\right\|^2, D_y(\infty) = \frac{1}{M}\sum_{i\in S}\frac{(\tilde{h}_i c_\infty + b_i - y_i)^2}{\tilde{h}_i c_\infty + b_i}. \tag{39}$$

If $\tilde{h}_i = 1$, $b_i = b$, then we have, as in the denoising case

$$D_y(\infty) = 2(y_{\text{rms}} - y_{\text{am}}). \tag{40}$$

Remark 2. For the denoising problem, if $y_{\text{rms}} > y_{\text{am}} + 1/2$, then the discrepancy equation (30) has a unique solution. In the case of deblurring, the value of the discrepancy function at $\beta = 0$ can be obtained by computing x^* by means of the fast algorithm discussed in the following section.

5. Algorithms

We consider first the optimization problem (15). From the first Karush–Kuhn–Tucker (KKT) condition $x^*\nabla f_0(x^*; y) = 0$, we obtain the fixed point equation

$$x^* = \frac{x^*}{h} H^T \left(\frac{y}{Hx^* + b}\right)^2, \tag{41}$$

which suggests the following iteration:

$$x^{k+1} = \frac{x^k}{h} H^T \left(\frac{y}{Hx^k + b}\right)^2 = x^k - \frac{2x^k}{h}\nabla_x f_0(x^k; y). \tag{42}$$

Convergence can be obtained by a line search along the descent direction but, from our point of view, the interesting feature of this iteration is that it defines a scaled gradient method, so that we can provide acceleration and convergence in the framework of the general approach known as a scaled-gradient projection (SGP) method [9]. The monotone SGP version of (42) is the algorithm we use for the computation of approximate solutions of the optimization problem (15).

In the case of regularized problems, we can introduce algorithms similar to (42) using the so-called split-gradient method [23]. It is based on a decomposition of the gradient of the penalty function in a positive and a negative parts:

$$-\nabla_x f_1(x) = U_1(x) - V_1(x) \tag{43}$$

with $U_1(x) \geqslant 0, V_1(x) \geqslant 0$ in the non-negative orthant. Then, from the first KKT condition one easily obtains the following fixed point equation:

$$x^* = \frac{x^*}{h + 2\beta V_1(x^*)} \left\{ H^T \left(\frac{y}{Hx^* + b} \right)^2 + 2\beta U_1(x^*) \right\}, \tag{44}$$

and (42) is replaced by the following iteration:

$$\begin{aligned} x^{k+1} &= \frac{x^k}{h + 2\beta V_1(x^k)} \left\{ H^T \left(\frac{y}{Hx^k + b} \right)^2 + 2\beta U_1(x^k) \right\} \\ &= x^k - \frac{2x^k}{h + 2\beta V_1(x^k)} \nabla_x f_\beta(x^k; y). \end{aligned} \tag{45}$$

Convergence of this algorithm can be obtained by introducing a step-length in the descent direction and performing a line search based, for instance, on the Armijo rule [23]. The limit is a minimizer of the regularized function. The convergence is not very fast; however, since it is a scaled-gradient method, acceleration and convergence can also be obtained in the framework of SGP. The choice of the functions $U_1(x), V_1(x)$ is largely arbitrary, but in [23] it is shown that a 'natural' choice exists for all the frequently used regularization functions. Examples of successful applications are given in [34, 36].

In conclusion, for the regularizers introduced in section 3, the arrays $V_1(x)$ can be chosen according to the following rules.

- Tikhonov regularization

$$V_1(x) = x. \tag{46}$$

- Tikhonov gradient regularization

$$V_1(x) = 4x. \tag{47}$$

- Hyper-surface regularization

$$[V_1(x)]_{j_1,j_2} = \left[2\psi'_\delta\left(D^2_{j_1,j_2}\right) + \psi'_\delta\left(D^2_{j_1,j_2-1}\right) + \psi'_\delta\left(D^2_{j_1-1,j_2}\right) \right]. \tag{48}$$

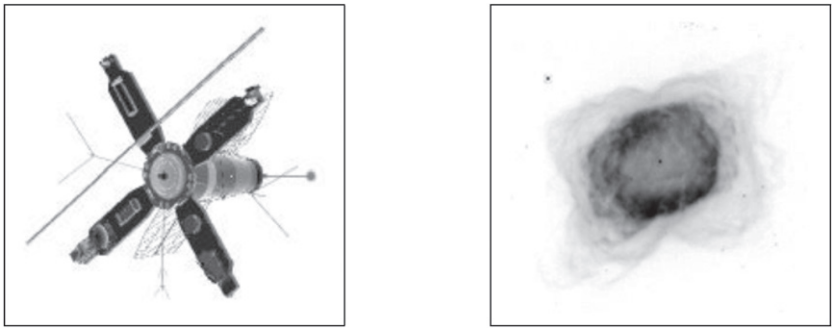
- Markov random field regularization

$$[V_1(x)]_{j_1,j_2} = \sum_{k_1,k_2 \in \mathcal{N}_{j_1,j_2}} \psi'_\delta \left[\left(\frac{x_{j_1,j_2} - x_{k_1,k_2}}{w_{k_1,k_2}} \right)^2 \right] \frac{x_{j_1,j_2}}{w_{k_1,k_2}}. \tag{49}$$

In these equations, $\psi'_\delta(t)$ is the derivative of $\psi_\delta(t)$.

Algorithm SGP-POS (scaled-gradient projection method)

Choose the starting point $x^{(0)} \geqslant 0$, set the parameters $\beta, \theta \in (0, 1)$, $0 < \alpha_{\min} < \alpha_{\max}$.
FOR $k = 0, 1, 2, \ldots$ do the following steps:
STEP 1. Choose the parameter $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$ and the scaling matrix S_k;
STEP 2. Projection: $y^{(k)} = P_+(x^{(k)} - \alpha_k S_k \nabla f(x^{(k)}))$;
If $y^{(k)} = x^{(k)}$ then stop, declaring that $x^{(k)}$ is a stationary point;
STEP 3. Descent direction: $d^{(k)} = y^{(k)} - x^{(k)}$;
STEP 4. Set $\lambda_k = 1$;
STEP 5. Backtracking loop:
If $f(x^{(k)} + \lambda_k d^{(k)}) \leqslant f(x^{(k)}) + \beta \lambda_k \nabla f(x^{(k)})^T d^{(k)}$ then
go to Step 6;
ELSE
set $\lambda_k = \theta \lambda_k$ and go to Step 5;
ENDIF
STEP 6. Set $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$.
END



(a) (b)

Figure 1. (a) The spacecraft and (b) the nebula NGC 7027.

We remark that all the algorithms described in this section can be obtained from similar gradient-like algorithms for the usual approach to Poisson data reconstruction by replacing the quotient $y/(Hx + b)$ with its square.

For the convenience of the reader, we give the algorithm SGP in its monotone version, with projection on the non-negative orthant (SGP-POS). The (diagonal) scaling matrices S_k for the different cases can be obtained from the previous equations. The step-lengths α_k are given by the two Barzilai–Borwein rules [6], modified in terms of the scaling and alternated according to a procedure proposed in [19]. Details are given in [9].

6. Numerical tests

We describe a few numerical tests performed for verifying the accuracy achievable with the approximate model investigated in this paper. We consider two reference objects: the frequently used spacecraft image, corresponding to an object with sharp details (figure 1(a)), and an image of the nebula NGC 7027 (figure 1(b)), an example of a diffuse astronomical object. Both objects are 256×256 (in figure 1 we show only their central parts), their maximum value is 255 and both are superimposed on a background $b = 1$. Moreover, for generating

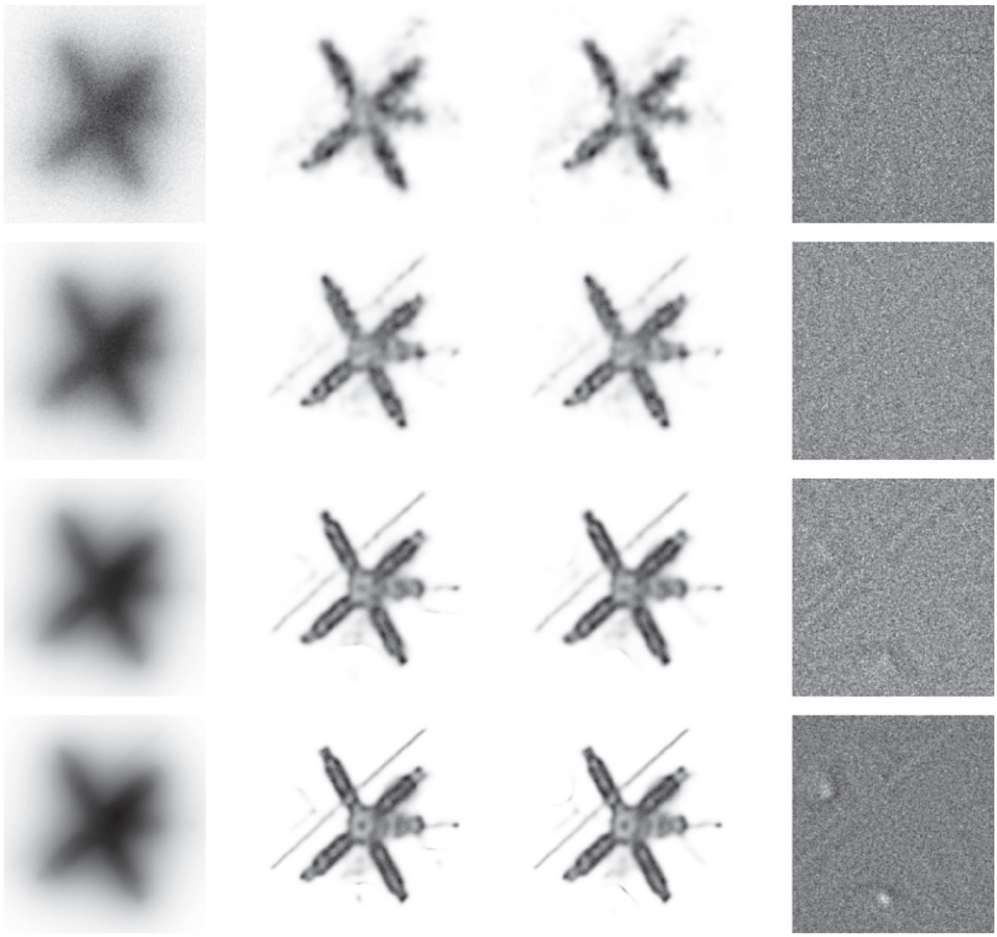


Figure 2. Non-regularized reconstructions of the spacecraft: (a) the blurred images, (b) the minimum rms reconstructions with the exact model, (c) those with the approximate model and (d) the normalized residuals in the case of the approximate model.

images with different noise levels, for each object we consider three other versions with the maximum values 2550, 25 500 and 255 000, respectively (and backgrounds 10, 100 and 1000), obtained by scaling the two original objects. Next, the four versions of each object are convolved with a PSF and then perturbed with Poisson noise (we did not add Gaussian noise). The PSF used simulates that taken by a ground-based telescope and is downloaded from http://www.mathcs.emory.edu/∼nagy/RestoreTools/index.html. For each image, we generate 25 different realizations of noise so that we have a total of 200 noisy images.

We first consider non-regularized reconstructions by comparing results obtained with the data-fidelity function (1) (referred to as the 'exact model') and those obtained with the data-fidelity function (7) (referred to as the 'approximate model'). We use the monotone SGP version of the RL algorithm (3) in the exact case and of the algorithm (42) in the approximate one.

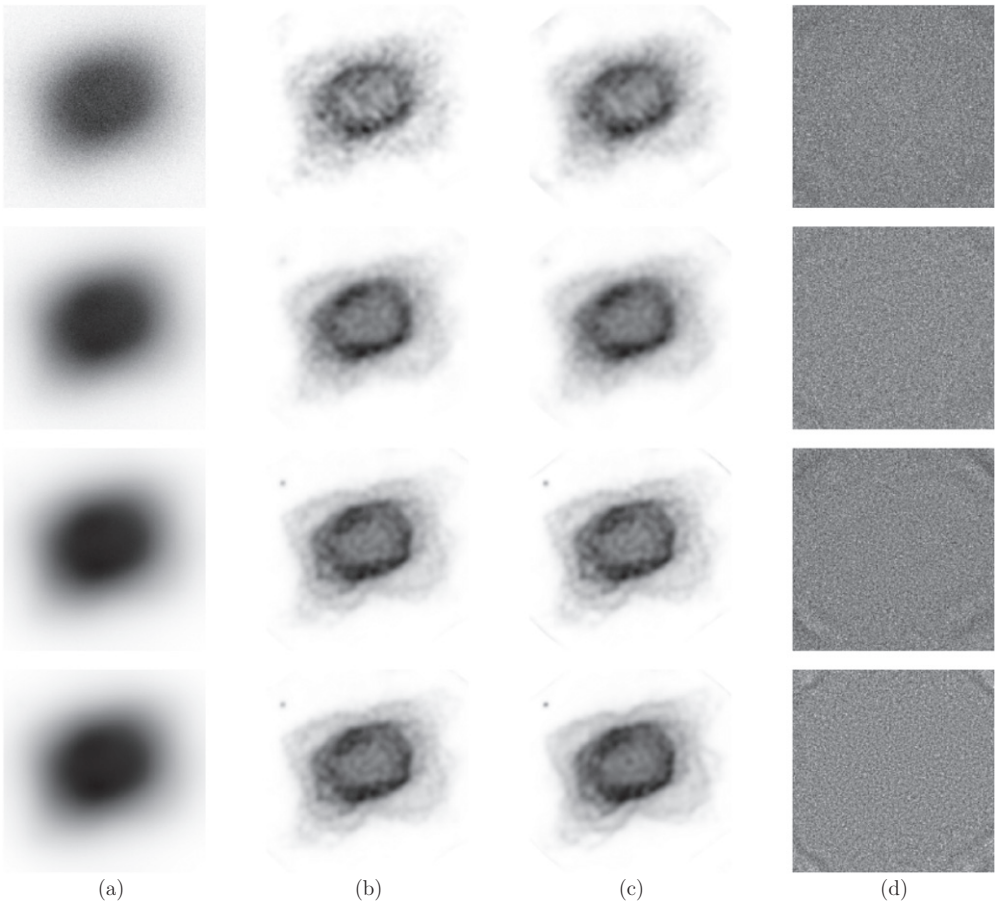


Figure 3. Non-regularized reconstructions of the nebula NGC 7027: (a) the blurred images, (b) the minimum rms reconstructions with the exact model, (c) those with the approximate model and (d) the normalized residuals in the case of the approximate model.

Early stopping of the iteration is considered in both cases. To this purpose two stopping criteria are used. Let $\rho^{(k)}$ and $D^{(k)}$ be defined as follows:

$$\rho^{(k)} = \frac{||x^k - \tilde{x}||}{||\tilde{x}||}, \qquad D^{(k)} = \frac{1}{M} \left\| \frac{Hx^k + b - y}{\sqrt{Hx^k + b}} \right\|^2, \tag{50}$$

where x^k is the result of the kth iteration and $\tilde{x}$ is the reference object. Then, the first criterion consists in computing at each iteration the relative rms error $\rho^{(k)}$ and stopping the iteration when this parameter reaches its minimum value. The second consists in computing the discrepancy $D^{(k)}$ and stopping the iteration when it crosses 1. Iteration is initialized with $x^0 = y_{\text{am}} - b$ and, in all cases, $D^{(0)} > 1$. In the approximate case, the value of $D^{(k)}$ is also the value of the objective function and therefore decreases for increasing iteration, since we use a monotone SGP algorithm.

The results are given in table 1 for the four images of the spacecraft with different noise levels. For each image, we report the average value and standard deviation of both the number of iterations and the reconstruction error, computed using the 25 realizations of noise. The

Table 1. Non-regularized reconstructions of the spacecraft: errors and iterations.

	Minimum error				Discrepancy			
	Exact		Approx		Exact		Approx	
	Iteration	Error (%)	Iteration	Error (%)	Iteration	Error (%)	Iteration	Error (%)
255	73 ± 19	40.1 ± 0.6	63 ± 6	40.2 ± 0.5	33 ± 14	43.9 ± 9.6	17 ± 1	54.2 ± 1.9
2550	186 ± 58	33.5 ± 0.4	178 ± 52	33.5 ± 0.4	366 ± 119	35.8 ± 2.2	103 ± 50	35.5 ± 1.4
25 500	465 ± 198	29.3 ± 0.3	395 ± 156	29.3 ± 0.2	593 ± 322	30.0 ± 1.1	581 ± 339	30.1 ± 1.2
255 000	1449 ± 376	26.9 ± 0.2	1439 ± 238	27.0 ± 0.1	1788 ± 553	27.3 ± 0.5	1742 ± 382	27.4 ± 0.5

Table 2. Non-regularized reconstructions of the nebula NGC 7027: errors and iterations.

	Minimum error				Discrepancy			
	Exact		Approx		Exact		Approx	
	Iteration	Error (%)	Iteration	Error (%)	Iteration	Error (%)	Iteration	Error (%)
255	33 ± 2	13.3 ± 0.3	33 ± 3	13.5 ± 0.3	24 ± 13	15.4 ± 3.2	14 ± 2	54.2 ± 1.9
2550	57 ± 5	10.4 ± 0.2	60 ± 7	10.4 ± 0.3	62 ± 23	10.9 ± 1.1	72 ± 24	11.1 ± 1.0
25 500	135 ± 17	7.4 ± 0.1	140 ± 23	7.5 ± 0.2	154 ± 36	7.5 ± 0.2	159 ± 17	7.6 ± 0.4
255 000	314 ± 35	5.5 ± 0.1	287 ± 56	5.6 ± 0.3	340 ± 37	5.6 ± 0.1	350 ± 32	5.6 ± 0.1

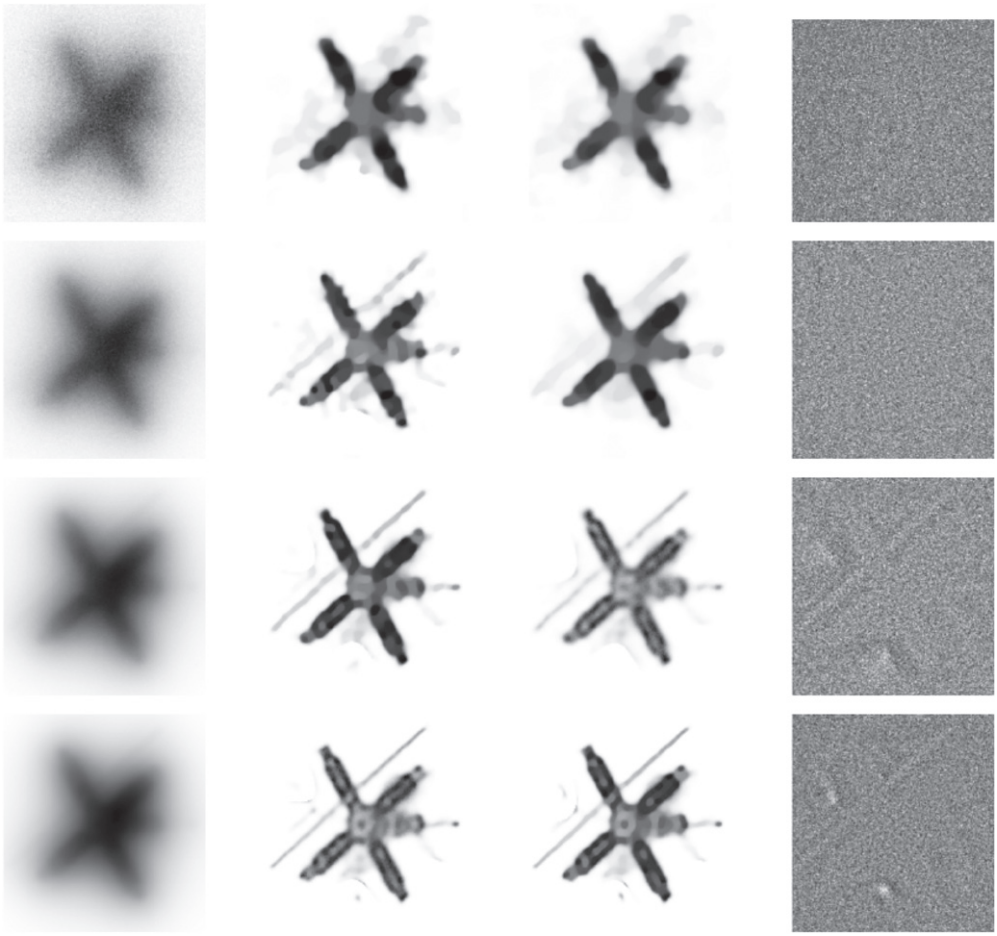


Figure 4. Regularized reconstructions of the spacecraft: (a) the blurred images, (b) the minimum rms reconstructions with the exact model, (c) those with the approximate model and (d) the normalized residuals in the case of the approximate model.

results obtained by stopping the iterations at the minimum error show that the two models are equivalent. The reconstruction error is weakly dependent on the noise realization even if the number of iterations is strongly varying. Stopping based on Bardsley and Goldes discrepancy principle works better for the exact model at the highest noise level, even if the standard deviation is larger than in the case of the approximate model.

In figure 2, we show a few reconstructions. In column (a), we show the four images with different noise levels, in columns (b) and (c) the reconstructions corresponding to the minimum rms error for the exact and the approximate model, respectively; finally, in the last column, we show the normalized residuals defined by

$$R^{(k)} = \frac{Hx^k + b - y}{\sqrt{Hx^k + b}}, \tag{51}$$

and computed in the case of the reconstructions of column (c). Artifacts are present at the lowest noise levels, due to the reconstruction method and not to the model.

The results for the nebula NGC 7027 are reported in table 2 and figure 3. Also, in this case the two models provide the same results with the same efficiency when iteration is stopped at the minimum error. If the discrepancy criterion is used, as in the case of the spacecraft, it provides a better result when applied to the exact model in the case of the highest noise levels. In the normalized residuals, some artifacts are observed at the corners of the image domain, presumably due to edge effects.

The second test consists in investigating possible improvements of the reconstructions by the use of regularized objective functions for the two models. To this purpose we focus on the penalty function we denoted as hyper-surface regularization, equation (24), with $\delta = 10^{-4}$, and we use the SGP algorithm with the scaling defined by equations (45) and (48). This scaling has already been successfully used in the case of denoising of Poisson data [36], and we use the same parameters of the algorithm described in that paper. For a given β, iteration is stopped when $|f_\beta(x^k; y) - f_\beta(x^{k-1}; y)| \leqslant 10^{-7} f_\beta(x^{k-1}; y)$. The choice of β is performed by computing x^*_β and using a secant-like method for solving equation (30), with a tolerance of 10^{-3}. In general, few steps (from 5 to 10) are required for getting the solution. We always find a solution since conditions of theorem 3 are satisfied with $d = 1$. Next, the value of β providing the minimum rms error is obtained by searching in an interval around the value provided by the discrepancy equation. Also, in this experiment we considered 25 different realization of noise for each test image.

In the case of the spacecraft, the reconstruction errors and the number of required iterations are reported in table 3. The average reconstruction errors are smaller than those obtained in the non-regularized case, with comparable standard deviations. We observe again a strong variation in the number of iterations for both models. When β is selected with the criterion of minimum rms error, the two models provide essentially the same results at the considered noise levels. As concerns the use of the discrepancy criterion, it provides acceptable results except at the highest noise level and, in general, it seems to be more accurate in the case of the exact model. The reconstructions and the normalized residuals are shown in figure 4. The residuals are still affected by strong artifacts, at least in the case of the lowest noise levels.

The reconstruction errors and the number of required iterations for the reconstruction of the nebula NGC 7027 are reported in table 4. By comparing with table 2, we do not find significant improvements with respect to the non-regularized reconstructions, even if the computational cost is now significantly higher. As concerns the use of the discrepancy principle, the remarks done in the case of the spacecraft apply also to the nebula. The reconstructions and the normalized residuals are shown in figure 5. The interesting point is that the residuals are free of artifacts and, therefore, from this point of view, the regularized reconstructions look more reliable than the non-regularized ones.

A few comments concerning the value of the regularization parameter are as follows. In all experiments the discrepancy principle of Bardsley and Goldes provides average values of the regularization parameter greater than those corresponding to the minimum rms error even if, in general, of the same order of magnitude. However, the standard deviations are very large both in the case of the minimum rms error and in the case of the discrepancy principle. In the case of the spacecraft, the standard deviation is of the order of 30% for the exact model and 50% for the approximate one for the values of β determined with the minimum rms error; in the case of discrepancy, the standard deviation is of the order of 70% for the exact model and 50% for the approximate one. Similar results hold true in the case of the nebula. These results compared with those obtained for the reconstruction error indicate that, even if the value of β is strongly dependent on the noise realization, the reconstruction error is much more stable, independent of the criterion used for selecting the regularization parameter.

Table 3. Regularized reconstructions of the spacecraft: errors and iterations.

	Minimum error				Discrepancy			
	Exact		Approx		Exact		Approx	
	Iteration	Error (%)	Iteration	Error (%)	Iteration	Error (%)	Iteration	Error (%)
255	247 ± 54	36.4 ± 0.6	395 ± 148	37.1 ± 0.2	367 ± 198	40.7 ± 4.5	614 ± 213	43.8 ± 1.2
2550	458 ± 138	30.7 ± 0.3	441 ± 118	30.4 ± 0.3	462 ± 210	32.2 ± 1.0	593 ± 200	32.6 ± 1.5
25 500	1308 ± 124	26.1 ± 0.2	1191 ± 242	26.6 ± 1.4	933 ± 221	26.9 ± 0.7	730 ± 55	29.6 ± 3.2
255 000	2190 ± 409	24.3 ± 0.8	2190 ± 409	24.3 ± 0.8	1700 ± 462	24.9 ± 1.0	1655 ± 145	25.4 ± 0.8

Table 4. Regularized reconstructions of the nebula NGC 7027: errors and iterations.

	Minimum error				Discrepancy			
	Exact		Approx		Exact		Approx	
	Iteration	Error (%)	Iteration	Error (%)	Iteration	Error (%)	Iteration	Error (%)
255	352 ± 153	13.5 ± 0.9	639 ± 139	15.3 ± 0.6	707 ± 356	16.6 ± 1.9	505 ± 241	22.4 ± 1.3
2550	409 ± 161	9.7 ± 0.2	340 ± 179	9.9 ± 0.2	673 ± 301	11.5 ± 1.4	692 ± 184	11.9 ± 1.0
25 500	530 ± 313	7.2 ± 0.1	466 ± 235	7.1 ± 0.1	1297 ± 450	8.2 ± 0.5	775 ± 195	8.3 ± 0.5
255 000	603 ± 150	5.4 ± 0.1	571 ± 208	5.5 ± 0.1	1022 ± 534	5.8 ± 0.2	1132 ± 451	6.0 ± 0.2

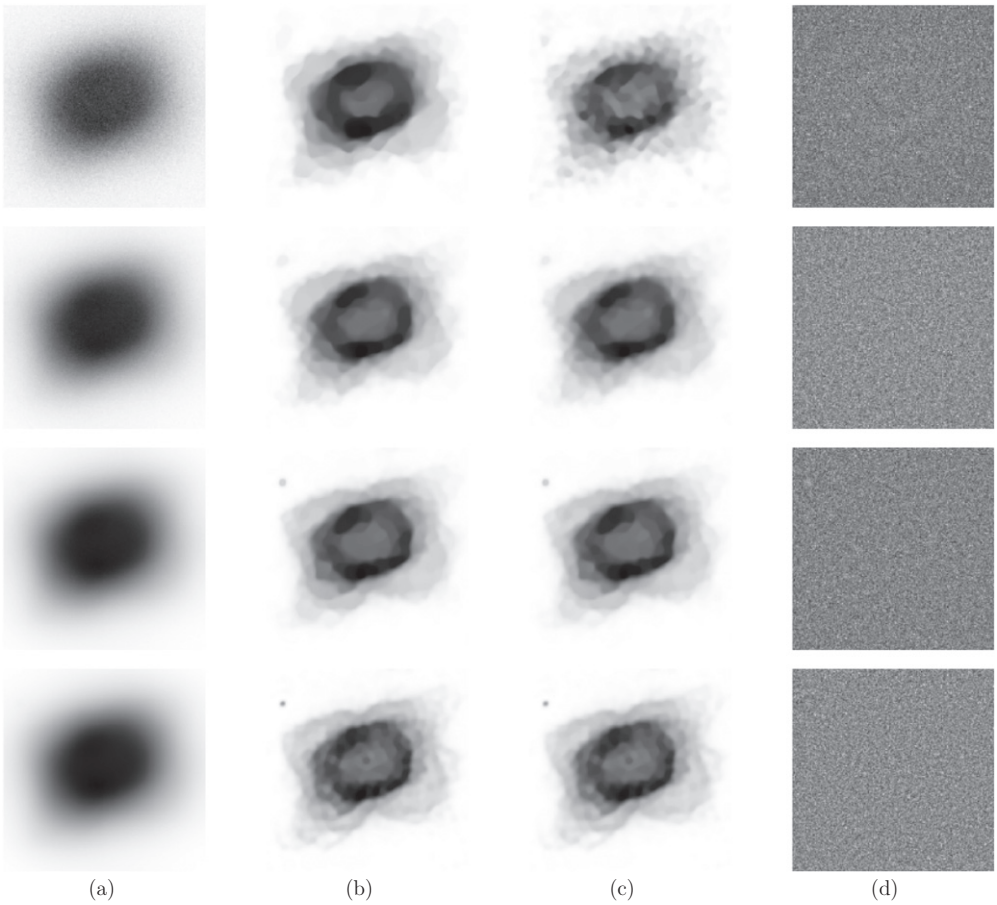


Figure 5. Regularized reconstructions of the nebula NGC 7027: (a) the blurred images, (b) the minimum rms reconstructions with the exact model, (c) those with the approximate model and (d) the normalized residuals in the case of the approximate model.

7. Concluding remarks

The approximate model discussed in this paper provides a sound theoretical basis of the discrepancy principle introduced by Bardsley and Goldes in an empirical way. Moreover, preliminary numerical experiments indicate that, at least for a moderate or large number of photons, the approximate model is equivalent to the exact one.

Both in the theoretical analysis and the numerics we assume that the imaging matrix is exactly known. Therefore an interesting problem, outside the scope of this paper, is to investigate if one of the two models is more robust than the other with respect to errors in the imaging matrix. This problem is obviously important in practical applications where the imaging matrix is never exactly known.

A few comments about the values of the regularization parameter β provided by the discrepancy principle of Bardsley and Goldes are as follows. For the highest noise levels considered in our experiments, it works better for the exact rather than for the approximate model even if the number of numerical experiments is too small for drawing some general

conclusion. More precisely, in the case of the exact model we compute the function $\tilde{D}_y(\beta) = 2f_0(x^*_{P,\beta}; y)/M$, where $x^*_{P,\beta}$ is the minimizer of (2). We do not have a proof of the monotonicity of this function but in our numerical experiments we find that this property holds true and that this function lies above $D_y(\beta)$. Therefore, it crosses 1 for a value of β smaller than that corresponding to the crossing of $D_y(\beta)$.

In addition, we recall that a different discrepancy principle is proposed in [8] for the exact model, based on the discrepancy function

$$D_y^{(P)}(\beta) = \frac{2}{M} f_0^{(P)}(x^*_{P,\beta}; y), \tag{52}$$

which is an increasing function of β. Thanks to the approximate model introduced in this paper, a fourth discrepancy function can be introduced which is defined by $\tilde{D}_y^{(P)}(\beta) = 2f_0^{(P)}(x^*_\beta; y)/M$. We do not have a proof of the monotonicity of this function, but numerical experiments indicate that it is also monotone and therefore can be used for obtaining other values of β.

In summary, we have two models and a discrepancy function for each one, leading to a discrepancy principle for the corresponding model. However, the discrepancy function of one model can be applied to the other and vice versa, so that, for each model we have two criteria for selecting the regularization parameter. We believe that the investigation of the applicability of these models and criteria to the reconstruction of real data is an interesting point for practical applications.

Appendix A

In this appendix, we give results concerning the existence of the solution of the discrepancy equation in the case of Tikhonov regularizer (22). We first remark that the case of denoising is not interesting. Indeed, the solution of the corresponding optimization problem can be obtained by solving second-order algebraic equations, one for each component of the image, and does not provide reliable results. Therefore, we focus on the case of deblurring.

Equation (31) has a unique solution if d satisfies conditions (37). Since the following limits are obvious:

$$\lim_{\beta\to 0} = x^*, \qquad \lim_{\beta\to\infty} = 0, \tag{A.1}$$

we have

$$D_y(0) = \frac{1}{M}\left\| \frac{Hx^* + b - y}{\sqrt{Hx^* + b}} \right\|^2, \qquad D_y(\infty) = \frac{1}{M}\sum_{i\in S} \frac{(y_i - b_i)^2}{b_i}, \tag{A.2}$$

and these quantities can be easily computed in practice since x^* can be obtained by pushing the SGP algorithm to convergence.

Appendix B

In this appendix, we prove the results concerning the existence of the solution of the discrepancy equation for the three regularizers given in terms of the first differences between adjacent values of the unknown object.

Proof of lemma 6. The limit of x^*_β for $\beta \to 0$ is obvious and it is also obvious that the limit for $\beta \to \infty$ is a constant array/cube. The problem is to compute the value of the constant. In

the case of *denoising* from the KKT condition on the gradient, thanks to the identity (27), we have

$$0 \leqslant \sum_{i\in S}[\nabla_x f_\beta(x^*_\beta; y)]_i = \sum_{i\in S}[\nabla_x f_0(x^*_\beta; y)]_i = \frac{M}{2} - \frac{1}{2}\sum_{i\in S}\left(\frac{y_i}{[x^*_\beta]_i}\right)^2, \tag{B.1}$$

and therefore, in the limit $x^*_\beta \to c_\infty$:

$$0 \leqslant \frac{M}{2} - \frac{1}{2}\sum_{i\in S}\left(\frac{y_i}{c_\infty}\right)^2, \tag{B.2}$$

which implies $c_\infty \geqslant y_{\rm rms} > 0$. Therefore, for β being sufficiently large, x^*_β is interior to the non-negative orthant and equality holds true in the previous equations. It follows that equation (34) is proved. □

In the case of *deblurring,* we can proceed in a similar way. Therefore, from the KKT condition on the gradient and identity (27), after some manipulation we obtain

$$0 \leqslant \sum_{j\in R}[\nabla_x f_0(c_\infty e_N; y)]_j = \frac{1}{2}\sum_{i\in S}\tilde{h}_i\left\{1 - \left(\frac{y_i}{\tilde{h}_i c_\infty + b_i}\right)^2\right\}. \tag{B.3}$$

The inequality is not satisfied by $c_\infty = 0$ if the condition of lemma 2 holds true, while it is satisfied by c_∞ sufficiently large. Therefore, c_∞ must be positive and equality holds true, thus proving equation (35). Equation (36) is a simple consequence of the previous one.

Proof of theorem 3. The inequalities (37) are obvious. Equation (30) has a unique solution if and only if they are satisfied by $d = 1$. Equations (38)–(40) can be obtained by simple computations from the results of lemma 6. □

Acknowledgments

This work was partially supported by Italian MUR (Italian Ministry for University and Research), PRIN2008, grant 2008T5KA4L: *Optimization Methods and Software for Inverse Problems*.

References

[1] Bardsley J M and Nagy J G 2006 Covariance-preconditioned iterative methods for nonnegatively constrained astronomical imaging *SIAM J. Matrix Anal. Appl.* **27** 1184–98
[2] Bardsley J M 2008 An efficient computational method for total-variation penalized Poisson likelihood estimation *Inverse Problems Imaging* **2** 167–85
[3] Bardsley J M and Luttman A 2009 Total variation-penalized Poisson likelihood estimation for ill-posed problems *Adv. Comput. Math.* **31** 35–59
[4] Bardsley J M and Goldes J 2009 Regularization parameter selection methods for ill-posed Poisson maximum likelihood estimation *Inverse Problems* **25** 095005
[5] Barrett H H and Meyers K J 2003 *Foundations of Image Science* (New York: Wiley) pp 1047–8
[6] Barzilai J and Borwein J M 1988 Two-point step size gradient methods *IMA J. Numer. Anal.* **8** 141–8
[7] Bertero M, Boccacci P, Desiderá G and Vicidomini G 2009 Image deblurring with Poisson data: from cells to galaxies *Inverse Problems* **25** 123006
[8] Bertero M, Boccacci P, Talenti G, Zanella R and Zanni L 2010 A discrepancy principle for Poisson data *Inverse Problems* **26** 105004
[9] Bonettini S, Zanella R and Zanni L 2009 A scaled gradient projection method for constrained image deblurring *Inverse Problems* **25** 015002
[10] Bonettini S and Ruggiero V 2010 On the uniqueness of the solution of image reconstruction problems with Poisson data *AIP Conf. Proc.* **1281** 1803–6

[11] Brune C, Sawatzky A and Burger M 2010 Primal and dual Bregman methods with application to optical nanoscopy *J. Comput. Vis.* **92** 211–29
[12] Carlavan M and Blanc-Féraud L 2010 Complex wavelet regularization for 3D confocal microscopy deconvolution *INRIA Research Report* 7366
[13] Carlavan M and Blanc-Féraud L 2011 Regularizing parameter estimation for Poisson noisy image restoration *Proc. 1st Int. ICST Workshop on New Computational Methods for Inverse Problems (Paris, May 2011)* at press
[14] 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
[15] Defrise M, Vanhove C and Liu X 2011 An algorithm for total variation regularization in high-dimensional linear problems *Inverse Problems* **27** 065002
[16] Dey N, Blanc-Féraud L, Zimmer C, Roux P, Kam Z, Olivo-Marin J C and Zerubia J 2006 Richardson–Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution *Microsc. Res. Tech.* **69** 260–6
[17] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)
[18] Formiconi A R, Pupi A and Passeri A 1989 Compensation of spatial system response in SPECT with conjugate gradient reconstruction technique *Phys. Med. Biol.* **34** 69–84
[19] Frassoldati G, Zanghirati G and Zanni L 2008 New adaptive stepsize selections in gradient methods *J. Indust. Manag. Optim.* **4** 299–312
[20] Geman S, Manbeck K M 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, MA: Academic) pp 93–130
[21] Green P J 1990 Bayesian reconstructions from emission tomography data using a modified EM algorithm *IEEE Trans. Med. Imaging* **9** 84–93
[22] Huesman R H, Gullberg G T, Greenberg W L and Budinger T F 1977 *User's Manual: Donner Algorithms for Reconstruction Tomography* (Berkeley, CA: Lawrence Berkeley Laboratory, University of California)
[23] Lantéri H, Roche M and Aime C 2002 Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms *Inverse Problems* **18** 1397–419
[24] 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
[25] Lucchese M and Borghese A 2009 Denoising of digital radiographic images with automatic regularization based on total variation *Lecture Notes Comput. Sci.* **5716** 711–20
[26] Lucy L B 1974 An iterative technique for the rectification of observed distributions *Astron. J.* **79** 745–54
[27] Natterer F and Wübbeling F 2001 *Mathematical Methods in Image Reconstruction* (Philadelphia: SIAM) pp 118–23
[28] Resmerita E, Engl H W and Iusem N 2007 The expectation-maximization algorithm for ill-posed integral equations: a convergence analysis *Inverse Problems* **23** 2575–88
[29] Richardson W H 1972 Bayesian based iterative method of image restoration *J. Opt. Soc. Am.* **62** 55–9
[30] Setzer S, Steidl G and Teuber T 2010 Deblurring Poissonian images by split Bregman techniques *J. Vis. Commun. Image Represent.* **21** 193–9
[31] Shepp L A and Vardi Y 1982 Maximum likelihood reconstruction for emission tomography *IEEE Trans. Med. Imaging* **MI-1** 113–22
[32] Snyder D L, Hammoud A M and White R L 1993 Image recovery from data acquired with a charge-coupled-device camera *J. Opt. Soc. Am.* **10** 1014–23
[33] Snyder D L, Helstrom C W, Lanterman A D, Faisal M and White R L 1995 Compensation for readout noise in CCD images *J. Opt. Soc. Am.* A **12** 272–83
[34] Vicidomini G, Boccacci P, Diaspro A and Bertero B 2009 Application of the split-gradient method to 3D image deconvolution in fluorescence microscopy *J. Microsc.* **234** 47–61
[35] Vogel C R 2002 *Computational Methods for Inverse Problems* (Philadelphia: SIAM)
[36] Zanella R, Boccacci P, Zanni L and Bertero M 2009 Efficient gradient projection methods for edge-preserving removal of Poisson noise *Inverse Problems* **25** 045010