Journal of Computational and Applied Mathematics 231 (2009) 236–248

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Non-negatively constrained image deblurring with an inexact interior point method

Silvia Bonettini [a,*], Thomas Serafini [b]

[a] Dipartimento di Matematica, Università di Ferrara, Italy

[b] Dipartimento di Matematica, Università di Modena e Reggio Emilia, Italy

ARTICLE INFO

Article history:
Received 17 July 2007
Received in revised form 6 October 2008

Keywords:
Image deblurring
Deconvolution methods
Interior point algorithms
Regularization techniques

ABSTRACT

Nonlinear image deblurring procedures based on probabilistic considerations have been widely investigated in the literature. This approach leads to model the deblurring problem as a large scale optimization problem, with a nonlinear, convex objective function and non-negativity constraints on the sign of the variables. The interior point methods have shown in the last years to be very reliable in nonlinear programs. In this paper we propose an inexact Newton interior point (IP) algorithm designed for the solution of the deblurring problem. The numerical experience compares the IP method with another state-of-the-art method, the Lucy Richardson algorithm, and shows a significant improvement of the processing time.

1. Introduction

In image deblurring problems, image formation is modeled as a first kind Fredholm integral equation which, after discretization, results in a large-scale system of the form

$$A\boldsymbol{x} = \boldsymbol{b}, \tag{1}$$

where A is an $N \times N$ matrix and $\boldsymbol{b} \in \mathbb{R}^N$ is given by the sum of two terms $\boldsymbol{b} = \boldsymbol{g} + \eta$: $\boldsymbol{g} \in \mathbb{R}^N$ is the blurred image that would have been recorded in absence of noise, $\eta \in \mathbb{R}^N$ is the cumulative effect of the noise processes affecting the acquisition of the final image [1]. Here and in the following, a two-dimensional image $\boldsymbol{x} \in \mathbb{R}^{n\times n}$ is represented as a vector $\boldsymbol{x} \in \mathbb{R}^N$, $N = n^2$ by stacking column by column. The image restoration problem is then to obtain an approximation of $\boldsymbol{x}$, knowing A and $\boldsymbol{b}$. The matrix A represents the physical effect of the instrument and, since the system (1) is given by the discretization of an ill-posed problem [1], it could be a very ill-conditioned matrix, with singular values decaying to zero, with no significant gap to estimate the rank. Thus, an iterative approach to the solution of (1) could be preferable.

Several iterative algorithms proposed in the literature are based on the maximization of the maximum likelihood (ML) function, that expresses the probability $p(\boldsymbol{b}|\boldsymbol{x})$ [2–6]. Such methods require a hypothesis on the image formation process. Two noise processes are generally considered: the Gaussian noise, that is independent of the value of the pixels, and the Poisson noise. Furthermore, a non-negativity constraint on the components x_i is imposed, due to the physical meaning of the variables.

In the Poisson noise case, one of the most popular algorithms is the Lucy–Richardson (LR), or Expectation Maximization (EM), algorithm [2,3,5]. The LR algorithm has a very simple form, does not require any step-length computation, and it is convergent to a positive maximum of the ML function [7], but it exhibits a low convergence speed, especially close to the

* Corresponding author.
E-mail addresses: bntslv@unife.it, bonettini.silvia@unimo.it (S. Bonettini), serafini.thomas@unimore.it (T. Serafini).

0377-0427/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2009.02.020

optimal solution. This behavior can be explained by observing that the LR algorithm actually is a scaled steepest descent method [8,9], thus its convergence rate is linear. To get a better convergence rate, an alternative of the gradient direction should be considered.

Recently, the Newton or quasi-Newton approach has been proposed for the non-negative image restoration problem [10–14]. However, the computation of the Newton direction could be very costly in terms of time, since it is defined as the solution of a linear system. Thus, the convenience of a Newton approach is preserved if the cost of the computation of the direction is balanced by a very fast convergence of the overall algorithm.

In this paper we propose a Newton interior point (IP) algorithm for the maximum likelihood problem that has been specialized to the image deblurring problem. The main advantage of the proposed method, is that the Newton direction can be computed inexactly, with an adaptive tolerance which preserves the global convergence.

We provide also the possibility to include an entropy penalization–regularization term. Furthermore, we analyze a recently-proposed variant of the LR algorithm for the regularized problem.

In our numerical experiments on astronomical test images, we show the rapid convergence of the algorithm compared to the LR method, on both the regularized and non-regularized cases.

The paper is organized as follows: in Section 2 we explain the statistical model and we describe in detail the nonlinear programming problem to be solved. Moreover, we describe the main features of the LR method. In Section 3 we present the interior point algorithm, focusing on the direction determination, the line-search technique and the initialization. Section 4 deals with the regularization techniques that can be applied in the ML problem and we consider a regularized variant of the LR method. The numerical experience comparing the interior point and the LR method on some astronomical test images, with and without regularization, is presented in Section 5. Our conclusion and perspectives are given in Section 6.

2. The maximum likelihood problem

In a probabilistic framework, the element b_i of the observed image $\boldsymbol{b}$ can be interpreted as the realization of a Poisson distribution with mean $(A\boldsymbol{x})_i = \sum_j A_{ij}x_j$. Accordingly,

$$p(b_i|\boldsymbol{x}) = \frac{\mathrm{e}^{-\sum_j A_{ij}x_j}\left(\sum_j A_{ij}x_j\right)^{b_i}}{b_i!}.$$

Moreover, each component (pixel) of the observed image $\boldsymbol{b}$ is subject to an independent Poisson process. The global likelihood of the observed image is therefore

$$p(\boldsymbol{b}|\boldsymbol{x}) = \prod_i \frac{\mathrm{e}^{-\sum_j A_{ij}x_j}\left(\sum_j A_{ij}x_j\right)^{b_i}}{b_i!}.$$

Thus, the maximum likelihood criterion leads to seeking for an approximate solution of the problem (1) that maximizes $p(\boldsymbol{b}|\boldsymbol{x})$, or, equivalently, that minimizes

$$-\log(p(\boldsymbol{b}|\boldsymbol{x})) = \sum_i \left(\sum_j A_{ij}x_j - b_i \log\left(\sum_j A_{ij}x_j\right) + \log(b_i!)\right).$$

By using the Stirling formula $\log(b_i!) = -b_i + b_i \log b_i$ we obtain the Csiszär divergence [15], that is a generalized form of the Kullback–Leibler measure of the distance between the blurred image $A\boldsymbol{x}$ and the observed image $\boldsymbol{b}$:

$$KL(A\boldsymbol{x}, \boldsymbol{b}) = \sum_i \left(\sum_j A_{ij}x_j - b_i - b_i \log \frac{\left(\sum_j A_{ij}x_j\right)}{b_i}\right).$$

As in [16,17], we consider a slight modification of the Csiszär divergence, adding a constant background term bg. In a more compact notation, the functional can be written as

$$J(\boldsymbol{x}) = \sum_i (A\boldsymbol{x})_i + bg - b_i - b_i \log \frac{(A\boldsymbol{x})_i + bg}{b_i}. \tag{2}$$

Taking into account of the physical meaning of the variables x_i, we impose a non-negativity constraint: therefore, the maximum likelihood problem becomes

$$\begin{array}{ll} \min & J(\boldsymbol{x}). \\ \boldsymbol{x} \geq 0 & \end{array} \tag{3}$$

The gradient and hessian of the objective function J are

$$\nabla J(\boldsymbol{x}) = A^{\mathrm{T}}\boldsymbol{e} - A^{\mathrm{T}}Y^{-1}\boldsymbol{b} \tag{4}$$

$$\nabla^2 J(\boldsymbol{x}) = A^{\mathrm{T}}BY^{-2}A, \tag{5}$$

where $\boldsymbol{e}$ is a vector whose N components are all equal to one, $B = diag(\boldsymbol{b})$ and $Y = diag((A\boldsymbol{x}) + bg)$.

The matrix A has positive entries and it is generally dense. However, considering the boundary conditions imposed for the discretization of the Fredholm integral equation, we can devise an important property of the matrix A. In this paper we will assume that periodic boundary conditions are imposed, so that the matrix A is block-circulant with circulant blocks BCCB.

This property is crucial in order to perform the matrix–vector products quickly: in fact, the products $A\boldsymbol{x}$ can be performed with a $\mathcal{O}(N \log N)$ complexity, by employing the Fast Fourier Transform (FFT) [18].

Since the components b_i of the observed image are non-negatives, the hessian matrix (5) is positive semidefinite in the feasible region defined by the constraints, thus the problem (3) is convex and any local minimum is also a global minimum. The Karush–Kuhn–Tucker (KKT) first order optimality conditions for (3) can be stated as follows:

$$\begin{aligned} &\boldsymbol{x}^{\mathrm{T}}\nabla J(\boldsymbol{x}) = 0 \\ &\nabla J(\boldsymbol{x}) \geq 0 \\ &\boldsymbol{x} \geq 0. \end{aligned} \tag{6}$$

If the matrix A is normalized such that $A^{\mathrm{T}}\boldsymbol{e} = \boldsymbol{e}$, the complementarity condition (6) becomes

$$\boldsymbol{x} = XA^{\mathrm{T}}Y^{-1}\boldsymbol{b},$$

where $X = diag(\boldsymbol{x})$. The Lucy–Richardson (LR) method [2,3,5] is based on the previous fixed point formulation, and its iteration is defined as

$$\boldsymbol{x}^{k+1} = X_k A^{\mathrm{T}} Y_k^{-1}\boldsymbol{b}. \tag{7}$$

Starting from a positive initial point, the non-negativity of A and $\boldsymbol{b}$ guarantees that all the successive iterates remain feasible and, furthermore, the convergence to a solution of (3) can be proved [7].

The LR method is attractive because of its convergence properties and its low computational cost, consisting in $\mathcal{O}(N \log N)$ operations (needed to perform the matrix vector product $A^{\mathrm{T}}Y\boldsymbol{b}$) per iteration. These nice features made the method one of the most popular for astronomical and medical image restoration problems. However, the main drawback of the method is a very slow convergence, that, in many cases, leads to hundreds or thousands of iterations to obtain an approximation of the solution of the desired quality. In fact, since the iteration (7) can be written as

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - X_k \nabla J(\boldsymbol{x}_k)$$

the LR method can be interpreted as a scaled steepest descent method, with the scaling matrix given by X_k, therefore its convergence rate is linear. On the other hand, the fixed point iteration (7) implies that the convergence rate of the LR method is governed by the spectral radius of the operator $G(\boldsymbol{x}) = XA^{\mathrm{T}}Y^{-1}\boldsymbol{b}$ in the solution. If the spectral radius is close to the unit, a very low convergence has to be expected.

This drawback of the LR method motivated us to propose an interior point method, which exploits the Newton direction instead of the gradient direction, for the problem (3).

3. A primal–dual approach

The method proposed in this section applies to every constrained nonlinear programming problem of the form

$$\begin{aligned} &\min \quad f(\boldsymbol{x}). \\ &\boldsymbol{x} \geq 0 \end{aligned} \tag{8}$$

In spite of what is usual in the interior point framework, we do not introduce slackness on the positivity constraint. Indeed, taking into account of the physical meaning of the variables and of the domain of the objective function (2), we design a feasible method, i.e. a method which generates a sequence $\{\boldsymbol{x}^k\}$ such that $\boldsymbol{x}^k > 0$ for any k. The iterate $\boldsymbol{x}^{k+1}$ is computed along the Newton direction by means of a line-search procedure, and the step length has to guarantee the positivity and the sufficient decrease of a merit function.

The main feature of this algorithm is that the Newton direction is computed inexactly by means of an iterative procedure associated with a termination rule guaranteeing the global convergence of the algorithm.

The Karush–Kuhn–Tucker (KKT) first-order optimality conditions of the problem (8) can be stated by introducing the Lagrange multiplier $\boldsymbol{w}$ for the constraints as follows:

$$\nabla f(\boldsymbol{x}) - \boldsymbol{w} = 0 \tag{9}$$

$$w_i x_i = 0 \quad i = i, \ldots, N \tag{10}$$

$$\boldsymbol{x}, \boldsymbol{w} \geq 0. \tag{11}$$

The variables x_i are the primal variables, while the Lagrange multipliers w_i are called dual variables. In a primal–dual approach, $\boldsymbol{x}$ and $\boldsymbol{w}$ represents the unknowns of the nonlinear system (9)–(10), subject to the non-negativity constraint (11). The interior point methods basic idea consists in solving the nonlinear system (9)–(10) generating a sequence of iterates $\{(\boldsymbol{x}^k, \boldsymbol{w}^k)\}$ which approximately solve the subproblems

$$(\mathcal{P}_k) \quad \begin{array}{l} \nabla f(\boldsymbol{x}) - \boldsymbol{w} = 0 \\ w_i x_i = \rho_k \quad i = i, \ldots, N \\ \boldsymbol{x}, \boldsymbol{w} > 0 \end{array} \tag{12}$$

where $\rho_k = \mu_k \sigma_k$, is the product of two positive parameters. The parameter $\mu_k > 0$ is a perturbation parameter and $\sigma_k \leq \sigma_{max} < 1$ is a positive forcing term. Given the point $(\boldsymbol{x}^k, \boldsymbol{w}^k)$ and the perturbation parameter ρ_k, the Newton direction is defined as the vector $\boldsymbol{d}^{k\mathrm{T}} = (\Delta \boldsymbol{x}^{k\mathrm{T}}, \Delta \boldsymbol{w}^{k\mathrm{T}})^{\mathrm{T}}$ which solves the linear system

$$\begin{pmatrix} \nabla^2 f(\boldsymbol{x}^k) & -I \\ W_k & X_k \end{pmatrix} \begin{pmatrix} \Delta \boldsymbol{x}^k \\ \Delta \boldsymbol{w}^k \end{pmatrix} = - \begin{pmatrix} \nabla f(\boldsymbol{x}^k) - \boldsymbol{w}^k \\ W_k X_k \boldsymbol{e} - \rho_k \boldsymbol{e} \end{pmatrix} \tag{13}$$

where X_k and W_k are diagonal matrices whose entries are the components of the vectors $\boldsymbol{x}^k$ and $\boldsymbol{w}^k$.

By eliminating the second block of equations, we get the *condensed system*

$$M_k \Delta \boldsymbol{x}^k = -\nabla f(\boldsymbol{x}^k) + \rho_k X_k^{-1} \boldsymbol{e} \tag{14}$$

$$\Delta \boldsymbol{w}^k = -\boldsymbol{w}^k - X_k^{-1} W_k \Delta \boldsymbol{x}^k + \rho_k X_k^{-1} \boldsymbol{e}, \tag{15}$$

where $M_k = \nabla^2 f(\boldsymbol{x}^k) + X_k^{-1} W_k$. The new point is computed as

$$\begin{array}{l} \boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha_k \Delta \boldsymbol{x}^k \\ \boldsymbol{w}^{k+1} = \boldsymbol{w}^k + \alpha_k \Delta \boldsymbol{w}^k \end{array} \tag{16}$$

where the step length along the direction $\boldsymbol{d}^{k\mathrm{T}}$ is given by the damping parameter $\alpha_k \in (0, 1]$, which has to ensure the positivity of the new point, i.e. $(\boldsymbol{x}^{k+1}, \boldsymbol{w}^{k+1}) > 0$. To this end, the step length is reduced as

$$\alpha = C \cdot \min_{\Delta x_i < 0, \Delta w_i < 0} (1, -x_i/\Delta x_i, -w_i/\Delta w_i), \tag{17}$$

where C is a parameter less than one.

Furthermore, the step length may be reduced in order to satisfy the two following *centrality conditions* (see [19]), which require that the new point satisfies

$$\min_i (x_i^k + \alpha_k \Delta x_i^k)(w_i^k + \alpha_k \Delta w_i^k) \geq (\tau_1/n)(\boldsymbol{x}^k + \alpha_k \Delta \boldsymbol{x}^k)(\boldsymbol{w}^k + \alpha_k \Delta \boldsymbol{w}^k) \tag{18}$$

$$(\boldsymbol{x}^k + \alpha_k \Delta \boldsymbol{x}^k)(\boldsymbol{w}^k + \alpha_k \Delta \boldsymbol{w}^k) \geq \tau_2 \|\nabla f(\boldsymbol{x}^k) - \boldsymbol{w}^k\|. \tag{19}$$

for some positive constants τ_1 and τ_2. The meaning of (18) and (19) can be explained by recalling that the k-th iterate of an interior point method is an approximate solution of the nonlinear system (12). Then, instead of performing several Newton steps to solve (12), we can improve the accuracy of such approximation moving along the Newton direction (13). It can be proved [20] that there exists a value α_k such that (18) and (19) are satisfied.

Finally, the damping parameter has to guarantee a sufficient decrease of the merit function

$$\varphi(\boldsymbol{x}, \boldsymbol{w}) = \sqrt{\|\nabla f(\boldsymbol{x}) - \boldsymbol{w}\|^2 + \|WX\boldsymbol{e}\|^2}. \tag{20}$$

Here and in the following $\|\cdot\|$ indicates the euclidean norm, hence our merit function is a measure of the violation of the KKT conditions (9)–(10). We adopt the following line-search rule, associated with the merit function (20),

$$\varphi(\boldsymbol{x}^k + \alpha_k \Delta \boldsymbol{x}^k, \boldsymbol{w}^k + \alpha_k \Delta \boldsymbol{w}^k) \leq (1 - \beta(1 - \delta_k - \sigma_k)) L_k, \tag{21}$$

where $\beta < 1$ is a positive parameter and L_k is a reference value defined as follows

$$L_k = \max_{\bar{k} \leq i \leq k} \varphi(\boldsymbol{x}^i, \boldsymbol{w}^i), \quad \bar{k} = \max(0, k - M). \tag{22}$$

The criterion (22) (see [21]) allows a non-monotone descent of the merit function. The monotonicity degree is set by the parameter M; for $M = 0$ we found the line-search rule proposed in [22], where the authors show that it includes the Armijo rule. As pointed out in [21,23], the non-monotone techniques can be useful in the step length computation, avoiding too strict requirements on the descent of the merit function. Indeed, every backtracking step requires a gradient evaluation, thus a good choice is a line-search rule that requires as few reductions as possible.

After computing the new point with the updating rule (16), the perturbation parameter is updated such that

$$\mu_{k+1} \in [\mu_{k+1}^{(1)}, \mu_{k+1}^{(2)}] \tag{23}$$

where

$$\mu_{k+1}^{(1)} = \frac{(\boldsymbol{w}^{k+1})^{\mathrm{T}}\boldsymbol{x}^{k+1}}{N} \quad \text{and} \quad \mu_{k+1}^{(2)} = \frac{\varphi(\boldsymbol{x}^{k+1}, \boldsymbol{w}^{k+1})}{\sqrt{N}}. \tag{24}$$

The choice $\mu^{(1)}$ for the perturbation parameter is typical of the interior point methods and plays an important role in the convergence analysis. The choice $\mu^{(2)}$ is related to the inexact Newton method [22,24]. More precisely, with the choice (23)–(24), the algorithm presented in this section can be considered an inexact Newton method with forcing term σ_k, and the convergence proof can be derived as a special case from the proofs in [20,21,25], where the local superlinear rate is also proved.

Thus we may say that the convergence property of the proposed algorithm is based on its inexact feature. In the following section we explain the practical meaning of it.

3.1. Computing the direction

We propose an algorithm in which the solution of the Newton system (13) can be computed in an inexact way. In order to preserve the global convergence properties of the whole algorithm, the tolerance of such approximation has to be carefully determined. In particular, by defining $\boldsymbol{r}^k$ the residual vector of the Newton system in condensed form (14)–(15), i.e. $\boldsymbol{r}^k = -\nabla f(\boldsymbol{x}^k) + \rho_k X_k^{-1}\boldsymbol{e} - M_k\Delta\boldsymbol{x}^k$, we compute an approximation of the Newton direction such that

$$\|\boldsymbol{r}^k\| \leq \delta_k \varphi(\boldsymbol{x}^k, \boldsymbol{w}^k), \tag{25}$$

where $0 < \delta_k \leq \delta_{max} < 1$. The previous condition, together with (23)–(24), allows us to include our algorithm in the inexact Newton framework (see [20,24,21]). The condition (25) is crucial from a theoretical point of view, since it guarantees the global convergence of the algorithm, but it has also an important practical meaning. Indeed, it can represent the stopping criterion for an iterative solver applied to the system (14).

We point out that the stopping rule (25) is adaptive in the sense that it requires a coarser approximation of the Newton direction in the first interior point iterations, with a decreasing of the tolerance while the solution is approached. The main computational effort of the interior point methods is the solution of the Newton system in order to compute the direction. Most of the existing interior point codes are based on the direct factorization of the coefficient matrix. In the other cases, when an iterative strategy is applied, the fast convergence of the inner solver is due to a suitable preconditioner. But also in this case, the available codes provide the direct factorization of the preconditioner.

The image deblurring problems involve very large and dense matrices, which are often available only as an operator, thus the direct factorization of the matrix M_k can not be considered.

For the reasons above, we consider an iterative solver which requires only matrix–vector products, as the Preconditioned Conjugate Gradient (PCG).

This choice is motivated by observing that the system (14) has a symmetric positive definite matrix, thus the PCG method finds a solution in at most N steps. It is well-known that the preconditioning techniques can significantly improve the performances of the PCG method: in particular, a good preconditioner choice would be a matrix P_k such that $P_k^{-1}M_k$ has clustered eigenvalues.

Preconditioning the CG method in the image reconstruction frame is a very challenging issue and many approaches have been proposed in the literature: we mention for example the FFT-based preconditioners [26,27], the constraint preconditioners [28,29] and the FSIP (Factorized Sparse Inverse Preconditioners) approach [13].

The preconditioning technique implies an additional computational complexity to the conjugate gradient iteration that should be balanced by a very fast convergence of the iterative linear solver. Moreover, every preconditioner choice has to take into account of the special features of the coefficient matrix. We point out that, for the application we are dealing with, the matrix A is dense, very ill-conditioned or singular, and the entries of the matrices $F_k = BY_k^{-2}$ and $G_k = X_k^{-1}W_k$ can exhibit very large variations.

On the other hand, some of the approaches mentioned above are suited for problems where the blurring operator leads to a very sparse matrix A, while other techniques may become very expensive or ineffective due to the presence of the nonconstant diagonal matrices F_k and G_k.

Thus, in order to design a general and relatively cheap linear solver, our choice for the preconditioner is the exact diagonal of M_k, which can be computed as

$$(M_k)_{ii} = \sum_j A_{ji}^2 f_i + g_i$$

where f_i is the i-th diagonal entry of the matrices F_k and G_k respectively. The choice of the diagonal preconditioner may be non-optimal, but, with a small computational effort, we found that the number of PCG iterations needed to fulfill the rule (25) is relatively small, depending on the outer iterations. Indeed, at the first interior point iterations, often one or two PCG iterations are performed, while in proximity of the solution, up to an hundred PCG iterations are required. We recall that one CG inner iteration has an $\mathcal{O}(N \log N)$ computational complexity, due to the four FFTs needed to compute the hessian–vector product. The additional cost of the diagonal preconditioner consists in two further FFTs to compute the diagonal of M_k at

each outer interior point iteration and in N quotients at each CG iteration for solving the diagonal system. The leading cost of one interior point iteration is then $\mathcal{O}(N \log N)$.

3.2. The starting point

The choice of the starting point for the dual variables $\boldsymbol{w}^0$, given an initial guess $\boldsymbol{x}^0$, is a well known problem in the interior point methods [30,31]. The Newton direction can be drastically shortened by the positivity requirement $\boldsymbol{w} > 0$ and produce poor progress toward the solution. This behavior can be sustained for many iterations, making the solution process inefficient. In this section we present a heuristic that can produce a good dual starting point. We assume that an initial guess $\boldsymbol{x}^0$ is assigned to the primal variable and that a preliminary value $\tilde{\rho}$ is provided. We propose to set the initial dual variable $\boldsymbol{w}^0$ as

$$\begin{aligned} &\operatorname{argmin} \|\nabla f(\boldsymbol{x}^0) - \boldsymbol{w}\|^2 + \|X_0 \boldsymbol{w} - \tilde{\rho}\boldsymbol{e}\|^2, \\ &\boldsymbol{w} > \epsilon \end{aligned} \tag{26}$$

where ϵ is a positive tolerance. The solution of the quadratic programming problem (26) is the vector $\boldsymbol{w}^0$ that minimizes the KKT conditions of the problem (12), for $\boldsymbol{x} = \boldsymbol{x}^0$. The multiplier $\boldsymbol{w}^0$ can be obtained by projecting the vector $(I + X_0^2)^{-1}(\nabla f(\boldsymbol{x}^0) + \tilde{\rho}\boldsymbol{e})$ on the constraint $\boldsymbol{w} > \epsilon$.

However, the requirement (26) could be too strict, since it produces a multiplier that fits very well the initial guess $\boldsymbol{x}^0$. Thus, accordingly with the inexact nature of the algorithm, we introduce a residual $\boldsymbol{t}$, and we choose

$$\boldsymbol{w}^0 = \operatorname*{argmin}_{\boldsymbol{w} > \epsilon} \|\nabla f(\boldsymbol{x}^0) - \boldsymbol{w} + \boldsymbol{t}\|^2 + \|X_0 \boldsymbol{w} - \tilde{\rho}\boldsymbol{e}\|^2. \tag{27}$$

In this case, the solution $\boldsymbol{w}^0$ is given by the projection of $(I + X_0^2)^{-1}(\nabla f(\boldsymbol{x}^0) + \tilde{\rho}\boldsymbol{e} + \boldsymbol{t})$ on the constraint $\boldsymbol{w} > \epsilon$.

Algorithm 1 Inexact interior point method for the solution of (8)

Choose the starting point $(\boldsymbol{x}^0, \boldsymbol{w}^0) > 0$, set the parameters τ_1 and τ_2; compute the parameters $\delta_{max}, \sigma_{max}$; set the parameter $\theta < 1$ and set the tolerance τ.

FOR $i = 0, 1, 2, \ldots$ DO THE FOLLOWING STEPS:

STEP 1. IF $\varphi(\boldsymbol{x}^k, \boldsymbol{w}^k) < \tau$, THEN go to Step 8, ENDIF
STEP 2. Choose the perturbation parameter μ_k as in (24) and the forcing terms $0 < \delta_k < \delta_{max}$, $0 < \sigma_k < \sigma_{max}$.
STEP 3. Compute the primal direction $\Delta\boldsymbol{x}^k$ such that (25).
STEP 4. Compute the dual direction $\Delta\boldsymbol{w}^k$as in (15).
STEP 5. Determine the damping parameter α_k as in (17) and reduce it until the centrality conditions (18)–(19) are satisfied.
STEP 6. Backtracking loop:
    IF (21) holds, THEN go to Step 7;
    ELSE set $\alpha_k = \theta\alpha_k$ and go to Step 6.
    ENDIF
STEP 7. Compute the new point with the rule (16).
STEP 8. END

4. Regularization techniques

Since the problem (1) is an ill-posed problem, the solution of (3) can exhibit large instabilities corresponding to noise amplification, visible in the restored solution as a checkerboard effect. Thus, the solution of (3) may not be a reliable approximation of the solution of the restoration problem. Furthermore, we point out that any optimization method applied to (3), by definition, will tend to produce the noise-dependent solution of (3).

A widely used technique to preserve the restored image from the noise amplification consists of an early stopping of the iterations of the methods.

As an alternative, a regularized model is introduced, imposing some smoothness on the solution by penalizing the objective function in (3). A general form of the penalized–regularized problem is

$$\begin{aligned} &\min \quad J(\boldsymbol{x}) + \gamma R(\boldsymbol{x}), \\ &\boldsymbol{x} \geq 0 \end{aligned} \tag{28}$$

where γ is a regularization parameter and $R(\boldsymbol{x})$ is a regularizing functional.

Depending on the relative weight of the penalty versus the log-likelihood functional, expressed by the regularization parameter, the solution is pulled from the unstable solution of (3) toward the smoother solution of (28).

The regularization techniques can also be introduced from a probabilistic point of view, by giving an *a priori* probability to the image $\boldsymbol{x}$ and its measurement $\boldsymbol{b}$ (see [4,32,33,6] and reference therein). The estimate of $\boldsymbol{x}$ is then obtained by maximizing

the *a posteriori* probability

$$p(\boldsymbol{x}|\boldsymbol{b}) = \frac{p(\boldsymbol{b}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{b})},$$

or a modification of it. Taking the $-\log$ of the previous expression, introducing a multiplier γ and ignoring the constant term, we can define the following functional [33,6]

$$f(\boldsymbol{x}, \gamma) = -\log(p(\boldsymbol{b}|\boldsymbol{x})) - \gamma \log p(\boldsymbol{x}), \tag{29}$$

which is related to the maximum a posteriori probability problem. If we assume that the original image is the realization of independent Poisson processes at each pixel, with a known mean $\boldsymbol{p}$, then we have

$$p(\boldsymbol{x}) = \prod_i \frac{\mathrm{e}^{-p_i}(p_i)^{x_i}}{x_i!}$$

and

$$\begin{aligned} R(\boldsymbol{x}) \equiv -\log(p(\boldsymbol{x})) &= \sum_i p_i - x_i \log p_i + \log(x_i!) \\ &= \sum_i x_i \log\left(\frac{x_i}{p_i}\right) - (x_i - p_i). \end{aligned} \tag{30}$$

Hence, an *a priori* information about the solution $\boldsymbol{x}$ of the problem (1), the vector $\boldsymbol{p}$, is incorporated in the restoration model (28) (see [33,4]).

The functional (30) is also known as an *entropy* term and its gradient and hessian are given by

$$\nabla R(\boldsymbol{x}) = \log(\boldsymbol{x}) - \log(\boldsymbol{p})$$
$$\nabla^2 R(\boldsymbol{x}) = X^{-1}.$$

Hence, $R(\boldsymbol{x})$ is convex in its domain. Actually, it represents the Kullback–Leibler distance between the solution $\boldsymbol{x}$ and its a priori estimate $\boldsymbol{p}$, denoted by $KL(\boldsymbol{x}, \boldsymbol{p})$. Since this measure of the distance is not symmetric, we could consider also $KL(\boldsymbol{p}, \boldsymbol{x}) = \sum p_i \log(p_i/x_i) + x_i - p_i$ [34], or a combination of the two distances, $\beta KL(\boldsymbol{x}, \boldsymbol{p}) + (1-\beta)KL(\boldsymbol{p}, \boldsymbol{x})$ [32].

Other choices for $R(\boldsymbol{x})$ are considered for example in [32,35] and reference therein. For sake of completeness, we mention the well-known Tikhonov regularization term $R(\boldsymbol{x}) = \frac{1}{2}\|D\boldsymbol{x} - \boldsymbol{p}\|^2$, where D is a matrix.

In our study, we apply the interior point method to the problem (28) where $R(\boldsymbol{x})$ is given by (30), but the general form of the method would allow to consider other choices for the regularization and also for the objective functional, depending on the noise process assumed.

In the following section, we consider a variant of the LR algorithm which applies to the regularized problem.

4.1. Some gradient methods for the regularized problem

Simple gradient methods have been proposed also for the regularized case (28) [32,35]. The general form of these algorithms is related to the scaled steepest descent method

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \alpha_k X_k S_k^{-1} \nabla f(\boldsymbol{x}_k; \gamma) \tag{31}$$

with different choices of the scaling matrix S_k. In [32], the authors propose the following scaling matrix

$$S_k = diag(A^{\mathrm{T}} U_k A\boldsymbol{x}^k + [H^{\mathrm{T}} U_k \boldsymbol{b}]^- - \gamma \log \boldsymbol{p}), \tag{32}$$

where U_k is a weighting matrix. Here and in the following $[\cdot]^-$ indicates the negative part of a real number, i.e. $[z]^- = -z$ if $z < 0$ and $[z]^- = 0$ if $z \geq 0$. Analogously, we denote by $[\cdot]^+$ the positive part, so that, for any real number z we have $z = [z]^+ - [z]^-$.

Starting from a positive initial guess, if $U_k = Y_k^{-1}$, then the unitary step size guarantees the feasibility of the iterates, as shown in [32]. If we provide a line search rule, as for example the Armijo one, for determining the step length α_k, the scaled steepest-descent method converges to a non-negative solution of (28) [9].

Assuming that $A\boldsymbol{b}$ has positive entries and that the normalization $A^{\mathrm{T}}\boldsymbol{e} = \boldsymbol{e}$ holds, then the scaling (32) consists in $S_k = 1 - \gamma \log \boldsymbol{p}$. Thus, the method (31) with unitary stepsize reduces to the LR method (7) when $\gamma = 0$.

For other choices of the weighting matrix, as for example the choice $U_k = I$ proposed in [32, Section 6.1.1], the feasibility of the iterate (31) is not guaranteed in general: thus, some technique should be applied in order to preserve the positivity. We observe that, if a projection is performed and the new iterate is computed as follows

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha_k[\boldsymbol{x}^k - X_k S_k^{-1} \nabla f(\boldsymbol{x}_k; \gamma)]^+, \tag{33}$$

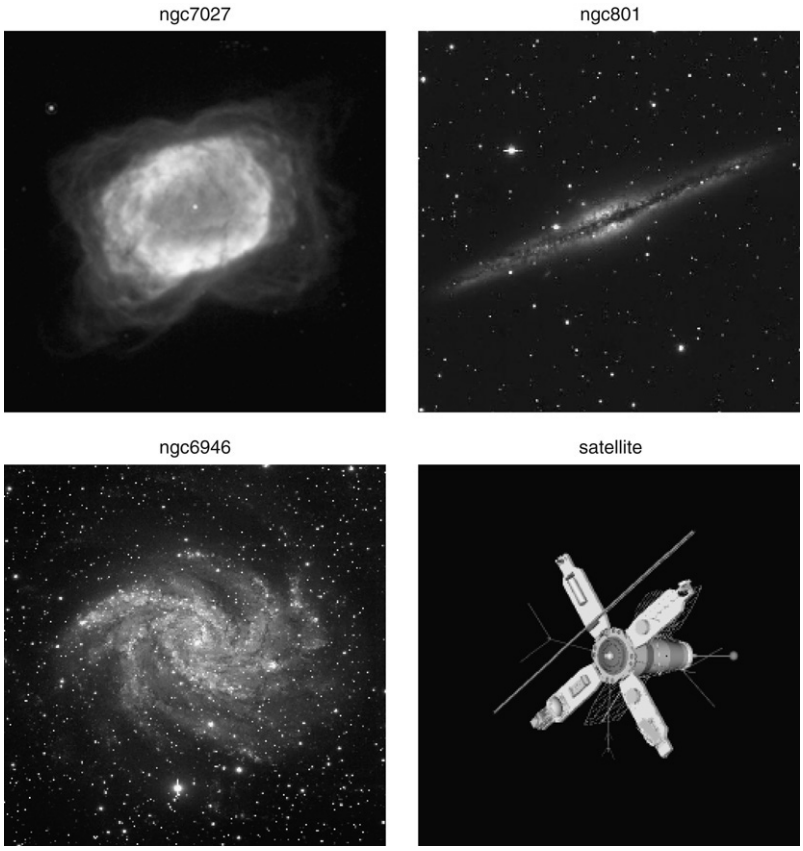

Fig. 1. Original images.

Table 1
Time comparison of the IP and LR algorithm.

	SNR (dB)	γ	IP			LR		
			ϵ_r	Time	Iter	ϵ_r	Time	Iter
ngc7027	48	0	0.0509821	33.7	13(399)	0.0512568	677	10 000
	33	8e−4	0.0735489	33.9	20(378)	0.0711466	70.76	600
ngc6946	48	0	0.205129	39.7	12(475)	0.239292	677	10 000
	33	8e−4	0.308028	9.9	8(76)	0.30939	70.98	600
ngc891	48	0	0.279646	18.3	8(194)	0.301251	677	10 000
	33	1e−4	0.327592	8.1	8(78)	0.338653	70.7	600
satellite	48	3e−4	0.259738	59.8	21(714)	0.26838	140.7	1 200
	33	8e−4	0.289965	25.1	14(279)	0.292606	141.0	1 200

then the resulting iteration is a special case of the scaled projection methods presented in [9, Sec. 2.3]. Hence, under the assumption that

$$c_1\|\boldsymbol{z}\|^2 \leq \boldsymbol{z}^{\mathrm{T}} X_k S_k^{-1} \boldsymbol{z} \leq c_2\|\boldsymbol{z}\|^2 \quad \forall \boldsymbol{z} \in \mathbb{R}^n,$$

for some positive scalars $c_1 c_2$, the method (33) converges to a stationary point of the problem (28), if α_k is chosen by the Armijo rule (see Proposition 2.3.4 in [9]). The methods (31) and its projected version (33) can be extended to other choices of the regularization functionals, as in [32,35].

5. Numerical experiments

The aim of our experimentation is to investigate if the method explained in Section 3 could be of practical interest for the image restoration problem and if it is a real improvement compared to a gradient method like the LR one. For this reason we compare the performance of the interior point algorithm with an existing and widely-used method, the LR method. More precisely, we consider the two cases, with and without the regularization term described in Section 4: in the second case, the comparison is between the classical LR method (7), while we choose the method (31)–(32) as a benchmark for the

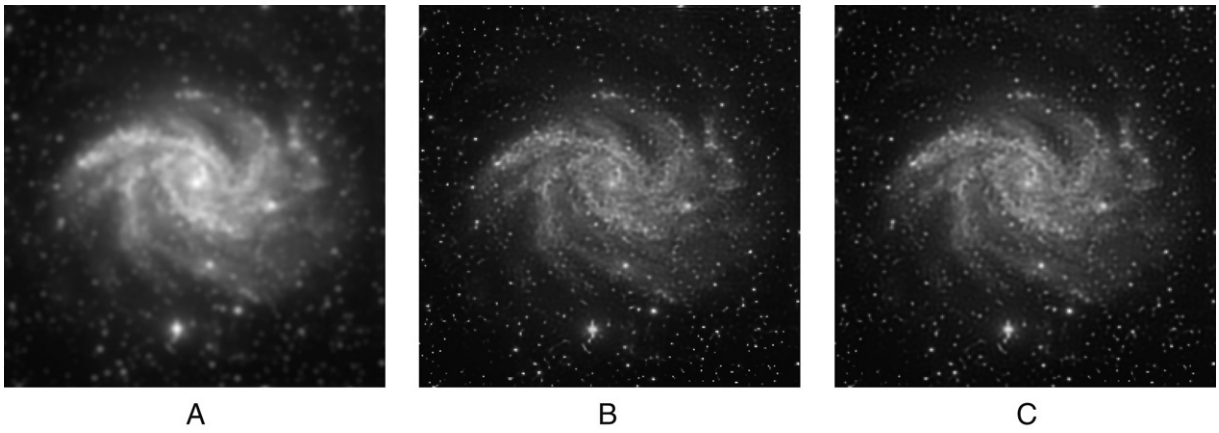

Fig. 2. Reconstruction of the test image ncg6946, SNR = 48. (A) blurred and noisy image; (B) IP reconstruction; (C) LR reconstruction.

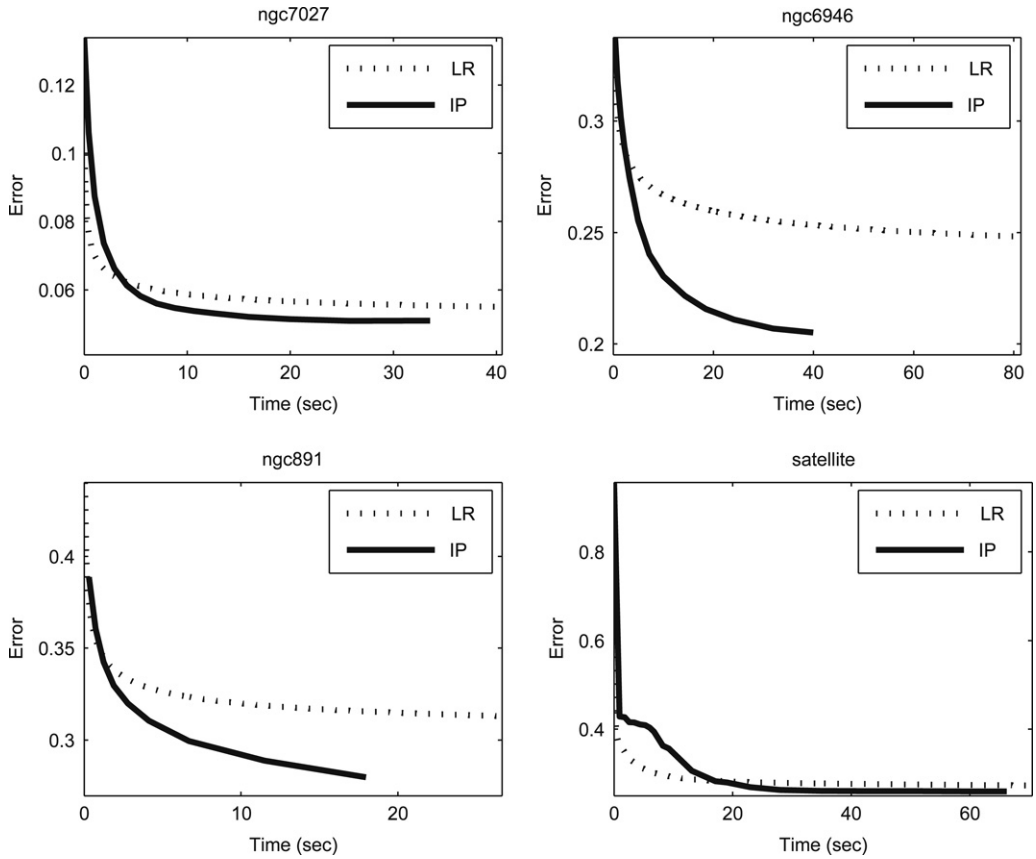

Fig. 3. Relative error with respect to the execution time: case SNR = 48.

regularized case. By doing so, the two methods applied to the same optimization problem, (3) or (28) respectively, can be compared.

The IP and LR algorithms, including the regularized versions, have been implemented in the Matlab environment, and run on an Intel Pentium M715 processor, 1.5 GHz, 512 MB DDR.

5.1. The test problems

We considered the images shown in Fig. 1 whose size is 256 × 256. All the images reported in this paper have been obtained with the Matlab 'imshow' function, thus the scale can vary. Images ngc6946 and ngc891 can be downloaded from

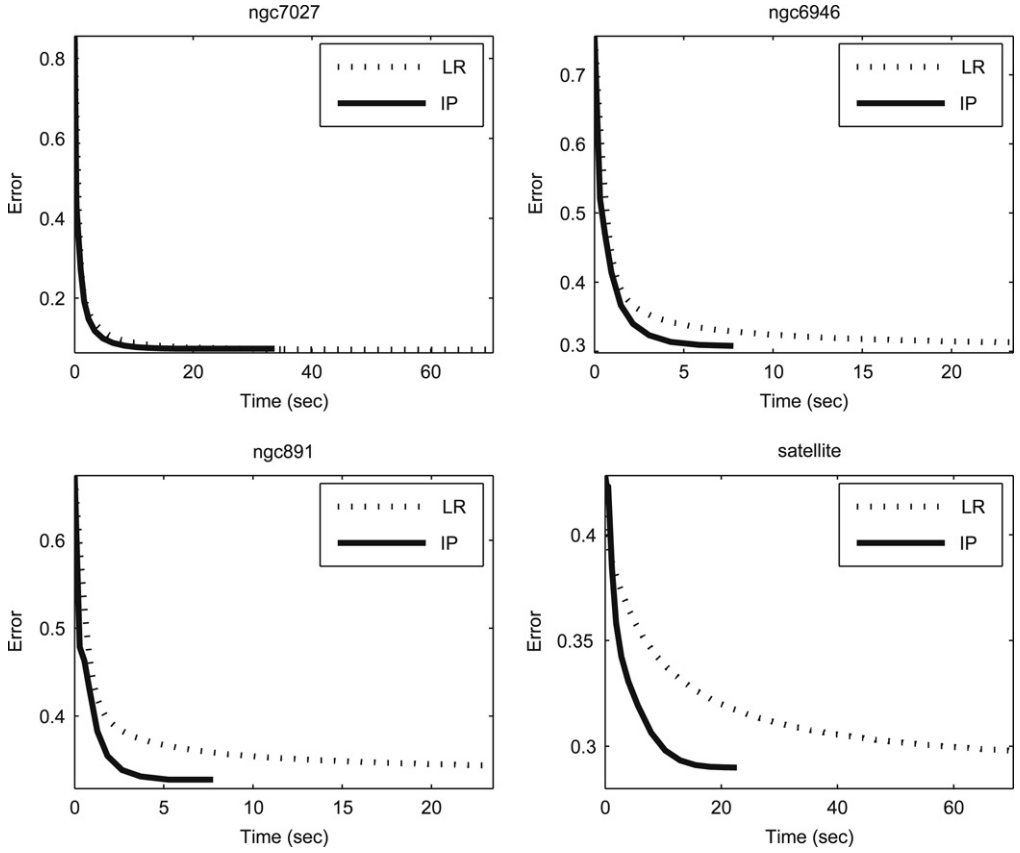

Fig. 4. Relative error with respect to the execution time: case SNR = 33.

http://medusa.as.arizona.edu/lbto/astronomical.htm. The images were cut and resized to a 256×256 array. In particular, for the image ngc6946 only the green channel has been selected.

The original images were artificially blurred by numerical convolution with a point-spread function which simulates a telescope [36]. It leads to a $n \times n$ matrix A with circular symmetry with respect its center, and the values along the radius are computed with

$$2\left[\frac{J_1(\pi\theta D/\lambda)}{\pi\theta D/\lambda}\right]^2,$$

where J_1 is the Bessel J function, D is 8.25, λ is 2.2×10^{-6}, for $\theta \in (0, 16.4]$. We approximate the value of the point-spread function for $\theta = 0$ with 0.5. The matrix obtained has been then normalized so that $A^{\mathrm{T}}\boldsymbol{e} = \boldsymbol{e}$.

The blurred images $\boldsymbol{g}$ are computed by numerical convolution of the true image with the point-spread function described above. They are successively corrupted by Poisson noise by means of the 'imnoise' Matlab function, yielding the simulated observed images $\boldsymbol{b}$ (Fig. 1). For each test image we consider two different noise levels which are listed in Table 1. The background parameter is $bg = 6.76 \times 10^{-9}$.

As index of the quality of the restored image we consider the relative error ϵ_r of the reconstructed image $\hat{x}$ with respect to the exact solution x

$$\epsilon_r = \frac{\|\hat{x} - x\|}{\|x\|}.$$

The settings of the methods are listed below.

5.2. Implementation details

The parameters $\delta_{max}, \sigma_{max}, \tau_1$ and τ_2 are computed at the first iteration as in [20]. The reduction factor for the backtracking loop is $\theta = 0.8$ and the parameter β in the line search rule (21) is set to 10^{-4}.

For the experiments, the monotonicity degree M has been set to 10 and the forcing term δ_k related to the inner stopping criterion has been fixed equal to δ_{max}. The other forcing term σ_k has been chosen as $\sigma_k = \min(0.5, \sigma_{max})$.

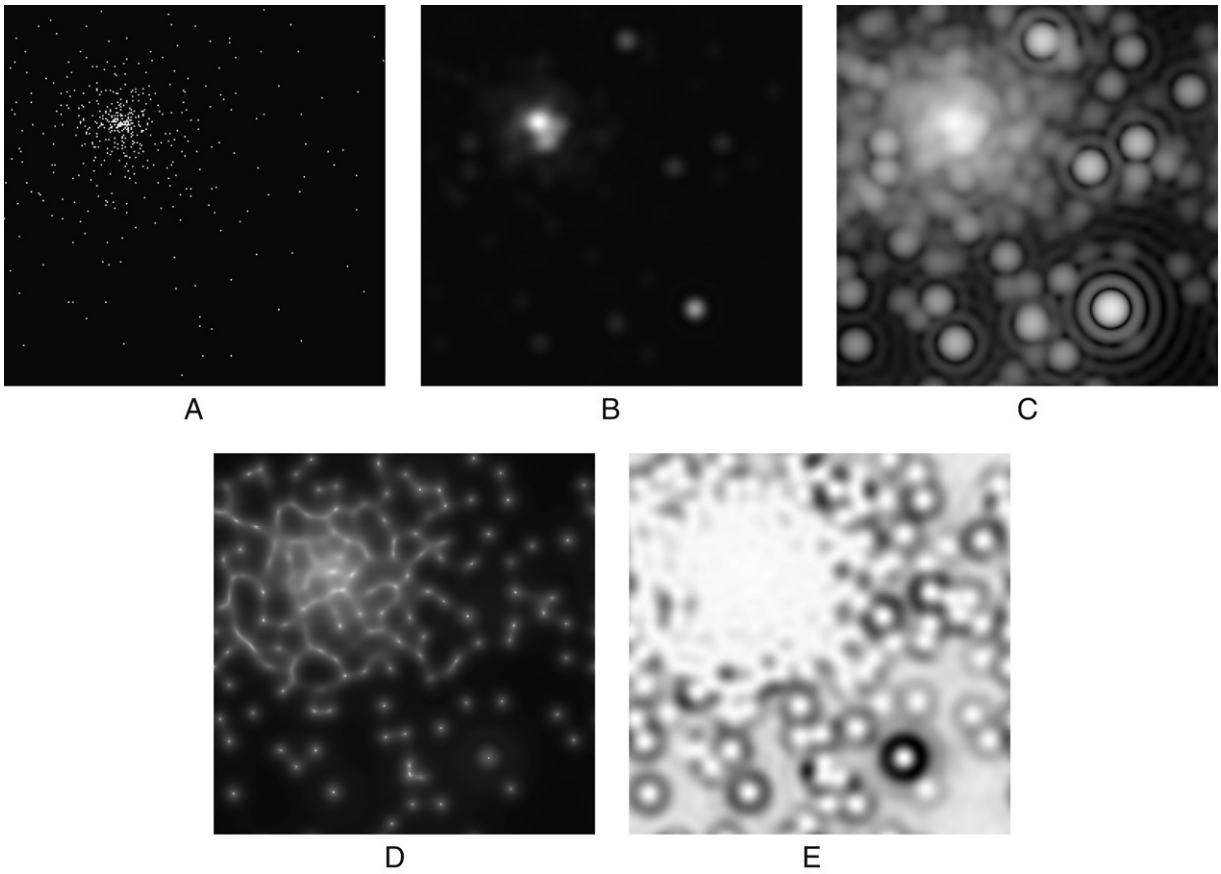

Fig. 5. Test image of a star cluster. (A) Original image in logarithmic scale, (B) noisy blurred in linear scale; (C) noisy blurred in logarithmic scale (SNR = 52.1979); (D) IP reconstructed in logarithmic scale; (E) LR reconstructed in logarithmic scale.

The stopping criterion of the whole interior point method has been set with the tolerance $\tau = 5 \times 10^{-3}$: this is not a very strict tolerance, but it takes into account of the ill-conditioning of the problem.

Furthermore, we limit the maximum number of PCG iterations to 100. The parameters for the initialization of the Lagrange multiplier in (27) are $\tilde{\rho} = 0.5$, $\boldsymbol{t} = 10^{-1}$ and $\epsilon = 10^{-4}$.

For the regularized case (28) we consider the variant (31)–(32) of the LR algorithm with $U_k = Y_k^{-1}$. In fact, we observed that the other choice $U_k = I$ produces a poorly-scaled direction, requiring many step-length reductions for the fulfillment of the Armijo rule.

The regularization parameter γ has been empirically determined, since it is out of the scope of this paper to find the optimal one. The a priori information in (30) has been chosen as $\boldsymbol{p} = \boldsymbol{b}$.

5.3. Numerical results

In Table 1 we report a comparison between the proposed IP and the LR methods, in terms of the relative error, iteration number and execution time on the test images in Fig. 1. For the IP method we report the number of external iterations and, in brackets, the total number of PCG iterations. For a visual comparison, we report also the noisy blurred and restored images for the case ngc6946 in Fig. 2.

After obtaining the IP approximate solution, we performed a number of LR iterations such that the relative error of the LR reconstruction was about the same, stopping the method anyway after 10 000 iterations or when no significant improvements are observed.

The comparison of the two algorithms is plotted also in Figs. 3 and 4, where the relative error is reported with respect to the execution time.

We considered also another type of astronomical image, shown in Fig. 5: this image is downloadable from ftp://ftp.stsci.edu/software/stsdas/testdata/restore/sims/star_cluster/ and it consists in 470 light sources. For this experiment we stopped the IP iterations when $\varphi(\boldsymbol{x}^k, \boldsymbol{w}^k) < 0.06$, achieving the relative residual $\epsilon_r = 0.5897$, after 36 outer and 1620 inner iterations, performed in 136 s. On the same test problem the LR algorithm needed 8000 iteration to

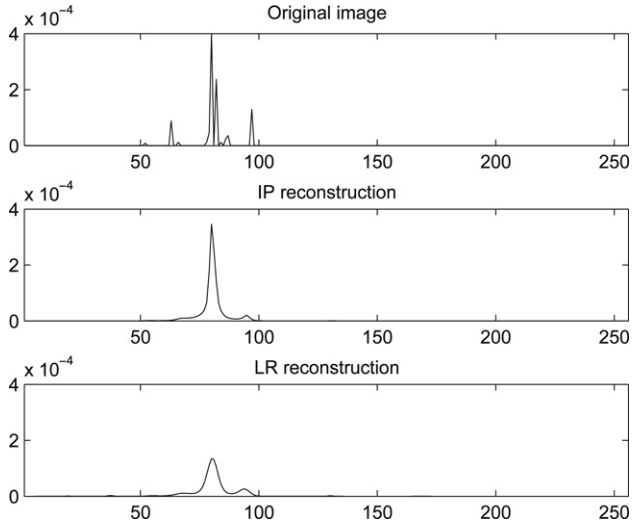

Fig. 6. Star cluster image, row 80:original and reconstructed images.

reach $\epsilon_r = 0.6146$ in 482 s. Since the image is very sparse, the relative error may not be a significant index of the reliability of the reconstruction. Thus, in Fig. 6, we report the row 80 of the original and reconstructed images.

We observe that the IP reconstruction detects better the intensity of the peaks and produces a sharper image. The same observation can be made also for the other rows. In order to depict the difference of the two reconstructions, we report also the IP and LR reconstructions in logarithmic scale.

6. Conclusions

We proposed an inexact interior point algorithm for deblurring noisy images. Also a regularized version of the algorithm is presented, which can be generalized to different kinds of noise processes. We compared our approach with the Lucy Richardson method, which is a reference method for image deblurring. We framed the analysis of the methods into an optimization point of view. The results show that the interior point approach can significantly improve the execution times needed for obtaining the same reconstruction image quality, as the reference method. We also presented some cases where the IP method was able to obtain a better relative error than the smallest Lucy Richardson's one.

As a final remark, the paper proved that the superlinear convergence properties of the IP method are a real improvement over the gradient methods and are of both theoretical and practical interest.

Acknowledgements

We are extremely grateful to professor Mario Bertero and professor Patrizia Boccacci of the University of Genova for fruitful discussions and for the image ngc7027 in Fig. 1. This work is supported by the PRIN2006 Project *Problemi inversi in Medicina e Astronomia*.

References

[1] M. Bertero, P. Boccacci, Introduction to Inverse Problems, Institute of Physics Publishing, Bristol, 1998.
[2] L.B. Lucy, An iterative technique for the rectification of observed distributions, Astronomical Journal 79 (1974) 745–754.
[3] W.H. Richardson, Bayesian–based iterative method of image restoration, Journal of the Optical Society of America A 62 (1972) 55–59.
[4] J. Llacer, J. Nuñez, in: R.L. White, R.J. Allen (Eds.), Iterative Maximum Likelihood and Bayesian Algorithms for Image Reconstruction in Astronomy, Space Telescope Science Insitute, Baltimore, MD, 1990, pp. 62–69.
[5] L.A. Shepp, Y. Vardi, Maximum likelihood reconstruction for emission tomography, IEEE Transaction on Medical Imaging 1 (1982) 113–122.
[6] R. Molina, J. Nuñez, F. Cortijo, J. Mateos, Image restoration in astronomy: A bayesian perspective, Signal Processing Magazine, IEEE 18 (2) (2001) 11–29.
[7] Y. Vardi, L. Shepp, L. Kaufman, A statistical model for positron emission tomography, Journal of the American Statistical Society 80 (389) (1985) 8–37.
[8] K. Kaufman, Implementing and accelerating the EM algorithm for positron emission tomography, IEEE Transaction on Medical Imaging 6 (1987) 37–51.
[9] D.P. Bertsekas, Nonlinear Progrmming, 2nd ed., Athena Scientific, 1999.
[10] M. Rojas, T. Steihaug, Large-scale optimization techniques for nonnegative image restorations, in: Advanced Signal Processing Algorithms, Architectures, and Implementations, XII SPIE Conference, No. 4791 2002, pp. 233–242.
[11] J. Bardsley, C. Vogel, A nonnegatively constrained convex programming method for image restoration, SIAM Journal on Scientific Computing 25 (2003) 1326–1343.
[12] J. Bardsley, J. Nagy, Covariance–preconditioned iterative methods for nonnegativaly constrained astronomical imaging, SIAM Journal on Matrix Analysis and Applications 27 (2006) 1184–1197.
[13] H. Fu, M. Ng, M. Nikolova, J. Barlow, Efficient minimization methods of mixed L1–L1 and L2–L1 norms for image restoration, SIAM Journal on Scientific Computing 27 (2006) 1881–1902.

[14] M. Hanke, J.G. Nagy, C. Vogel, Quasi-Newton approach to nonnegative image restorations, Linear Algebra and its Applications 316 (1–3) (2000) 223–236.
[15] I. Csiszär, Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems, Annals of Statistics 19 (1991) 2032–2066.
[16] B. Anconelli, M. Bertero, P. Boccacci, M. Carbillet, H. Lanteri, Reduction of boundary effects in multiple image deconvolution with an application to LBT LINC–NIRVANA, Astronomy & Astrophysics 448 (2006) 1217–1224.
[17] M. Bertero, P. Boccacci, A simple method for the reduction of boundary effects in the Lucy–Richardson approach to image deconvolution, Astronomy & Astrophysics 437 (2005) 369–374.
[18] P. Davis, Circulant Matrices, John Wiley & Sons, 1979.
[19] A.S. El-Bakry, R.A. Tapia, T. Tsuchiya, Y. Zhang, On the formulation and theory of Newton interior–point method for nonlinear programming, Journal of Optimization Theory and Applications 89 (1996) 507–541.
[20] S. Bonettini, E. Galligani, V. Ruggiero, Inner solvers for interior point methods for large scale nonlinear programming, Computational Optimization and Applications 37 (1) (2007) 1–34.
[21] S. Bonettini, A nonmonotone inexact Newton method, Optimization Methods and Software 20 (4–5) (2005) 475–491.
[22] S.C. Eisenstat, H.F. Walker, Globally convergent inexact Newton methods, SIAM Journal on Optimization 4 (1994) 393–422.
[23] T. Serafini, G. Zanghirati, L. Zanni, Gradient projection methods for quadratic programs and applications in training support vector machines, Optimization Methods and Software 20 (2–3) (2005) 343–378.
[24] C. Durazzi, On the Newton interior–point method for nonlinear programming problems, Journal of Optimization Theory and Applications 104 (2000) 73–90.
[25] S. Bonettini, E. Galligani, V. Ruggiero, An inexact newton method combined with hestenes multipliers' scheme for the solution of the karush-kuhn-tucker systems, Applied Mathematics and Computation 168 (2005) 651–676.
[26] R.H. Chan, M.K. Ng, Conjugate gradient methods for Toeplitz systems, SIAM Review 38 (3) (1996) 427–482.
[27] M.K. Ng, Nonlinear image restoration using FFT–based conjugate gradient methods, in: Proceedings of the 1995 International Conference on Image Processing, vol. 2, 1995, pp. 2041–2044.
[28] M. Benzi, M. Ng, Preconditioned iterative methods for weighted Toeplitz least squares problems, SIAM Journal on Matrix Analysis and Applications 27 (4) (2006) 1106–1124.
[29] L. Lukšan, C. Matonoha, J. Vlček, Interior–point method for nonlinear nonconvex optimization, Numerical Linear Algebra with Applications 11 (2004) 431–453.
[30] M. Gertz, J. Nocedal, A. Sartenaer, A starting-point strategy for nonlinear interior methods, Applied Mathematics Letters 17 (2004) 945–953.
[31] E.A. Yildirim, S.J. Wright, Warm start strategies in interior-point methods for linear programming, SIAM Journal on Optimization 12 (2002) 782–810.
[32] H. Lanteri, M. Roche, C. Aime, Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms, Inverse Problems 18 (2002) 1397–1419.
[33] A.S. Carasso, Linear and nonlinear image deblurring: A documented study, SIAM Journal on Numerical Analysis 36 (6) (1999) 1659–1689.
[34] C.L. Byrne, J. Graham-Eagle, Iterative image reconstruction algorithms based on cross-entropy minimization, in: M.A. Fiddy (Ed.), in: Proc. SPIE, Inverse Problems in Scattering and Imaging, vol. 1767, 1992, pp. 83–92.
[35] B. Anconelli, M. Bertero, P. Boccacci, M. Carbillet, H. Lanteri, Iterative methods for the reconstruction of astronomical images with high dynamic range, Journal of Computational and Applied Mathematics 198 (2007) 231–331.
[36] M. Carbillet, S. Correia, P. Boccacci, M. Bertero, Restoration of interferpmetric images–The case–study of the Large Binocular Telescope, Astronomy & Astrophysics 387 (2002) 744–757.