

Numerical Methods for Parameter Estimation in Poisson Data Inversion

Luca Zanni · Alessandro Benfenati · Mario Bertero · Valeria Ruggiero

Received: 28 June 2014 / Accepted: 11 December 2014
© Springer Science+Business Media New York 2014

Abstract In a regularized approach to Poisson data inversion, the problem is reduced to the minimization of an objective function which consists of two terms: a data-fidelity function, related to a generalized Kullback–Leibler divergence, and a regularization function expressing a priori information on the unknown image. This second function is multiplied by a parameter β , sometimes called regularization parameter, which must be suitably estimated for obtaining a sensible solution. In order to estimate this parameter, a discrepancy principle has been recently proposed, that implies the minimization of the objective function for several values of β . Since this approach can be computationally expensive, it has also been proposed to replace it with a constrained minimization, the constraint being derived from the discrepancy principle. In this paper we intend to compare the two approaches from the computational point of view. In particular, we propose a secant-based method for solving the discrepancy equation arising in the first approach; when this root-finding algorithm can be combined with an efficient solver of the inner

minimization problems, the first approach can be competitive and sometimes faster than the second one.

Keywords Poisson noise · Regularization parameter estimation · Discrepancy principle · Hypersurface regularization · Seminorm regularization

1 Introduction

In several applications of imaging such as emission tomography, microscopy, and astronomy, the image is formed by counting the number of photons which come from the sources and hit the image domain. The fluctuations of the number of counts in a given pixel or voxel are described by a Poisson distribution which, after the seminal paper of Shepp and Vardi [42], is considered the main source of noise and is taken as the starting point of a maximum likelihood (ML) approach to image restoration.

Since the ML problem is ill-posed, a regularized approach, based on the analogy between image modeling and statistical mechanics, is proposed by Geman and Geman [29]. The assignment of an energy distribution to an image leads to a prior for the pixel values, which is given by Gibbs distribution. This distribution combined with the likelihood derived from the image model provides a posterior distribution for those values. By taking the negative logarithm of this posterior distribution, a penalized ML estimator can be obtained, that leads to the minimization of the following function [6]

$$f_{\beta}(x; y) = f_0(x; y) + \beta f_1(x), \quad \beta \geq 0, \quad (1)$$

where

L. Zanni (✉)
Dipartimento di Scienze Fisiche, Informatiche e Matematiche,
Università di Modena e Reggio Emilia, 41125 Modena, Italy
e-mail: luca.zanni@unimore.it

A. Benfenati · V. Ruggiero
Dipartimento di Matematica e Informatica,
Università di Ferrara, 44122 Ferrara, Italy
e-mail: alessandro.benfenati@unife.it

V. Ruggiero
e-mail: valeria.ruggiero@unife.it

M. Bertero
Dipartimento di Informatica, Bioingegneria,
Robotica e Ingegneria dei Sistemi, Università di Genova,
16146 Genova, Italy
e-mail: bertero@disi.unige.it

– the first term is given by

$$f_0(x; y) = D(y; Hx + b), \tag{2}$$

$D(y; z)$ being the generalized Kullback–Leibler (KL) divergence defined by

$$D(y; z) = \sum_{i=1}^m \left\{ y_i \ln \frac{y_i}{z_i} + z_i - y_i \right\}, \tag{3}$$

where $z_i > 0$ and $y_i \ln \frac{y_i}{z_i} = 0$, when $y_i = 0$; this function coincides, except for an additive constant, with the negative logarithm of the likelihood in the case of Poisson data [42]; it is nonnegative, convex, and coercive on the domain [7];

- $y \in \mathbf{R}^m$ is the detected image while $x \in \mathbf{R}^n$ is the source image to be estimated;
- $H \in \mathbf{R}^{m \times n}$ is the imaging matrix satisfying the following conditions

$$H_{i,j} \geq 0, \quad \sum_{i=1}^m H_{i,j} > 0, \tag{4}$$

the last condition being frequently replaced by the more restrictive one $H^T \mathbf{1}_m = \mathbf{1}_n$, $\mathbf{1}_m \in \mathbf{R}^m$, $\mathbf{1}_n \in \mathbf{R}^n$ being the vectors with all entries equal to 1;

- b is a nonnegative background affecting the detected image; in any case we assume $Hx + b > 0$;
- $f_1(x)$ is the potential function attributed to the image and it is assumed to be nonnegative and convex; it may be non-differentiable;
- β is a regularization parameter controlling the relative weight of the functions $f_0(x; y)$ and $f_1(x)$.

In general y, b, H are the data of the problem, while β is not known and, in principle, should be estimated in a suitable way for obtaining a sensible estimate of the unknown source image x^* . Since x^* is nonnegative, the problem is the constrained minimization of $f_\beta(x; y)$ on the nonnegative orthant.

The structure of the function (1) is similar to that of the standard Tikhonov-regularized functionals: the function $f_0(x; y)$ plays the role of a misfit or data-fidelity function analogous to the least-square function in the Tikhonov approach, while the potential $f_1(x)$ is analogous to the regularization function. For this reason β is often called *regularization parameter* even if some concepts of Tikhonov regularization theory, such as the regularization algorithm definition, do not apply to the case of Poisson data. It is sufficient to remark that the assumption of “noise tending to zero”, typically used in Tikhonov theory, here must be replaced by the statement “number of counts tending to infinity”.

Thanks to the previous assumptions the function $f_\beta(x; y)$ is also nonnegative, convex, and coercive on the nonnegative orthant; therefore, it has global minimizers for each value of $\beta \geq 0$. Since the role of this parameter is crucial, the selection of its value is an important issue in the application of the previous approach to Poisson data reconstruction.

A plethora of methods for the selection of the regularization parameter has been proposed in the case of Tikhonov theory, the most popular and also theoretically sound being the so-called *Morozov discrepancy principle* [24]. In the case of Poisson data only a few have been proposed so far [2, 7]. In this paper we focus on the second one [7] because it is more similar to Morozov discrepancy principle. It can be formulated as follows.

As shown in [46, 47], if x^* denotes the true source image corresponding to the detected data y and the values of x^* are large, then the expected value of $f_0(x^*; y)$ is approximately given by $m/2$, m being the number of pixels or voxels in the image domain. Therefore, if we introduce the normalized discrepancy function

$$\mathcal{D}_H(x; y) = \frac{2}{m} D(y; Hx + b), \tag{5}$$

and we denote as x_β a nonnegative minimizer of (1) it is quite natural to propose the following model for the selection of β , equivalent to require that the discrepancy corresponding to the selected minimizer is close to that of the true source image.

Model 1 Select the value of β such that

$$\mathcal{D}_H(x_\beta; y) = \eta, \tag{6}$$

where $x_\beta \geq 0$ is the minimizer of (1) and η is a given number close to 1. If β is a solution of the Eq. (6), then $\bar{x} = x_{\bar{\beta}}$ is called solution of Model 1.

The model is meaningful because, for several potential functions $f_1(x)$ used in practice, the minimizer x_β is unique and, as proved in [7], $\mathcal{D}_H(x_\beta; y)$ is an increasing function of β , thus ensuring the uniqueness of the solution of the discrepancy equation (6). We will also denote Model 1 as the *Crossing Model*.

It has been remarked by several authors [17, 18, 44] that the approach of Model 1 can lead to highly time-consuming computations since it can require the solution of several minimization problems; therefore, a *Constrained Model* has been proposed and formulated as follows.

Model 2 For a given η , solve the problem

$$\min_{x \geq 0} f_1(x), \quad \text{subject to } \mathcal{D}_H(x; y) \leq \eta. \tag{7}$$

Any solution \bar{x} of this problem is called solution of Model 2. It is obvious that in such a case only one minimization is required.

The main purpose of this paper is to investigate the two models, both in the case of a differentiable and in the case of a non-differentiable potential function $f_1(x)$. Under suitable conditions discussed in Sect. 2, the models provide the same parameter estimation; here we compare through a wide computational study the efficiency of the numerical procedures involved in the two approaches. For each model we use state of the art computational techniques.

In particular our main contribution is the introduction of an effective secant-type solver for the discrepancy equation (6). This solver, named Modified Dai–Fletcher (MDF) method, is able to reduce the number of function evaluations and, consequently, the number of required solutions of the non-negative minimization problems. This leads to efficient solution of the discrepancy equation when very effective inner solvers are available for the nonnegative minimization problems.

We consider the case of a differentiable regularization (HS) and the non-differentiable one (TV). In the first case we show that MDF, combined with an efficient solver for the inner optimization problems, enables Model 1 to be less time demanding than Model 2. On the other hand, in the case of TV regularization, Model 1 can be competitive with Model 2 but the latter is definitely more efficient in some cases.

The paper is organized as follows. In Sect. 2, we recall the results proved in [44] concerning the relations between the two models in the case $y > 0$ (Sect. 2.1). Moreover it is shown that, in the case of a differentiable potential function $f_1(x)$, it is possible to generalize the results to the case $y \geq 0$, which is relevant in medical imaging and microscopy; proofs are given in Appendix 1. In Sect. 2.2, we recall the results concerning existence and uniqueness of the discrepancy equation (6) and we show that these results can be extended to seminorm regularization. In Sect. 3, we discuss the solvers for the two models and, in particular, the proposed root-finding method for solving the discrepancy equation (6) is presented. The setting of the main parameters involved by the solvers is detailed in Appendix 2. In Sect. 4 the numerical experiments are described. These are based on two different source images, both well known: the cameraman and the spacecraft. In both cases three noise levels are considered by changing the total number of counts. Finally, the main conclusions on the numerical efficiency of the two models are drawn in Sect. 5.

Throughout the paper we denote by $\text{lev}_\alpha f(t)$ the level set of a function $f(t)$, i.e., $\{t; f(t) \leq \alpha\}$, and by i_C the indicator function of a closed and convex set C . Furthermore, given two vectors x and y , $y > 0$, the operator $\frac{x}{y}$ is to be intended component-wise.

2 Relations Between the Two Models

In this section we show, using and extending results proved in [44], that the two models have exactly the same solutions for a suitable class of potential functions. Some of the proofs required for the extension of the results are given in Appendix 1. Moreover we report and extend results, proved in [7], concerning Model 1.

For a unified treatment of the cases we are investigating if it is convenient to set

$$f_1(x) = R(Lx), \tag{8}$$

where $R(t)$ is a convex and nonnegative function and L is a suitable matrix. We recall that any seminorm on \mathbf{R}^n can be written as $\|L \cdot\|$, and therefore it can be referred by the notation (8).

2.1 Relations Between Penalized and Constrained Convex Problems

In order to have notations close to those used in [44], we set $\tau = \eta \frac{m}{2}$ and $\lambda = \frac{1}{\beta}$. Moreover we introduce the quantities

$$\tau_0 = \min_{x \geq 0} D(y; Hx + b), \tag{9}$$

$$\tau_L = \min_{x \geq 0, x \in \mathcal{N}(L)} D(y; Hx + b),$$

where $\mathcal{N}(L)$ is the null space of the matrix L in Eq. (8).

We consider the following problems, strictly connected to the Models (6) and (7):

$$\min_{x \geq 0} R(Lx) + \lambda D(y; Hx + b), \quad \lambda \geq 0, \tag{10}$$

and

$$\min_{x \geq 0} R(Lx), \quad \text{subject to } D(y; Hx + b) \leq \tau, \tag{11}$$

with $\tau \geq \tau_0$. Moreover in this work we focus on the following regularization functions:

- Tikhonov-like:

$$R(Lx) = \frac{1}{2} \|Lx\|_2^2, \tag{12}$$

where L is the identity or a first/second-order finite difference operator;

- Hypersurface potential (HS) [35]:

$$R(Lx) = \sum_{i=1}^n \left\| \begin{pmatrix} L_i x \\ \delta \end{pmatrix} \right\|_2 - n\delta, \tag{13}$$

where δ is a positive parameter, L_i is the $d \times n$ discrete gradient operator for a d -dimensional image ($d = 1, 2, 3$) and $L = (L_1^T, \dots, L_n^T)^T$, $L \in \mathbf{R}^{p \times n}$, $p = nd$;
 – discrete Total Variation (TV):

$$R(Lx) = \sum_{i=1}^n \|L_i x\|_2. \tag{14}$$

When $R(Lx)$ is (12) or (13), we have a differentiable function while (14) is a non-differentiable version of (13). In all the three cases, R satisfies a set of general assumptions, applicable to other regularization functions such as, for example, $\|Lx\|_1$ or Markov random-field regularizations [29]. We denote as **Assumption REG** the following assumptions, also satisfied by any seminorm:

- $R : \mathbf{R}^p \rightarrow \mathbf{R}$ is a proper, convex, continuous function, such that $R(t) \geq 0$ and $R(t) = 0 \Leftrightarrow t = 0$;
- the level sets $\text{lev}_\alpha R(t)$ of $R(t)$ are nonempty and bounded for $\alpha \geq 0$;
- $L \in \mathbf{R}^{p \times n}$, $p \geq n$.

Under **Assumption REG**, it is evident that $\mathcal{N}(L)$ is the set of minimum points of $R(Lx)$.

Theoretical results about the relations between the convex problems (10) and (11) can be found in [33]; we point out that, when the minimizer of a general penalized problem $\lambda F(x) + G(x)$ is not unique, $F(x)$ can assume different values at the minimum points; furthermore, it may happen that the multiplier of the constrained problem $\min_{x \in \text{lev}_\tau(F(x))} G(x)$ is not unique [21].

In [44], the authors specialize these general results to the problems (10) and (11), devising the conditions that assure their equivalence when $R(Lx)$ is a seminorm and $y > 0$. The discussion can be easily extended to a differentiable $R(Lx)$ satisfying the **Assumption REG** (as, for example, Tikhonov-like regularization or Hypersurface potential). For sake of completeness, we report these results, requiring only **Assumption REG** for $R(Lx)$.

Theorem 1 [44, Theorem 3.2] *Let $H \in \mathbf{R}^{m \times n}$ be such that*

$$\mathcal{K} = \{x \geq 0 : Hx + b > 0\} \neq \emptyset \tag{15}$$

and

$$\{x \geq 0\} \not\subseteq \mathcal{N}(H). \tag{16}$$

*We assume $L \in \mathbf{R}^{p \times n}$ with $\mathcal{N}(L) \cap \mathcal{N}(H) = \{0\}$ and $R(Lx)$ satisfying the **Assumption REG**. Then we have*

- (i) *the problems (10) with $\lambda \geq 0$ and (11) with $\tau \geq \tau_0$ have a solution;*

- (ii) *under the assumption $y > 0$, if \hat{x} and \tilde{x} are solutions of (10) for a given $\lambda > 0$, then $R(L\hat{x}) = R(L\tilde{x})$ and $H\hat{x} = H\tilde{x}$, i.e., $D(y; H\hat{x} + b) = D(y; H\tilde{x} + b)$;*
- (iii) *let $\text{argmin}_{x \geq 0} D(y; Hx + b) \cap \mathcal{N}(L) = \emptyset$ and $\tau_0 < \tau < \tau_L$; under the assumption $y > 0$, if \hat{x} and \tilde{x} are solutions of (11) for a given τ , then $R(L\hat{x}) = R(L\tilde{x})$ and $H\hat{x} = H\tilde{x}$, with $D(y; H\hat{x} + b) = D(y; H\tilde{x} + b) = \tau$.*

In presence of nonzero background, the assumption (16) allows to exclude the trivial case when the nonnegative orthant is a subset of $\mathcal{N}(H)$. Obviously, when H satisfies the conditions (4), the assumptions (15) and (16) hold.

Using part (iii) of Theorem 1 and a relationship between the value of λ and the corresponding solutions of problem (10) (the extension of this relationship to the case of differentiable regularization term satisfying **Assumption REG** is given in Appendix 1), the following theorem states the relationship between problems (10) and (11).

Theorem 2 [44, Theorem 3.4] *Let $H \in \mathbf{R}^{m \times n}$ be such that (15)–(16) hold and $L \in \mathbf{R}^{p \times n}$ with $\mathcal{N}(L) \cap \mathcal{N}(H) = \{0\}$. Let $\text{argmin}_{x \geq 0} D(y; Hx + b) \cap \mathcal{N}(L) = \emptyset$ and $\tau_0 < \tau < \tau_L$. When $R(Lx)$ is a seminorm regularization or a differentiable function satisfying the **Assumption REG**, if \tilde{x} is a solution of (11), then there exists a unique $\bar{\lambda} > 0$ such that \tilde{x} is a solution of (10).*

In conclusion Theorem 2, combined with part (ii) and part (iii) of Theorem 1, states that Model 2 has solutions and that they are all the solutions of Model 1.

For suitable differentiable regularizations, the previous results can be established under assumptions weaker than those of the previous Theorem 2: more precisely one can remove the assumption that all the values of y are strictly positive. Before introducing these results, we recall the conditions for the existence and uniqueness of the solution of the problem (10) for $\lambda > 0$ in some special cases and we extend part (iii) of Theorem 1. The proofs are reported in Appendix 1.

Let $I_1 = \{i \mid y_i > 0\}$ and $I_2 = \{1, 2, \dots, m\} - I_1$. The cardinality of I_1 is denoted by m_1 . We denote by y_{I_1} the vector of nonzero entries of y and by H_{I_1} and H_{I_2} the submatrices of H given by the rows with indices in I_1 and I_2 , respectively. The following Lemma is a generalization of a result proved in [14].

Lemma 1 *Let $H \in \mathbf{R}^{m \times n}$ be such that the assumptions (15)–(16) hold and $L \in \mathbf{R}^{p \times n}$. Let $\mathcal{N}(L) \cap \mathcal{N}(H_{I_1}) = \{0\}$, $I_1 \neq \emptyset$. If $R(Lx)$ satisfies the **Assumption REG** and $R(t)$ is a differentiable function with positive-definite Hessian, then the problem (10) for $\lambda > 0$ has a unique solution.*

Lemma 1 holds for Tikhonov-like regularization, HS potential and Markov random-field functions.

Under the assumptions of Lemma 1, the following relation between λ and the unique solution \hat{x} of the problem (10) holds:

$$\lambda = \frac{\langle \nabla_t R(t)|_{t=L\hat{x}}, L\hat{x} \rangle}{\left\langle \frac{yI_1}{H_1\hat{x}+b}, H_1\hat{x} \right\rangle - \langle 1_m, H\hat{x} \rangle}, \tag{17}$$

where the quotient of the vectors is intended component-wise. Furthermore, part (iii) of Theorem 1 can be restated under more general assumptions.

Theorem 3 *Let $H \in \mathbf{R}^{m \times n}$ be such that the assumptions (15)–(16) hold, $L \in \mathbf{R}^{p \times n}$ and let $R(Lx)$ satisfy **Assumption REG**. Let $\operatorname{argmin}_{x \geq 0} D(b, Hx + b) \cap \mathcal{N}(L) = \emptyset$, and $\tau_0 < \tau < \tau_L$; under the assumption $\mathcal{N}(L) \cap \mathcal{N}(H_{I_1}) = \{0\}$, $I_1 \neq \emptyset$, if \hat{x} and \tilde{x} are solution of (11), then $R(L\hat{x}) = R(L\tilde{x})$ and $D(y; H\hat{x} + b) = D(y; H\tilde{x} + b) = \tau$, with $H_{I_1}\hat{x} = H_{I_1}\tilde{x}$ and $\langle 1_{m-m_1}, H_{I_2}\tilde{x} \rangle = \langle 1_{m-m_1}, H_{I_2}\hat{x} \rangle$.*

Finally, following the same arguments used in [44] for proving Theorem 2, Theorem 3 enables to extend the results of Theorem 2 to the case $y \geq 0$ for differentiable regularization functions satisfying Lemma 1.

Theorem 4 *Let $H \in \mathbf{R}^{m \times n}$ be such that the assumptions (15)–(16) hold and $L \in \mathbf{R}^{p \times n}$. Let $\mathcal{N}(L) \cap \mathcal{N}(H_{I_1}) = \{0\}$, $I_1 \neq \emptyset$. Let $\operatorname{argmin}_{x \geq 0} D(b, Hx + b) \cap \mathcal{N}(L) = \emptyset$, and $\tau_0 < \tau < \tau_L$. Assume that $R(Lx)$ satisfies **Assumption REG** and $R(t)$ is a differentiable function with positive-definite Hessian. If \tilde{x} is a solution of (11), then there exists a unique $\bar{\lambda} > 0$ such that \bar{x} is a solution of (10).*

2.2 Existence and Uniqueness of the Solution of the Discrepancy Equation

In [7], conditions for the existence and uniqueness of the solution of the discrepancy equation (6) are devised for denoising and deblurring of images corrupted by Poisson noise, when $f_1(x) \equiv R(Lx)$ is the Tikhonov-like regularization (12) or the HS potential.

Using the results of Sect. 2.1, the discrepancy equation can be used also when $R(Lx)$ is a seminorm regularization. We recall that x_β denotes a solution of the penalized problem

$$\min_{x \geq 0} f_\beta(x; y) \equiv D(y; Hx + b) + \beta R(Lx). \tag{18}$$

The following theorem states the conditions assuring the well-definiteness of the function $\mathcal{D}_H(x_\beta; y)$ for $y > 0$.

Theorem 5 *Let $H \in \mathbf{R}^{m \times n}$ be such that (15)–(16) hold and $L \in \mathbf{R}^{p \times n}$ with $\mathcal{N}(L) \cap \mathcal{N}(H) = \{0\}$. When $y > 0$ and $R(Lx)$ satisfies the **Assumption REG**, $\mathcal{D}_H(x_\beta; y)$ is well defined for $\beta > 0$.*

Proof Given $\beta > 0$, Theorem 1 assures that the problem (10) with $\lambda = \frac{1}{\beta}$ has at least a solution and, if we consider two different solutions of the problem \hat{x} and \tilde{x} , we have $D(y; H\hat{x} + b) = D(y; H\tilde{x} + b)$. Therefore for all $\beta > 0$, $\mathcal{D}_H(x_\beta; y)$ is well defined. \square

Finally, Theorem 6 establishes the uniqueness of the solution of Model 1 for seminorm regularization (i) and restates the results already obtained in [7] (ii).

Theorem 6 *Let $H \in \mathbf{R}^{m \times n}$ be such that (15)–(16) hold and $L \in \mathbf{R}^{p \times n}$ with $\mathcal{N}(L) \cap \mathcal{N}(H) = \{0\}$.*

Assume $\tau = \eta \frac{m}{2}$ such that $\tau_0 < \tau < \tau_L$ (i.e. $\operatorname{argmin}_{x \geq 0} D(y; Hx + b) \cap \mathcal{N}(L) = \emptyset$). Thus we have:

- (i) *under the assumption $y > 0$, the solution $\bar{\beta}$ of the discrepancy equation combined with a seminorm regularization exists and is unique;*
- (ii) *let $R(Lx)$ be a differentiable regularization term satisfying **Assumption REG**, such that the Hessian of $R(t)$ is positive definite; under the assumption $\mathcal{N}(L) \cap \mathcal{N}(H_{I_1}) = \{0\}$, $I_1 \neq \emptyset$, the solution $\bar{\beta}$ of the discrepancy equation combined with the regularization term $R(Lx)$ exists and is unique; furthermore the vector $x_{\bar{\beta}}$ corresponding to the solution is unique.*

Proof (i) We consider the problem (11) with $\tau = \eta \frac{m}{2}$ and $R(Lx)$ given by a seminorm. For part (iii) of Theorem 1, the solution \bar{x} of this problem exists and $D(y; H\bar{x} + b) = \tau$. For Theorem 2, there exists a unique $\bar{\lambda} > 0$ such that \bar{x} is a solution of (10) with $\bar{\lambda} = 1/\bar{\beta}$ and this value does not depend on \bar{x} . Then $\bar{\beta} = 1/\bar{\lambda}$ is the unique solution of the discrepancy equation combined with the seminorm regularization.

(ii) We consider the problem (11) with $\tau = \eta \frac{m}{2}$ and $R(Lx)$ satisfying **Assumption REG**. For Theorem 3, the solution \bar{x} of this problem exists and $D(y; H\bar{x} + b) = \tau$. Since $R(t)$ is a differentiable function with positive-definite Hessian, Theorem 4 assures that there exists a unique $\bar{\lambda} > 0$ such that \bar{x} is a solution of (10) with $\bar{\lambda} = 1/\bar{\beta}$ and this value does not depend on \bar{x} . Then $\bar{\beta} = 1/\bar{\lambda}$ is the unique solution of the discrepancy equation. Furthermore, from Lemma 1, the solution $x_{\bar{\beta}}$ of (10) for a given $\bar{\beta}$ is unique. \square

In [7], under the assumption of part ii) of the previous theorem, for differentiable function with positive-definite Hessian $R(t)$ and under the assumption (4) on H , the authors give the conditions that assure $\tau_0 < \eta \frac{m}{2} < \tau_L$ for HS and Tikhonov-like regularization. For example, for image deconvolution ($n = m$ and $H1_n = 1_n$), the condition $\frac{n}{2} < \tau_L$ is satisfied if

$$\frac{1}{n} \sum_{i \in I_1} y_i \log y_i > \frac{1}{2} + \bar{y} \log \bar{y}, \tag{19}$$

where $\bar{y} = \frac{\sum_{i \in I_1} y_i}{n}$.

3 Numerical Methods

In this section we briefly describe the solvers used in the two models and, in particular, we propose a very efficient root-finding method for the solution of the discrepancy equation.

3.1 The ADMM Solver for the Constrained Model

In [44], the authors propose to estimate the regularization parameter by solving the constrained problem (11) with the well-known alternating direction method of multipliers (ADMM) [28,30]. The problem (11) is equivalent to

$$\begin{aligned} \min_x F(x, z) + \frac{1}{2\gamma} \|Ax - c - z\|_2^2, \\ \text{subject to } Ax = z + c, \end{aligned} \tag{20}$$

where γ is a positive parameter, $F(x, z) = i_{\text{lev}_\tau D(y; z_1)} + i_{z_2 \geq 0} + R(\bar{z})$ and $z = (z_1^T, z_2^T, \bar{z}^T)^T$. The linear constraint depends on $R(Lx)$; for example, for discrete TV we have

$$\begin{pmatrix} H \\ I \\ L \end{pmatrix} x = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} + \begin{pmatrix} -b1_m \\ 0 \\ 0 \end{pmatrix} \tag{21}$$

with $\bar{z} = z_3$, while for HS potential the constraint is

$$\begin{pmatrix} H \\ I \\ L \\ 0 \end{pmatrix} x = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_\delta \end{pmatrix} + \begin{pmatrix} -b1_m \\ 0 \\ 0 \\ -\delta 1_n \end{pmatrix} \tag{22}$$

with $\bar{z} = (z_3^T, z_\delta^T)^T$.

The basic idea of ADMM method is to compute a saddle point of the Augmented Lagrangian $\mathcal{L}(x, z, p)$ of the problem (20), decoupling the minimization step of the Augmented Lagrangian method in a sequence of minimizations with respect to the different sets of variables x and z_k [9]. The application of ADMM for the constrained problem (11) is analyzed in [44] for a seminorm regularization: the authors describe the algorithm in detail and prove the convergence of the sequence of multipliers obtained for the inner subproblem in z_1 to the multiplier λ of $D(y; Hx + b) \leq \tau$. In view of the theorems of the previous section, these results can be easily extended to suitable differentiable regularization functions. We mention that the subproblem concerning the minimization with respect to x requires the solutions of a linear system that can be obtained by fast transforms; the minimization

with respect to z_1 , that is $\min_{z_1 \in \text{lev}_\tau(D(z_1; y))} \frac{1}{2\gamma} \|z_1 - v^{(i)}\|_2^2$ ($v^{(i)}$ depending on the current iteration), requires the solution of a nonlinear equation. This can be obtained by few iterations of the Newton’s method [17].

It is well known that the performance of ADMM can be strongly dependent on the choice of the parameter γ . In order to reduce this drawback, in [32,45] strategies for adaptively adjusting γ at any iteration have been proposed. The convergence of this adaptive version of ADMM is proved, assuming that γ becomes fixed after a finite number of iterations. Then the choice of the updating strategy is based on the numerical effectiveness. The procedure proposed in [32] is based on the control of the l_2 norm of two vectors, the so-called primal and dual residual:

$$\begin{aligned} r^{(i)} &= Ax^{(i)} - z^{(i)} - c, \\ s^{(i)} &= \frac{1}{\gamma} A^T \sigma^{(i)}, \end{aligned} \tag{23}$$

with $\sigma^{(i)} = z^{(i)} - z^{(i-1)}$. Indeed the vectors $r^{(i)}$ and $\sigma^{(i)}$ are crucial components of the upper bound of the absolute error between the objective function at the current iterate and its minimum value. In the experiments of Sect. 4, we use the following updating procedure:

$$\gamma^{(i+1)} = \begin{cases} \frac{\alpha}{\gamma^{(i)}} & \text{if } \|r^{(i)}\| > \mu \|\sigma^{(i)}\| \text{ and } i \leq k_{\max} \\ \alpha \gamma^{(i)} & \text{if } \|\sigma^{(i)}\| > \mu \|r^{(i)}\| \text{ and } i \leq k_{\max} \\ \gamma^{(i)} & \text{otherwise} \end{cases} \tag{24}$$

where α and μ are positive values greater than 1 and $\gamma^{(0)}$ is a prefixed positive value. From the practical point of view, this adaptive version of ADMM appears less dependent on the parameter settings than the standard ADMM approach. Finally, we mention that other methods have been adapted in [44] for the numerical solution of the constrained problem, as Arrow–Hurwitz method [1] or its extrapolated version [20]. As in the case of ADMM, the performance of these algorithms depends on a suitable selection of two prefixed parameters. Then, in the numerical experiments described in Sect. 4, for Model 2 we focus the attention on ADMM.

3.2 The MDF Solver for the Crossing Model

The solution $\bar{\beta}$ of the discrepancy equation (6) can be approximated by solving the root-finding problem

$$\mathcal{F}(\beta) = \mathcal{D}_H(x_\beta; y) - \eta = 0, \tag{25}$$

where $\mathcal{F}(\beta)$ is a continuous increasing function. We solve this problem by a specialized version of the algorithm proposed in [22], called Modified Dai–Fletcher (MDF) method. The root-finding solver [22] consists in two phases: a bracketing phase to determine the extremes $0 < \beta_l < \beta_u$

of an interval containing the root and a secant phase for the root approximation. The implementation of the bracketing phase has been changed according to the special regularization framework we are considering. In particular, we have taken into account that the evaluation of the discrepancy equation in β is generally less expensive for large values of β , due to the special form of the penalized problem providing x_β . For this reason, when we start from β such that $\mathcal{F}(\beta) < 0$, we look for β_l and β_u by means of a secant-like approach, allowing steps in the interval $[d\beta, 10d\beta]$, where $d\beta$ denotes the previous step; on the other hand, when we move from β such that $\mathcal{F}(\beta) > 0$, we reduce the tentative β by a constant factor $\omega \in (0, 1)$, for preserving the positivity of β and for avoiding the evaluation of the discrepancy function in too small values. The bracketing phase is sketched as in Algorithm 1.

Algorithm 1 Bracketing Phase of Algorithm MDF

INPUT: a tentative $\beta > 0$, an initial step $d\beta > 0$, and $\omega \in (0, 1)$.

```

IF  $\mathcal{F}(\beta) < 0$ 
     $\beta_l = \beta, \beta = \beta + d\beta,$ 
    While  $\mathcal{F}(\beta) < 0$ 
         $s = \min(1, \max((\mathcal{F}(\beta_l) - \mathcal{F}(\beta)) / \mathcal{F}(\beta), 0.1)),$ 
         $\beta_l = \beta, d\beta = d\beta/s, \beta = \beta + d\beta,$ 
    end
     $\beta_u = \beta,$ 
ELSE
     $\beta_u = \beta, \beta = \beta_u \omega,$ 
    While  $\mathcal{F}(\beta) > 0$ 
         $\beta_u = \beta, \beta = \beta_u \omega,$ 
    end
     $\beta_l = \beta,$ 
END
OUTPUT:  $\beta_l, \beta_u$  such that  $\bar{\beta} \in [\beta_l, \beta_u]$ .

```

The second phase of the MDF method is essentially the secant-based strategy described in [22]: starting from the values β_l and β_u provided by the bracketing phase, the root $\bar{\beta}$ is approximated by exploiting standard secant steps or modified steps designed to accelerate the convergence remote from the solution. For completeness, we report the main steps of the MDF secant phase in Algorithm 2. The stopping rule used in the MDF secant phase is

$$|\mathcal{F}(\beta_k)| \leq \epsilon_1 \quad \text{or} \quad (26)$$

$$(|\beta_k - \beta_{k-1}| \leq \epsilon_2 \beta_k \quad \text{and} \quad |\mathcal{F}(\beta_k)| \leq 10\epsilon_1),$$

where $\beta_{(k)}$ denotes the value of β at the k -th iteration of Algorithm 2 and ϵ_1, ϵ_2 are two small positive constants.

Algorithm 2 Secant Phase of Algorithm MDF

```

INPUT:  $\beta_l, \beta_u$  such that  $\bar{\beta} \in [\beta_l, \beta_u]$ .
 $s = (\mathcal{F}(\beta_u) - \mathcal{F}(\beta_l)) / \mathcal{F}(\beta_u), d\beta = (\beta_u - \beta_l) / s, \beta = \beta_u - d\beta,$ 
WHILE (stopping rule is not satisfied)
    IF  $\mathcal{F}(\beta) > 0$ 
        IF  $s \leq 2$ 
             $\beta_u = \beta, s = (\mathcal{F}(\beta_u) - \mathcal{F}(\beta_l)) / \mathcal{F}(\beta_u),$ 
             $d\beta = (\beta_u - \beta_l) / s, \beta = \beta_u - d\beta,$ 
        ELSE
             $s = \max((\mathcal{F}(\beta_u) - \mathcal{F}(\beta)) / \mathcal{F}(\beta), 0.1),$ 
             $d\beta = (\beta_u - \beta) / s, \beta_u = \beta,$ 
             $\beta = \max(\beta - d\beta, 0.75\beta_l + 0.25\beta),$ 
             $s = (\beta_u - \beta_l) / (\beta_u - \beta),$ 
        END
    ELSE
        IF  $s \geq 2$ 
             $\beta_l = \beta, s = (\mathcal{F}(\beta_u) - \mathcal{F}(\beta_l)) / \mathcal{F}(\beta_u),$ 
             $d\beta = (\beta_u - \beta_l) / s, \beta = \beta_u - d\beta,$ 
        ELSE
             $s = \max((\mathcal{F}(\beta_l) - \mathcal{F}(\beta)) / \mathcal{F}(\beta), 0.1),$ 
             $d\beta = (\beta - \beta_l) / s, \beta_l = \beta,$ 
             $\beta = \min(\beta + d\beta, 0.75\beta_u + 0.25\beta),$ 
             $s = (\beta_u - \beta_l) / (\beta_u - \beta),$ 
        END
    END
END
OUTPUT:  $\beta$  approximation of  $\bar{\beta}$ 

```

From a practical point of view, the MDF performance strictly depends on the effectiveness of the minimization method used for obtaining the solution x_β of the penalized problem (18). In our experience, two suitable solvers for (18) are the Scaled Gradient Projection (SGP) method [15], for the case of differentiable objective functional, and the PIDSplit+ algorithm [25,41], that can work also in case of non-differentiable problems. The SGP method has been widely used for optimization problems in imaging [4, 10, 12, 19, 38, 48] and has given rise to interesting accelerations of well-known reconstruction algorithms such as the Richardson–Lucy method [36, 39] and the Image Space Reconstruction Algorithm [23]. Its iteration can be described as

$$y^{(l)} = P_+(x^{(l)} - \alpha_l S_l \nabla f(x^{(l)})),$$

$$x^{(l+1)} = x^{(l)} + \theta_l (y^{(l)} - x^{(l)}), \quad \theta_l \in (0, 1], \quad (27)$$

where $P_+(\cdot)$ denotes the projection operator onto the non-negative orthant, θ_l is the line-search parameter, $\alpha_l > 0$ is a steplength parameter and S_l denotes a positive-definite diagonal scaling matrix. Here, a standard monotone line-search strategy is used for defining the parameter θ_l , while special updating rules for α_l and S_l are exploited for achieving improvements in the convergence rate. In order to update

α_l we use the strategy recently proposed in [37], that is derived by adapting to the scaled gradient projection methods the steplength rule introduced in [26] for steepest descent methods in unconstrained optimization. This updating rule aims at capturing second-order information by exploiting the gradients computed in a limited number of previous iterations and, in case of quadratic unconstrained optimization, it reduces to compute special approximations of the Hessian’s eigenvalues, called Ritz values [31]. The computational studies reported in [26] on unconstrained optimization problems emphasize that the new rule enables worthwhile convergence rate improvements in comparison with the well-known Barzilai–Borwein (BB) rule [3]. In particular, in case of the penalized problem (18), we found that the SGP version exploiting this Ritz-like steplength updating strategy generally exhibits better convergence than the original SGP method [37], that was based on an adaptive alternation of the two BB rules [15,27].

A crucial role for the SGP performance is also played by the diagonal scaling matrix S_l [8,11]. When a diagonal scaling is exploited, the challenging task consists in achieving convergence rate improvements without increasing the cost of each iteration, as usually done by quasi-Newton non-diagonal scaling techniques. To this end, we follow the idea introduced in [34], that defines the diagonal entries $s_i, i = 1, \dots, n$, of S_l through a splitting of the objective gradient:

$$s_i = \min \left(s_{\text{supp}}, \max \left(s_{\text{low}}, \frac{x_i^{(l)}}{V_i^D(x^{(l)}; y) + \beta V_i^R(x^{(l)}; y)} \right) \right),$$

where $0 < s_{\text{low}} < s_{\text{supp}}$ are positive constants (see [13] for a convenient choice of these values) and the vectors $V^D(x; y)$ and $V^R(x; y)$ are such that

$$\begin{aligned} \nabla(D(y; Hx + b)) &= V^D(x; y) - U^D(x; y), \\ \nabla(R(Lx)) &= V^R(x; y) - U^R(x; y), \\ V^D(x; y) &> 0, \quad U^D(x; y) \geq 0, \\ V^R(x; y) &> 0, \quad U^R(x; y) \geq 0. \end{aligned} \tag{28}$$

The choice of the functions $V^D(x; y)$ and $V^R(x; y)$ is not unique and in this paper we use the choice proposed in [46]. This strategy avoids significant additional cost and for suited choices of the splitting can yield useful improvement in the SGP convergence rate. Such a behavior has been observed in several imaging problems, where the splittings (28) were derived in a quite natural way by the structure of the gradient [5,15,46]. In this work, we are interested in solving the problem (18) in case of the HS regularization function (13) and, following the suggestions available in literature, we define S_l by assuming $V^D(x; y) = H^T 1_m$ [34] and $V^R(x; y)$ as in [46].

An alternative inner solver, suitable also in case of non-differentiable regularization functions, is the PIDSplit+ algo-

rithm [25,41]; this is a specialized version of the ADMM method, that requires to reformulate the problem (18) in the form (20), with $F(x, z) = D(y; z_1) + i_{z_2 \geq 0} + \beta R(\bar{z})$ and a linear constraint $Az = z + c$, depending on the regularization term (for example (21) or (22) for TV or HS, respectively). Also in this case, the minimization with respect to x involves the solution of a linear system, obtained by fast transforms, while the minimization with respect to the other sets of variables z_k and the update of the related multiplier vectors are less expensive. These last tasks are performed by closed formulas requiring at most matrix-vector products or a projection on the nonnegative orthant. Moreover, as already highlighted for ADMM, the effectiveness of the method depends on the choice of the parameter γ . Then, also for this algorithm, we can adaptively adjust this parameter by the same procedure described in (24).

For what concerns the use of the above inner solvers within the MDF root-finding method, we suggest adopting a careful setting for their stopping tolerance and starting vector. When the solvers are called within the Bracketing Phase, we found convenient to exploit a mild stopping tolerance and the detected image y as starting approximation of x_β while, in the Secant Phase, a more severe stopping tolerance is forced and the last x_β is used as starting vector for a new call of the inner solver. In Appendix 2, we summarize the parameters introduced in the previous algorithms and report the values used in our numerical experiments.

4 Numerical Experiments

This section is devoted to compare the effectiveness of the two approaches for the regularization parameter estimation previously described. The computational study is carried out on test problems arising from deblurring of images corrupted by Poisson noise. In particular, we compare the numerical procedure MDF for the solution of the discrepancy equation (6) with respect to the ADMM approach for solving the constrained model (7). The numerical experiments described in this section have been performed in the MATLAB environment (R2013a) on a workstation equipped with 2 Intel Xeon Six-Core CPUs at 3.1GHz and 188GB of RAM. In the experiments we consider a set of test problems, where the Poisson noise has been simulated by the `imnoise` function in the Matlab Image Processing Toolbox. The considered test problems are described in the following.

cameraman: the test problem is similar to that used in [44]: the original 256×256 image has values scaled in $[0, 3000]$, the image has been corrupted by Gaussian blur (with standard deviation 1.3) and contaminated by Poisson noise; the background is 0 and the values of y are in the range $[75, 2853]$. This test problem is called

Mcameraman. A second test problem, named *Lcameraman*, is generated by multiplying the object by 10, by convolving the image with the same Point Spread Function (PSF) and then by perturbing with Poisson noise; in this case the values of the detected image are in the range [888, 28146]. A third test problem, named *Scameraman*, is generated by proceeding as for the *Lcameraman* but the object is multiplied by $\frac{1}{3}$; the values of the detected image are in the range [19, 938].

spacecraft: the original image is a 256×256 image with sharp details, whose values are in the range [0, 2550] and the background term b is set to 10; following [43], the used PSF simulates that taken by a ground-based telescope and is downloaded from www.mathcs.emory.edu/nagy/RestoreTools/index.html. The values of the detected image are in the range [5, 1135]. This test problem is called *Sspacecraft*. A second test problem, named *Mspacecraft*, is generated by multiplying the object and the background by 10, by convolving the image with the same PSF and then by perturbing with Poisson noise; in this case the values of the detected image are in the range [134, 10766]. A third test problem, named *Lspacecraft*, is generated by proceeding as for the *Mspacecraft* but the object and the background are multiplied by 100; the values of the detected image are in the range [1585, 105383].

Figure 1 shows the original images and the corresponding blurred and noisy images for the test problems *Scameraman* and *Sspacecraft*.

In the subsequent tables, we report the results obtained with one noise realization for each of the six test problems. We checked that our results do not change significantly if the noise realization is changed.

4.1 Estimation of the Regularization Parameter for Hypersurface Potential

The first set of experiments concerns the case in which the regularization term is the HS function (13), with $\delta = 10^{-4} \max_i(y_i)$. Moreover, here and in the following we assume $\eta = 1$.

As concerns Model 1, we solve the discrepancy equation (6) with MDF, the root-finding solver described in Sect. 3.2. In Fig. 2, we show the plot of $\mathcal{D}_H(x_\beta; y) - 1$ with respect to β for the test problems *Mcameraman* and *Sspacecraft*. We observe that, for the second test problem (and, in general, for the *spacecraft* test problems), the behavior of this function around its zero makes very hard the localization of the crossing value. On the other hand, for *Mcameraman* and the related test problems, the computation of the root of $\mathcal{D}_H(x_\beta; y) - 1$ appears easier.

The stopping rule in MDF solver is given by (26), with $\epsilon_1 = 5 \times 10^{-4}$, $\epsilon_2 = 5 \times 10^{-3}$. For what concerns the

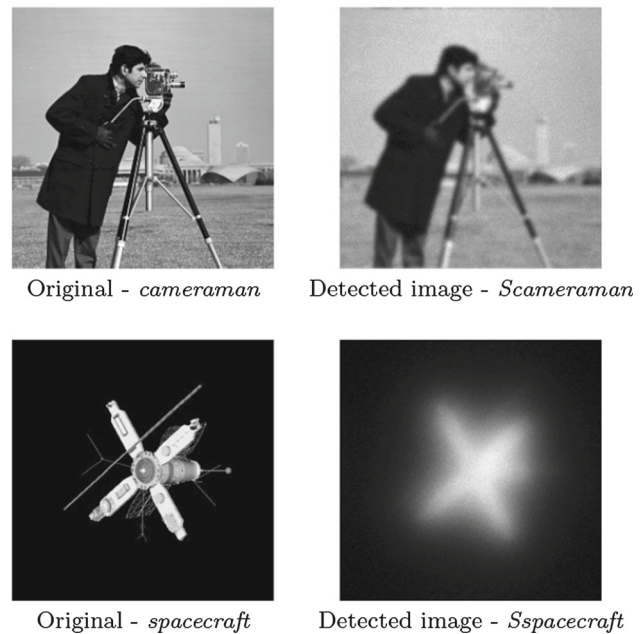


Fig. 1 Original images and detected data for the test problems *Scameraman* and *Sspacecraft*

inner subproblems solvers, in the case where SGP is used, the iterative procedure is stopped when the following criterion is satisfied

$$|f_{\beta_i}(x^{(l)}; y) - f_{\beta_i}(x^{(l-1)}; y)| \leq \epsilon_{inn} |f_{\beta_i}(x^{(l)}; y)| \quad (29)$$

or when a maximum number of iterations, equal to 5000, is performed without obtaining the required accuracy. We recall that f_{β_i} denotes the objective function of the penalized problem (18) with $\beta = \beta_i$. As described in Sect. 3.2, the ϵ_{inn} is a mild tolerance in the Bracketing Phase and it becomes more severe in the MDF Secant Phase, up to a lower bound ϵ_{maxinn} , reported in the caption of the tables concerning Model 1.

When PIDSplit+ is used as inner method, the stopping criterion (29) is coupled with a specific rule for the alternating direction multipliers methods [16], involving current values of the primal and dual residuals (23):

$$\|r^{(l)}\| \leq \epsilon_{pri} \text{ and } \|s^{(l)}\| \leq \epsilon_{dual}, \quad (30)$$

where ϵ_{pri} and ϵ_{dual} are positive feasibility tolerances. These tolerances are defined by means of an absolute and a relative criterion:

$$\epsilon_{pri} = \epsilon_a \sqrt{m + 4n} \quad (31)$$

$$+ \epsilon_r \max(\|Ax^{(l)}\|_2, \|z^{(l)}\|_2, \|c\|_2),$$

$$\epsilon_{dual} = \epsilon_a \sqrt{n} + \epsilon_r \|A^T p^{(l)}\|_2,$$

where $p^{(l)}$ is the current value of the multiplier of the linear constraint $Ax = z + c$; we fix $\epsilon_a = \epsilon_r = \epsilon_1$. Furthermore, in order to avoid the matrix-vector operations $A^T \sigma^{(l)}$ and

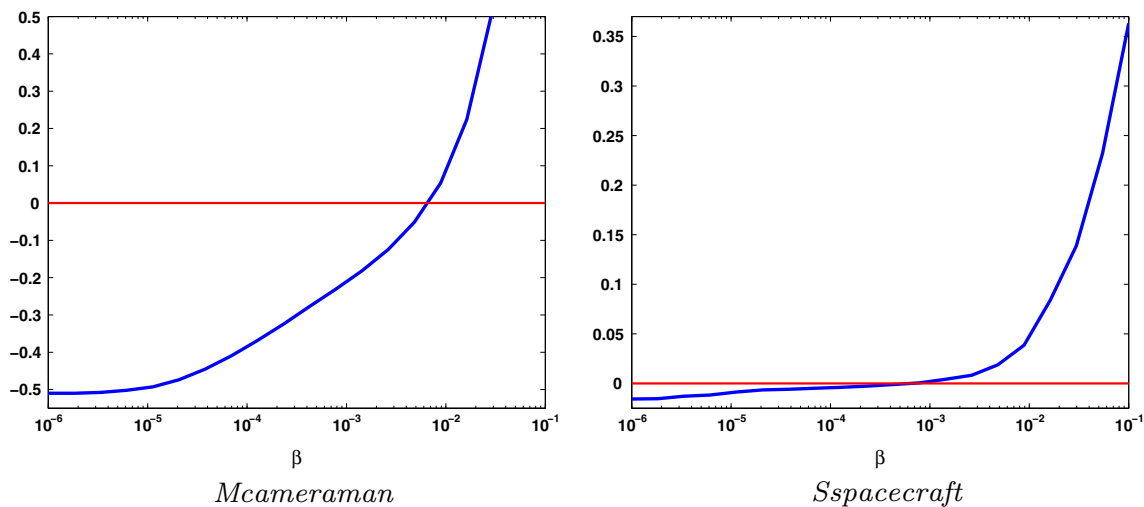


Fig. 2 Behavior of $(\mathcal{D}_H(x_\beta; y) - 1)$ versus β for the test problems *Mcameraman* and *Sspacecraft*

$A^T p^{(l)}$ (arising in the computation of $s^{(l)}$ and ϵ_{dual} , respectively), we overestimate $\|A^T \cdot\|_2$ with $\sqrt{10} \|\cdot\|_2$, where $\sqrt{10}$ is an approximation of $\|A^T\|_2$. Then PIDSplit+ is stopped when one of the conditions (30) or (29) is satisfied or when a maximum number of iteration, equal to 5000, is performed without obtaining the required accuracy. The effectiveness of PIDSplit+ strongly depends on the choice of γ ; following [41], for the solution of the subproblem corresponding to β_i , the value of γ is set equal to $\frac{a}{\beta_i}$ where $a > 0$ is a fixed parameter in MDF; the selection of a suitable value for a requires a trial process. To overcome this drawback, we exploit the updating procedure (24), implementing an adaptive version of PIDSplit+; in this case, the initial value $\gamma^{(0)}$ for the i -th subproblem of MDF is set equal to $\frac{1}{\beta_i}$ while the parameters of the updating procedure are $\mu = 5, \alpha = 2, k_{\text{max}} = 3500$.

Tables 1 and 2 show the results obtained when the regularization parameter is estimated using MDF with inner solver SGP and PIDSplit+, respectively. In the tables, k is the number of iterations required by MDF to satisfy the stopping criterion, k_{tot} is the total number of iterations of the inner solver, β_k is the obtained estimate, $\mathcal{D}_k \equiv \mathcal{D}_H(x_{\beta_k}; y)$, err is the relative reconstruction error $\frac{\|x_{\beta_k} - x^*\|_2}{\|x^*\|_2}$, where x^* is the original object, and, finally, time denotes the execution time in seconds; in particular, for each case, we report the average execution time over ten runs. In Table 2, we show the results of PIDSplit+ corresponding to the value of γ for which we observed the best efficiency and those obtained with the adaptive PIDSplit+. The updating procedure (24) within PIDSplit+ increases the performance of MDF; indeed for each subproblem, the procedure allows to devise suitable values for $\gamma^{(0)}$ and to obtain within a few iterations an approximate solution, satisfying the mild criterion (30) rather than the stopping rule (29). A comparison of the results reported in these two tables demonstrates that SGP is definitely more

efficient than PDSplit+ as inner solver, at least with the exception in the *Lspacecraft* case. The small variations in the values of β_k and in the errors resulting from the comparison of the two tables are not significant as it concerns the quality of the reconstructed images.

The reported results with SGP are obtained by assuming a severe inner tolerance. Therefore we repeated the experiments by reducing this parameter. Table 3 reports the results of the MDF solver when a very weak inner tolerance is used in SGP. Unlike *Lspacecraft*, that requires $\epsilon_{\text{maxinn}} = 10^{-7}$, the other test problems are solved with $\epsilon_{\text{maxinn}} = 10^{-6}$; the restored images are obtained with high efficiency. In spite of this weak inner tolerance, the relative errors do not show significant changes with respect to those of Table 1, since the variations are at most equal to 3%.

As concerns Model 2, it is solved by ADMM solver. For the stopping of the method, before verifying a condition similar to (26), it has to be verified that the current iterate is an approximate solution to the minimization problem; indeed, it could happen that the constraint is satisfied but the solution of the problem is not yet sufficiently accurate. Therefore, when the current iterate $x^{(k)}$ satisfies the condition (30), the solver switches to check that the following stopping rule is satisfied,

$$|\mathcal{D}_H(x^{(k)}; y) - \eta| \leq \epsilon_1 \text{ or } (|\beta_k - \beta_{k-1}| \leq \epsilon_2 |\beta_k| \text{ and } |\mathcal{D}_H(x^{(k)}; y) - \eta| \leq 10\epsilon_1)$$

with $\beta_i = \frac{1}{\lambda^{(i)}}$. The maximum number of ADMM iterations is fixed as 5000. Furthermore, at each step of ADMM, the constrained least-squares subproblem related to z_1 is solved by the Newton's method with a stopping tolerance of 10^{-12} on the distance between two successive approximations. In our experiments, after the initial iterations of ADMM, four

Table 1 Model 1—Regularization term: HS potential

Test problem	k	k_{tot}	β_k	\mathcal{D}_k	err	Time
<i>Scameraman</i>	7	784	1.245×10^{-2}	1.0000	1.157×10^{-1}	14.44
<i>Mcameraman</i>	8	831	6.687×10^{-3}	0.9998	8.563×10^{-2}	15.11
<i>Lcameraman</i>	8	1054	1.236×10^{-3}	1.0004	7.112×10^{-2}	18.91
<i>Sspacecraft</i>	13	2882	5.513×10^{-4}	0.9985	3.017×10^{-1}	49.90
<i>Mspacecraft</i>	14	6475	7.635×10^{-5}	0.9990	2.584×10^{-1}	111.37
<i>Lspacecraft</i>	15	9709	8.053×10^{-6}	0.9969	2.301×10^{-1}	172.78

MDF combined with SGP, $\epsilon_{\text{maxinn}} = 5 \times 10^{-8}$. k is the number of MDF iterations, k_{tot} is the total number of inner solver iterations, β_k is the estimate of β , $\mathcal{D}_k \equiv \mathcal{D}_H(x_{\beta_k}; y)$, err is the relative reconstruction error, time is the execution time in seconds

Table 2 Model 1—Regularization term: HS potential

Test problem	γ	k	k_{tot}	β_k	\mathcal{D}_k	err	Time
<i>Scameraman</i>	$50/\beta_i$	6	732	1.258×10^{-2}	1.0001	1.156×10^{-1}	28.43
<i>Mcameraman</i>	$10/\beta_i$	11	1080	6.359×10^{-3}	1.0008	8.511×10^{-2}	41.90
<i>Lcameraman</i>	$500/\beta_i$	7	682	1.167×10^{-3}	1.0032	7.205×10^{-2}	27.39
<i>Sspacecraft</i>	$50/\beta_i$	4	747	1.000×10^{-4}	0.9997	3.107×10^{-1}	27.45
<i>Mspacecraft</i>	$75/\beta_i$	8	8765	1.024×10^{-4}	0.9999	2.653×10^{-1}	334.56
<i>Lspacecraft</i>	$100/\beta_i$	5	6480	1.000×10^{-5}	0.9999	2.309×10^{-1}	255.95
<i>Scameraman</i>	Adaptive	7	535	1.214×10^{-2}	1.0004	1.155×10^{-1}	21.14
<i>Mcameraman</i>	Adaptive	13	590	6.630×10^{-3}	0.9986	8.516×10^{-2}	24.03
<i>Lcameraman</i>	Adaptive	8	684	1.230×10^{-3}	1.0009	6.974×10^{-2}	27.30
<i>Sspacecraft</i>	Adaptive	8	1801	5.415×10^{-5}	0.9997	3.216×10^{-1}	70.43
<i>Mspacecraft</i>	Adaptive	17	3486	1.005×10^{-6}	1.0009	2.796×10^{-1}	135.96
<i>Lspacecraft</i>	Adaptive	13	4108	7.965×10^{-7}	0.9966	2.381×10^{-1}	159.81

MDF combined with PIDSplit+, $\epsilon_{\text{maxinn}} = 5 \times 10^{-8}$. γ is the parameter appearing in (20) and the other notations are as in Table 1

Table 3 Model 1—Regularization term: HS potential

Test problem	k	k_{tot}	β_k	\mathcal{D}_k	err	Time
<i>Scameraman</i>	8	497	1.231×10^{-2}	0.9996	1.158×10^{-1}	9.36
<i>Mcameraman</i>	8	546	6.599×10^{-3}	1.0002	8.571×10^{-2}	10.11
<i>Lcameraman</i>	10	747	1.241×10^{-3}	1.0008	7.070×10^{-2}	13.75
<i>Sspacecraft</i>	6	1115	3.075×10^{-4}	0.9998	3.142×10^{-1}	19.51
<i>Mspacecraft</i>	5	1461	1.000×10^{-5}	0.9996	2.811×10^{-1}	26.48
<i>Lspacecraft</i>	7	4434	1.000×10^{-6}	0.9995	2.619×10^{-1}	81.15

MDF combined with SGP, $\epsilon_{\text{maxinn}} = 10^{-6}$ except for *Lspacecraft* where $\epsilon_{\text{maxinn}} = 10^{-7}$. For the notation description see Table 1

Newton’s steps enable to compute an approximation of $\lambda^{(i)}$ within the required tolerance.

Table 4 shows the results obtained by means of ADMM with a prefixed value of γ . As in the previous tables, k denotes the number of iterations, β_k the computed estimate of the regularization parameter, err the relative reconstruction error, and time the execution time in seconds. We observe how the effectiveness of ADMM is strongly dependent on the value of γ . Furthermore, in Table 5 we report the results

obtained by ADMM combined with the updating procedure (24), with $\alpha = 2$, $\mu = 10$, and $k_{\text{max}} = 3500$. Here $\gamma^{(0)}$ denotes the initial value of the parameter. This adaptive version of ADMM allows to obtain satisfactory solutions of Model 2 without having to find a suitable value for γ . Also in this case the effectiveness of the method can depend on the computational setting, for example on the value of $\gamma^{(0)}$, but this dependence appears less relevant than for the standard ADMM.

Table 4 Model 2—Regularization term: HS potential

γ	k (Time)	β_k	err	k (Time)	β_k	err	k (Time)	β_k	err
	<i>Scameraman</i>			<i>Mcameraman</i>			<i>Lcameraman</i>		
5	835 (44.43)	1.257×10^{-2}	1.153×10^{-1}	1753 (94.11)	6.714×10^{-3}	8.517×10^{-2}	5000	*	*
10	425 (22.65)	1.251×10^{-2}	1.154×10^{-1}	878 (49.23)	6.714×10^{-3}	8.517×10^{-2}	5000	*	*
50	974 (50.09)	1.346×10^{-2}	1.161×10^{-1}	375 (22.21)	6.888×10^{-3}	8.577×10^{-2}	1139 (60.32)	1.234×10^{-3}	7.083×10^{-2}
150	2974 (158.96)	1.347×10^{-2}	1.160×10^{-1}	1215 (66.91)	6.908×10^{-3}	8.581×10^{-2}	434 (23.57)	1.249×10^{-3}	7.129×10^{-2}
200	3969 (216.68)	1.348×10^{-2}	1.160×10^{-1}	1627 (94.51)	6.908×10^{-3}	8.581×10^{-2}	620 (31.53)	1.252×10^{-3}	7.150×10^{-2}
	<i>Sspacecraft</i>			<i>Mspacecraft</i>			<i>Lspacecraft</i>		
0.05	782 (44.13)	4.260×10^{-5}	3.471×10^{-1}	1436 (82.05)	1.845×10^{-6}	2.916×10^{-1}	2971 (166.57)	7.032×10^{-8}	2.498×10^{-1}
0.1	774 (41.27)	8.506×10^{-5}	3.477×10^{-1}	1442 (85.26)	3.687×10^{-6}	2.915×10^{-1}	2976 (165.86)	1.406×10^{-7}	2.497×10^{-1}
0.5	807 (47.90)	4.167×10^{-4}	3.476×10^{-1}	1501 (86.93)	1.832×10^{-5}	2.909×10^{-1}	3017 (171.53)	7.019×10^{-7}	2.495×10^{-1}
1	957 (51.33)	8.035×10^{-4}	3.434×10^{-1}	1589 (91.59)	3.630×10^{-5}	2.901×10^{-1}	3072 (165.61)	1.401×10^{-6}	2.493×10^{-1}
2.5	2372 (122.26)	1.623×10^{-3}	3.275×10^{-1}	1988 (116.41)	8.745×10^{-5}	2.872×10^{-1}	3262 (188.88)	3.476×10^{-6}	2.484×10^{-1}

ADMM with different values of γ (see (20)). k is the number of ADMM iteration, β_k is the estimate of β , err is the relative reconstruction error

Table 5 Model 2—Regularization term: HS potential

Test problem	k	β_k	\mathcal{D}_k	err	Time
$\gamma^{(0)} = 0.5$					
<i>Scameraman</i>	461	1.234×10^{-2}	0.9987	1.152×10^{-1}	26.03
<i>Mcameraman</i>	452	6.695×10^{-3}	0.9997	8.535×10^{-2}	26.31
<i>Lcameraman</i>	965	1.255×10^{-3}	1.0050	7.151×10^{-2}	56.08
<i>Sspacecraft</i>	807	4.167×10^{-4}	1.0050	3.476×10^{-1}	47.81
<i>Mspacecraft</i>	1501	1.832×10^{-5}	1.0050	2.909×10^{-1}	84.71
<i>Lspacecraft</i>	3017	7.019×10^{-7}	1.0050	2.495×10^{-1}	174.36
$\gamma^{(0)} = 1$					
<i>Scameraman</i>	650	1.253×10^{-2}	0.9997	1.154×10^{-1}	37.67
<i>Mcameraman</i>	451	6.695×10^{-3}	0.9997	8.535×10^{-2}	26.30
<i>Lcameraman</i>	583	1.241×10^{-3}	1.0031	7.113×10^{-2}	33.51
<i>Sspacecraft</i>	957	8.035×10^{-4}	1.0050	3.434×10^{-1}	54.18
<i>Mspacecraft</i>	1589	3.630×10^{-5}	1.0050	2.901×10^{-1}	91.15
<i>Lspacecraft</i>	3072	1.401×10^{-6}	1.0050	2.493×10^{-1}	175.50

Adaptive version of ADMM. $\gamma^{(0)}$ is the initial value for the sequence defined in (24), $\mathcal{D}_k \equiv \mathcal{D}_H(x_{\beta_k}; y)$, time is the execution time in seconds. For the other notations see Table 4

From Tables 1, 2, 3, 4, and 5, we can draw the following remarks.

- The effectiveness of the two approaches for the estimation of the regularization parameter appears very similar; indeed, the time for the solution of the minimization problems involved in Model 1 is comparable to that required for the constrained problem. This is evident from Tables 2 and 5 where the penalized subproblems of the Model 1 and the Model 2 are faced with solvers belonging to the same class of methods.

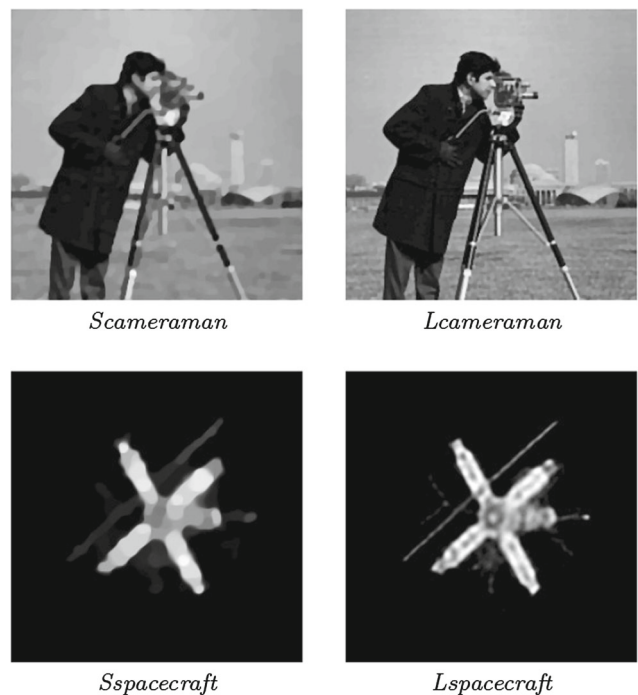


Fig. 3 Some restoration results obtained with HS regularization

- For differentiable regularization term, as HS, there exist efficient iterative algorithms, especially tailored for the solution of penalized problems; this is the case of SGP. When MDF is coupled with this solver, it can be very efficient, as follows by a comparison of Tables 3 and 5, and it allows Model 1 to be less time consuming than Model 2.

Examples of the reconstructions obtained with the HS regularization are given in Fig. 3 for the test problems *Scameraman*, *Lcameraman*, *Sspacecraft*, and *Lspacecraft*.

Table 6 Model 1—Regularization term: TV

Test problem	k	k_{tot}	β_k	\mathcal{D}_k	err	Time
Scameraman	7	535	1.207×10^{-2}	1.0003	1.152×10^{-1}	20.24
Mcameraman	11	600	6.575×10^{-3}	0.9972	8.514×10^{-2}	22.81
Lcameraman	8	715	1.230×10^{-3}	1.0006	6.973×10^{-2}	27.40
Sspacecraft	7	1744	1.000×10^{-5}	0.9998	3.5456×10^{-1}	62.67
Mspacecraft	17	3314	1.005×10^{-6}	1.0009	2.796×10^{-1}	120.95
Lspacecraft	7	2687	1.000×10^{-7}	0.9999	2.401×10^{-1}	95.22

MDF combined with adaptive PIDSplit+, $\epsilon_{\text{maxinn}} = 5 \times 10^{-8}$ and $\gamma^{(0)} = \frac{1}{\beta_i}$, where β_i is the current value of the i -th MDF iteration. The other notations are explained in Table 1

Table 7 Model 2—Regularization term: TV

Test problem	k	β_k	\mathcal{D}_k	err	Time
$\gamma^{(0)} = 0.5$					
Scameraman	448	1.232×10^{-2}	0.9986	1.152×10^{-1}	25.58
Mcameraman	442	6.689×10^{-3}	0.9997	8.534×10^{-2}	24.86
Lcameraman	966	1.254×10^{-3}	1.0050	7.152×10^{-2}	55.07
Sspacecraft	807	4.167×10^{-4}	1.0050	3.476×10^{-1}	44.36
Mspacecraft	1501	1.832×10^{-5}	1.0050	2.909×10^{-1}	83.96
Lspacecraft	3017	7.019×10^{-7}	1.0050	2.495×10^{-1}	169.63
$\gamma^{(0)} = 1$					
Scameraman	640	1.253×10^{-2}	0.9997	1.154×10^{-1}	37.05
Mcameraman	441	6.689×10^{-3}	0.9997	8.534×10^{-2}	24.91
Lcameraman	583	1.240×10^{-3}	1.0031	7.113×10^{-2}	33.42
Sspacecraft	956	8.036×10^{-4}	1.0050	3.435×10^{-1}	53.37
Mspacecraft	1589	3.629×10^{-5}	1.0050	2.901×10^{-1}	85.98
Lspacecraft	3072	1.401×10^{-6}	1.0050	2.493×10^{-1}	165.87

Adaptive version of ADMM. For the notation description see Table 4

4.2 Estimation of the Regularization Parameter for TV Regularization

In this section we describe the results obtained for the test problems by the two models combined with TV regularization. For Model 1, the discrepancy equation (6) is solved by MDF, using as inner solver the adaptive version of PIDSplit+. The stopping criteria for MDF and for the inner solver and the computational setting are the same as that of the experiments on HS. For Model 2, the constrained problem is solved by the adaptive version of ADMM. Also for these experiments, we use the same stopping rules and the computational setting of those described in the previous section.

Tables 6 and 7 show the results obtained for the two models. The analysis of these results leads to confirm the remarks of the previous section: when the MDF algorithm and the Constrained Model exploit a similar optimization solver, the effectiveness of the two approaches appears comparable. In particular, for some applications, we obtain with Model 2

a satisfactory reconstruction in a shorter time. Finally we observe that, with respect to the reconstruction errors, the HS and TV regularizations produce similar solutions.

5 Concluding Remarks

In this paper we investigate two approaches, defined as Model 1 and Model 2, for estimating the regularization parameter in the framework of the Bayesian approach to Poisson data inversion.

Following a discrepancy principle recently proposed for the regularization parameter estimation, the first model asks for the solution of a nonlinear discrepancy equation, while the second model requires the solution of a constrained minimization problem whose constraint is derived from the discrepancy principle. In order to apply the two approaches, a sequence of minimization problems with simple nonnegativity constraints must be faced in the first model, while an optimization solver able to handle more general nonlinear inequality constraints is necessary for solving the minimization problem arising in the second model.

We first extend some results on the relations between the two models; next we introduce an efficient root-finding solver, called MDF, for solving the discrepancy equation introduced by Model 1 and, finally, we compare through a wide computational study the efficiency of the numerical procedures involved by the two models.

The numerical experiments are carried out on two well-known test images and two different regularizations: a differentiable one (HS) and a non-differentiable one (TV). In the first case we show that MDF, combined with an efficient solver for the inner optimization problems, enables Model 1 to be less time demanding than Model 2. On the other hand, in the case of TV regularization, Model 1 can be competitive with Model 2 but the latter is definitely more efficient in some cases.

Acknowledgments This research is partially supported by the project *Learning meets time: A new computational approach for learning in*

dynamic systems, contract RBF12M3A, funded by the Italian Ministry of Education, University and Research *Futuro in Ricerca 2012* program and by the Italian Spinner2013 PhD project *High-complexity inverse problems in biomedical applications and social systems*.

Appendix 1

As already stated, the proof of Theorem 2 can be extended along the lines given in [44] to the case of a differentiable function $R(Lx)$ satisfying the **Assumption REG**. However one point needs a few additional remarks, namely the relation between a solution \bar{x} of problem (10) and the value of $\bar{\lambda}$. Under the assumptions of Theorem 1, using the same arguments of the proof of Lemma 3.3 in [44], it is possible to state that each solution \bar{x} of (10) does not belong to $\mathcal{N}(L)$ and the following relation with $\bar{\lambda}$ holds:

$$\bar{\lambda} = \frac{\langle p_1, L\bar{x} \rangle}{\left\langle \frac{y}{H\bar{x}+b} - 1_m, H\bar{x} \right\rangle}, \tag{32}$$

where $p_1 \in \partial R(L\bar{x})$.

When $R(Lx)$ is a seminorm, i.e., $R(Lx) = \|Lx\|$, since $\bar{x} \notin \mathcal{N}(L)$ and $\|p_1\|_* = 1$ (where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$), we have $\langle p_1, L\bar{x} \rangle = \|L\bar{x}\| > 0$ and, consequently,

$$\bar{\lambda} = \frac{R(L\bar{x})}{\left\langle \frac{y}{H\bar{x}+b} - 1_m, H\bar{x} \right\rangle}. \tag{33}$$

Then, for part ii) of Theorem 1, in the case of the seminorm regularization, $\bar{\lambda}$ is uniquely determined.

When $R(L\bar{x})$ is differentiable, $p_1 = \nabla_t R(t)|_{t=L\bar{x}}$ and the Eq. (32) can be written as:

$$\bar{\lambda} = \frac{\langle \nabla_t R(t)|_{t=L\bar{x}}, L\bar{x} \rangle}{\left\langle \frac{y}{H\bar{x}+b} - 1_m, H\bar{x} \right\rangle}. \tag{34}$$

Then, as in the previous case, for differentiable $R(t)$ satisfying the **Assumption REG**, $\bar{\lambda}$ is uniquely determined.

In particular, for Tikhonov-like regularization,

$$\langle \nabla_t R(t)|_{t=L\bar{x}}, L\bar{x} \rangle = \langle \bar{x}, \nabla_x R(L\bar{x}) \rangle = \|L\bar{x}\|_2^2.$$

For HS regularization, taking into account the expression of the gradient [14], we have $\langle \nabla_t R(t)|_{t=L\bar{x}}, L\bar{x} \rangle = \sum_{i=1}^n \frac{\bar{x}^T L_i^T L_i \bar{x}}{\sqrt{\|L_i \bar{x}\|^2 + \delta^2}}$.

Proof of Lemma 1 (which is a generalization of results in [14]) Under the assumption $\mathcal{N}(L) \cap \mathcal{N}(H_{I_1}) = \{0\}$, $I_1 \neq \emptyset$ and $\lambda > 0$, we prove that the objective function of the problem (10) is strictly convex by showing that the intersection

between the null spaces of the Hessian of $D(y; Hx + b)$ and $R(Lx)$, i.e., the null space of the objective function, is trivial.

Indeed we have that $\mathcal{N}(\nabla^2 D(y; Hx + b)) = \mathcal{N}(H_{I_1})$. On the other hand, the Hessian matrix of $R(Lx)$ is given by $L^T \nabla^2 R(Lx)L$. Since $\nabla^2 R(t)$ is positive definite for any t , $\mathcal{N}(\nabla^2 R(Lx)) = \mathcal{N}(L) = \{x = \operatorname{argmin}_x R(Lx)\}$. As a consequence of the assumption $\mathcal{N}(L) \cap \mathcal{N}(H_{I_1}) = \{0\}$, the objective function is strictly convex and the minimum point is unique. \square

Proof of Theorem 3 When \hat{x} and \tilde{x} are two solutions of problem (11), it is evident that $R(L\hat{x}) = R(L\tilde{x})$. We have to prove that $D(y; H\hat{x} + b) = \tau$ for any solution \hat{x} of the problem. Since \hat{x} is a solution of (11), $\operatorname{lev}_\tau(D(y; Hx + b)) \cap i_{x \geq 0} \neq \emptyset$ and $R(Lx)$ is continuous on its domain \mathbf{R}^n , there exists $v \in L^T \partial R(L\hat{x})$ (or $v = L^T \nabla R(L\hat{x})$ if $R(Lx)$ is differentiable) and a nonnegative scalar μ [40, §28] such that,

$$\begin{aligned} (v + \mu \nabla D(y; H\hat{x} + b))\hat{x} &= 0, \\ v + \mu \nabla D(y; H\hat{x} + b) &\geq 0, \\ \hat{x} &\geq 0. \end{aligned} \tag{35}$$

$$\begin{aligned} \mu(D(y; H\hat{x} + b) - \tau) &= 0, \\ D(y; H\hat{x} + b) &\leq \tau, \\ \mu &\geq 0. \end{aligned}$$

Let $v = L^T p$ for a suitable $p \in \partial R(t)|_{t=L\hat{x}}$ (or $p = \nabla_t R(t)|_{t=L\hat{x}}$). Since $\tau < \tau_L$, $\hat{x} \notin \mathcal{N}(L)$, and, consequently, v is different from 0 (indeed $\operatorname{argmin} R(Lx) = \mathcal{N}(L)$). Then, if $D(y; H\hat{x} + b) < \tau$, since $\mu(D(y; H\hat{x} + b) - \tau) = 0$, it follows $\mu = 0$; therefore, $v\hat{x} = 0$ and the problem is reduced to the minimum of $R(Lx)$ on the nonnegative orthant. Then $x = 0$ is a solution as well as \hat{x} . But $0 \in \mathcal{N}(L)$ and $0 \notin \operatorname{lev}_\tau D(y; Hx + b)$. Then $D(y; H\hat{x} + b) = \tau$.

Now, we prove that, if \hat{x} and \tilde{x} are solutions of (11), then $H_{I_1}\hat{x} = H_{I_1}\tilde{x}$; we write $\hat{x} = \hat{x}_1 + \hat{x}_2$ and $\tilde{x} = \tilde{x}_1 + \tilde{x}_2$ with $\hat{x}_1, \tilde{x}_1 \in \mathcal{R}(H_{I_1}^T)$, $\hat{x}_1 \neq \tilde{x}_1$, and $\hat{x}_2, \tilde{x}_2 \in \mathcal{N}(H_{I_1})$. Let $x = \mu\hat{x} + (1 - \mu)\tilde{x}$, where $\mu \in (0, 1)$ is chosen so that $x \geq 0$. Then

$$\begin{aligned} D(y; Hx + b) &= D(y_{I_1}; H_{I_1}(\mu\hat{x} + (1 - \mu)\tilde{x}) + b) \\ &\quad + \langle 1_{m-m_1}, H_{I_2}(\mu\hat{x} + (1 - \mu)\tilde{x}) + b \rangle \\ &= D(y_{I_1}, H_{I_1}(\mu\hat{x}_1 + (1 - \mu)\tilde{x}_1) + b) \\ &\quad + \mu \langle 1_{m-m_1}, H_{I_2}\hat{x} + b \rangle \\ &\quad + (1 - \mu) \langle 1_{m-m_1}, H_{I_2}\tilde{x} + b \rangle \\ &< \mu(D(y_{I_1}, H_{I_1}\hat{x}_1 + b) + \langle 1_{m-m_1}, H_{I_2}\hat{x} \rangle) \\ &\quad + (1 - \mu)(D(y_{I_1}; H_{I_1}\tilde{x}_1 + b) \\ &\quad + \langle 1_{m-m_1}, H_{I_2}\tilde{x} + b \rangle) \\ &= \mu\tau + (1 - \mu)\tau = \tau, \end{aligned} \tag{36}$$

where the strict inequality follows from the strict convexity of $D(y_{I_1}; H_{I_1}x + b)$ on $\mathcal{R}(H_{I_1}^T)$. Then we have $D(y; Hx + b) < \tau$. On the other hand, we obtain

$$\begin{aligned}
 R(Lx) &= R(\mu L\hat{x} + (1 - \mu)L\tilde{x}) \\
 &\leq \mu R(L\hat{x}) + (1 - \mu)R(L\tilde{x}) \\
 &= R(L\tilde{x}) = R(L\hat{x})
 \end{aligned}
 \tag{37}$$

so that x should be a minimizer of (11). This is impossible since any minimizer has to fulfill $D(y; Hx + b) = \tau$.

Finally, since $D(y; H\hat{x} + b) = D(y; H\tilde{x} + b) = \tau$ and $H_{I_1}\hat{x} = H_{I_1}\tilde{x}$, it follows

$$\langle 1_{m-m_1}, H_{I_2}\tilde{x} \rangle = \langle 1_{m-m_1}, H_{I_2}\hat{x} \rangle
 \tag{38}$$

□

Appendix 2

In this Appendix we specify the setting of the main parameters involved in the methods proposed for facing the two estimation models.

The MDF Method for the Crossing Model

- Stopping criterion: formula (26) is used with $\epsilon_1 = 5 \times 10^{-4}$, $\epsilon_2 = 5 \times 10^{-3}$, and a maximum number of 40 iterations is fixed.
- Starting value for β : 0.1. In the bracketing phase, $d\beta = \omega = 0.1$.
- The computation of the function $\mathcal{F}(\beta)$ is carried out by an inner solver. The tolerance $\epsilon_{inn}^{(i)}$ of the inner solver stopping criterion at the i -th call depends on a parameter ϵ_{maxinn} as follows:

– at the first call of the inner solver

$$\epsilon_{inn}^{(0)} = \max\{\epsilon_{maxinn}, \min\{10^3\epsilon_{maxinn}, 0.1\epsilon_1\}\};$$

– at the i -th call of the bracketing phase, we have

$$\epsilon_{inn}^{(i)} = \max\{\epsilon_{maxinn}, 0.2\epsilon_{inn}^{(i-1)}\};$$

– finally in the secant phase,

$$\epsilon_{inn}^{(i)} = \epsilon_{maxinn}.$$

ϵ_{maxinn} is set equal to 5×10^{-8} in all cases, except in Table 3.

- We consider the following inner solvers:
 - SGP

The maximum number of allowed iterations is 5000. The stopping criterion is

$$|f_{\beta_i}(x^{(l)}; y) - f_{\beta_i}(x^{(l-1)}; y)| \leq \epsilon_{inn}^{(i)} |f_{\beta_i}(x^{(l)}; y)|$$

with $\epsilon_{inn}^{(i)}$ corresponding to one of the choices indicated above and β_i being the current value of β at the i -th MDF iteration.
 - PIDSplit+

The maximum number of allowed iterations is 5000. The stopping criterion is:

$$\begin{aligned}
 (\|r^{(l)}\| \leq \epsilon_{pri} \text{ and } \|s^{(l)}\| \leq \epsilon_{dual}) \\
 \text{or} \\
 |f_{\beta_i}(x^{(l)}; y) - f_{\beta_i}(x^{(l-1)}; y)| \leq \epsilon_{inn}^{(i)} |f_{\beta_i}(x^{(l)}; y)|,
 \end{aligned}$$

where ϵ_{pri} and ϵ_{dual} are specified in (31) with $\epsilon_a = \epsilon_r = \epsilon_1$.

Formula (20) shows the role of the parameter γ .

– Adaptive PIDSplit+

The maximum number of allowed iterations is 5000. The stopping criterion is:

$$\begin{aligned}
 (\|r^{(l)}\| \leq \epsilon_{pri} \text{ and } \|s^{(l)}\| \leq \epsilon_{dual}) \\
 \text{or} \\
 |f_{\beta_i}(x^{(l)}; y) - f_{\beta_i}(x^{(l-1)}; y)| \leq \epsilon_{inn}^{(i)} |f_{\beta_i}(x^{(l)}; y)|,
 \end{aligned}$$

where ϵ_{pri} and ϵ_{dual} are specified in (31) with $\epsilon_a = \epsilon_r = \epsilon_1$.

The updating rule for $\gamma^{(l)}$ is given in (24), with $\alpha = 2$, $\mu = 5$, $k_{max} = 3500$; $\gamma^{(0)} = \frac{1}{\beta_i}$.

The ADMM Method for the Constrained Model

- The constrained problem for computing $\frac{1}{\beta_i}$ at the i -th iteration is solved by a Newton's iteration with a stopping tolerance 10^{-12} on the distance between two successive approximations.
- ADMM is stopped when the following two conditions are satisfied:

$$(\|r^{(k)}\| \leq \epsilon_{pri} \text{ and } \|s^{(k)}\| \leq \epsilon_{dual})$$

and

$$\begin{aligned}
 (|\mathcal{D}_H(x^{(k)}; y) - \eta| \leq \epsilon_1 \text{ or} \\
 (|\beta_k - \beta_{k-1}| \leq \epsilon_2 |\beta_k| \text{ and } |\mathcal{D}_H(x^{(k)}; y) - \eta| \leq 10\epsilon_1));
 \end{aligned}$$

here ϵ_{pri} and ϵ_{dual} , depending on $x^{(k)}$, are specified in (31), with $\epsilon_a = \epsilon_r = \epsilon_1$. As in MDF, $\epsilon_1 = 5 \times 10^{-4}$, $\epsilon_2 = 5 \times 10^{-3}$.

- Maximum number of allowed iterations: 5000.
- Two versions of ADMM are implemented:
 - standard version: the value of γ is a priori fixed;
 - adaptive version: $\gamma^{(i)}$ is given by the updating rule (24) with $\alpha = 2$, $\mu = 10$, $k_{max} = 3500$; $\gamma^{(0)}$ is set as 0.5 or 1.

References

1. Arrow, K.J., Hurwicz, L., Uzawa, H.: *Studies in Linear and Non-linear Programming*. Stanford University Press, Stanford (1958)
2. Bardsley, J.M., Goldes, J.: Regularization parameter selection methods for ill-posed Poisson maximum likelihood estimation. *Inverse Problems* **25**, 095005 (2009)
3. Barzilai, J., Borwein, J.: Two point step size gradient methods. *IMA J. Numer. Anal.* **8**, 141–148 (1988)
4. Ben Hadj, S., Blanc-Feraud, L., Aubert, G., Engler, G.: Blind restoration of confocal microscopy images in presence of a depth-variant blur and poisson noise. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, pp. 915–919 (2013)
5. Benvenuto, F., Zanella, R., Zanni, L., Bertero, M.: Nonnegative least-squares image deblurring: improved gradient projection approaches. *Inverse Problems* **26**(2), 025004 (2010)
6. Bertero, M., Boccacci, P., Desiderá, G., Vicidomini, G.: Image deblurring with Poisson data: from cells to galaxies. *Inverse Problems* **25**, 123006 (2009)
7. Bertero, M., Boccacci, P., Talenti, G., Zanella, R., Zanni, L.: A discrepancy principle for Poisson data. *Inverse Problems* **26**, 104004 (2010)
8. Bertsekas, D.: *Nonlinear Programming*. Athena Scientific, Belmont (1999)
9. Bertsekas, D., Tsitsiklis, J.: *Parallel and Distributed Computation: Numerical Methods*. Prentice-Hall, Englewood Cliffs (1988)
10. Bonettini, S., Cornelio, A., Prato, M.: A new semiblind deconvolution approach for Fourier-based image restoration: an application in astronomy. *SIAM J. Imaging Sci.* **6**, 1736–1757 (2013)
11. Bonettini, S., Landi, G., Loli Piccolomini, E., Zanni, L.: Scaling techniques for gradient projection-type methods in astronomical image deblurring. *Int. J. Comput. Math.* **90**(1), 9–29 (2013)
12. Bonettini, S., Prato, M.: Nonnegative image reconstruction from sparse Fourier data: a new deconvolution algorithm. *Inverse Problems* **26**, 095001 (2010)
13. Bonettini, S., Prato, M.: A new general framework for gradient projection methods. Submitted [arXiv:1406.6601v1](https://arxiv.org/abs/1406.6601v1) (2014)
14. Bonettini, S., Ruggiero, V.: On the uniqueness of the solution of image reconstruction problems with Poisson data, vol. 1281. In: Simos, T.E. et al. (ed.) *Proceedings of ICNAAM 2010, AIP Conference Proceedings*, pp. 1803–1806, AIP (2010)
15. Bonettini, S., Zanella, R., Zanni, L.: A scaled gradient projection method for constrained image deblurring. *Inverse Problems* **25**, 015002 (2009)
16. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends R Mach. Learn.* **3**(1), 1–122 (2011)
17. Carlván, M., Blanc-Féraud, L.: Two constrained formulations for deblurring Poisson noisy images. In: *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Brussels, Belgium (2011)
18. Carlván, M., Blanc-Féraud, L.: Sparse Poisson noisy image deblurring. *IEEE Trans. Image Process.* **21**(4), 1834–1846 (2012)
19. Cavicchioli, R., Chau, C., Blanc-Feraud, L., Zanni, L.: ML estimation of wavelet regularization hyperparameters in inverse problems. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, pp. 1553–1557 (2013)
20. Chambolle, A., Pock, T.: A first-order primal–dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**, 120–145 (2011)
21. Ciak, R., Shafei, B., Steidl, G.: Homogeneous penalizers and constraints in convex image restoration. *J. Math. Imaging Vis.* **47**(3), 210–230 (2012)
22. Dai, Y.H., Fletcher, R.: New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds. *Math. Program. Ser. A* **106**, 403–421 (2006)
23. Daube Witherspoon, M.E., Muehlener, G.: An iterative image space reconstruction algorithm suitable for volume ECT. *IEEE Trans. Med. Imaging* **5**(2), 61–66 (1986)
24. Engl, H.W., Hanke, M., Neubauer, A.: *Regularization of Inverse Problems*. Kluwer Academic Publishers, Dordrecht (1996)
25. Figueiredo, M.A.T., Bioucas-Dias, J.M.: Restoration of Poissonian images using alternating direction optimization. *IEEE Trans. Image Process.* **19**, 3133–3145 (2010)
26. Fletcher, R.: A limited memory steepest descent method. *Math. Program.* **135**(1–2), 413–436 (2012)
27. Frassoldati, G., Zanni, L., Zanghirati, G.: New adaptive stepsize selections in gradient methods. *J. Ind. Manag. Optim.* **4**(2), 299–312 (2008)
28. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite elements approximations. *Comput. Math. Appl.* **2**, 17–40 (1976)
29. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**(6), 721–741 (1984)
30. Glowinski, R., Marrocco, A.: Sur l’approximation, par elements finis d’ordre un, et la resolution, par penalisation-dualité, d’une classe de problèmes de dirichlet non lineares. *Revue Française d’Automatique, Infomatique, et Recherche Opérationnelle* (1975)
31. Golub, G., Van Loan, C.: *Matrix Computations*, 3rd edn. The Johns Hopkins Press, Baltimore (1996)
32. He, B.S., Yang, Y., Wang, S.L.: Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities. *J. Optim. Theory Appl.* **106**(2), 337–356 (2000)
33. Hiriart, J.B., Lemarechal, C.: *Convex Analysis and Minimization Algorithms*, vol. 1. Springer, Berlin (1993)
34. Lantéri, H., Roche, M., Aime, C.: Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms. *Inverse Problems* **18**(5), 1397–1419 (2002)
35. Le, T., Chartrand, R., Asaki, T.J.: A variational approach to reconstructing images corrupted by Poisson noise. *J. Math. Imaging Vis.* **27**, 257–263 (2007)
36. Lucy, L.B.: An iterative technique for the rectification of observed distributions. *Astron. J.* **79**, 745–754 (1974)
37. Porta, F., Zanella, R., Zanghirati, G., Zanni, L.: Limited-memory scaled gradient projection methods for real-time image deconvolution in microscopy. *Commun. Nonlinear Sci. Numer. Simul.* **21**, 112–127 (2015)
38. Prato, M., Cavicchioli, R., Zanni, L., Boccacci, P., Bertero, M.: Efficient deconvolution methods for astronomical imaging: algorithms and IDL-GPU codes. *Astron. Astrophys.* **539**, A133 (2012)
39. Richardson, W.H.: Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am. A* **62**(1), 55–59 (1972)
40. Rockafellar, R.T.: *Convex Analysis*. Princeton University Press, Princeton (1970)
41. Setzer, S., Steidl, G., Teuber, T.: Deblurring Poissonian images by split Bregman techniques. *J. Vis. Commun. Image R.* **21**, 193–199 (2010)
42. Shepp, L.A., Vardi, Y.: Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging* **1**(2), 113–122 (1982)
43. Staglianò, A., Boccacci, P., Bertero, M.: Analysis of an approximate model for Poisson data reconstruction and a related discrepancy principle. *Inverse Problems* **27**, 125003 (2011)

44. Teuber, T., Steidl, G., Chan, R.H.: Minimization and parameter estimation for seminorm regularization models with I-divergence constraints. *Inverse Problems* **29**, 035007 (2013)
45. Wang, S.L., Liao, L.Z.: Decomposition method with a variable parameter for a class of monotone variational inequality problems. *J. Optim. Theory Appl.* **109**(2), 415–429 (2001)
46. Zanella, R., Boccacci, P., Zanni, L., Bertero, M.: Efficient gradient projection methods for edge-preserving removal of Poisson noise. *Inverse Problems* **25**, 045010 (2009)
47. Zanella, R., Boccacci, P., Zanni, L., Bertero, M.: Corrigendum: efficient gradient projection methods for edge-preserving removal of Poisson noise. *Inverse Problems* **29**, 119501 (2013)
48. Zanella, R., Zanghirati, G., Cavicchioli, R., Zanni, L., Boccacci, P., Bertero, M., Vicidomini, G.: Towards real-time image deconvolution: application to confocal and sted microscopy. *Scientific Reports* **3** (2013)



Luca Zanni born in 1965, graduated in Mathematics cum laude in 1990 at the University of Modena, Italy. Since 2005, he is full professor of Numerical Analysis at the University of Modena and Reggio Emilia. From 2010 to 2012 he was the Director of the Department of Mathematics and currently he is Vice Director of the Department of Physics, Computer Science, and Mathematics. His research fields are numerical optimization, parallel computing, statistical learning, and

inverse problems. In these fields, he studied acceleration techniques for gradient-type iterative methods and he developed parallel software for solving large-scale optimization problems arising from the areas of image reconstruction and machine learning.



Alessandro Benfenati born in 1986 in Bologna, graduated in Mathematics summa cum laude at the University of Bologna in 2011. From 2012 he is a Doctorate Student at the University of Ferrara, in the Mathematics and Computer Science department; he is taking part in the project *High-complexity inverse problems in biomedical applications and social systems* (<http://www.ipbioss.unimore.it/site/home.html>) funded by the Italian Spinner Consor-

tium. His research interests lie in convex analysis focusing on Bregman Distance-based methods, optimization theory, with particular regards to ill-posed problems, numerical methods for optimization, parallel computing, and support vector machines. His major contributions regard the development and employment of optimization algorithms, related to image restoration problems in frameworks such as biomedicine, microscopy, and astronomy; moreover, he applied some numerical methods to face problems regarding Statistical Mechanics applied to Social Science.



Mario Bertero born in 1938, received the “laurea” in Physics from the University of Genova in 1960 and the “libera docenza” in Theoretical Physics in 1968. After several research and teaching appointments at the Universities of Genova, Bonn, and Bruxelles, in 1981 he became professor of Mathematics at the University of Genova and from 1995 he is professor of Information Science at the same University. He was editor-in-chief of the journal *Inverse Problems* from 1989 to

1994. His first research activity was on the theory of Regge poles and its applications to Nuclear Physics, on the quantum theory of potential scattering, and on the optical model of nuclear reactions. Next he started to work on the theory and applications of inverse and ill-posed problems both from the point of view of a probabilistic approach and from that of Tikhonov regularization theory.



Valeria Ruggiero received the Laurea degree in Mathematics (cum laude) from the University of Ferrara in 1978. In 1981–2000 she was researcher and then associate professor of Numerical Analysis at the Universities of Ferrara and Modena. Since 2000, she is Full Professor of Numerical Analysis at the Department of Mathematics and Computing Science in the University of Ferrara. Actually she is Director of the National Group for Scientific Computation of the Istituto

Nazionale di Alta Matematica (INdAM). Her research interests are focused in numerical methods for large-scale systems, nonlinear optimization, parallel computing, and related applications.