Accelerated gradient methods for constrained image deblurring

Journal of Physics: Conference Series **135** (2008) 012022 doi:10.1088/1742-6596/135/1/012022

Accelerated gradient methods for constrained image deblurring

S Bonettini[1], R Zanella[2], L Zanni[2], M Bertero[3]

[1] Dipartimento di Matematica, Università di Ferrara, Via Saragat 1, Building B, I-44100 Ferrara, Italy

[2] Dipartimento di Matematica Pura e Applicata, Università di Modena e Reggio Emilia, Via G. Campi 213/B, I-41100 Modena, Italy

[3] Dipartimento di Informatica e Scienze dell'Informazione, Università di Genova, Via Dodecaneso 35, I-16145 Genova, Italy

E-mail: `silvia.bonettini@unife.it`

Abstract. In this paper we propose a special gradient projection method for the image deblurring problem, in the framework of the maximum likelihood approach. We present the method in a very general form and we give convergence results under standard assumptions. Then we consider the deblurring problem and the generality of the proposed algorithm allows us to add a energy conservation constraint to the maximum likelihood problem. In order to improve the convergence rate, we devise appropriate scaling strategies and steplength updating rules, especially designed for this application. The effectiveness of the method is evaluated by means of a computational study on astronomical images corrupted by Poisson noise. Comparisons with standard methods for image restoration, such as the expectation maximization algorithm, are also reported.

1. Introduction

Nonlinear image deblurring procedures based on the maximum likelihood criterion have been widely investigated in literature. This approach leads to model the deblurring problem as a constrained large scale optimization problem, with a nonlinear, convex objective function. In particular, in the case of Poisson noise, an approximation of the original image can be obtained by computing a vector $\boldsymbol{x} \in \mathbb{R}^n$ that solves the minimization problem

$$\begin{aligned} \min \quad & J(\boldsymbol{x}) \\ \text{sub. to} \quad & \boldsymbol{x} \geq 0, \end{aligned} \tag{1}$$

in which the functional $J(\boldsymbol{x})$ is the Kullback–Leibler divergence of $(A\boldsymbol{x} + bg)$ from the observed noisy image $\boldsymbol{b}$, where $A \in \mathbb{R}^{n\times n}$ is the blurring operator, satisfying the conditions $A_{ij} \geq 0$, $\sum_{j=1}^n A_{ij} > 0 \; \forall\, i$, $\sum_{i=1}^n A_{ij} = 1 \; \forall\, j$, and bg denotes a constant background term. In details, the functional J is defined as

$$J(\boldsymbol{x}) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} A_{ij} x_j + bg - b_i - b_i \ln \frac{\sum_{j=1}^n A_{ij} x_j + bg}{b_i} \right). \tag{2}$$

One of the most popular algorithms for astronomical and medical image restoration problems is the Expectation–Maximization (EM) [15] or Richardson–Lucy method [12],[14], which applies to the problem (1) and consists in the iteration rule

$$\boldsymbol{x}^{(k+1)} = X_k A^T Y_k^{-1} \boldsymbol{b}, \tag{3}$$

where $X_k = \mathrm{diag}(\boldsymbol{x}^{(k)})$ and $Y_k = \mathrm{diag}(A\boldsymbol{x}^{(k)} + bg)$. The EM method is attractive because of its convergence properties and its low computational cost per iteration, but it exhibits a very slow convergence rate, that, in many cases, leads to thousands of iterations to obtain a satisfactory approximation of the solution. For this reason, some modified EM versions have been proposed in order to improve the convergence rate of the method (3), as for example the accelerated version introduced in [4], that is also implemented in the MATLAB *deconvlucy* function (EM_MATLAB) of the *Image Processing* toolbox. However, as far as we know, no convergence proof of this accelerated algorithm is available.
In this study we propose a different approach that allows us to design efficient acceleration strategies provided by a rigorous theoretical foundation. We introduce our algorithm by observing that, since

$$\nabla J(\boldsymbol{x}^{(k)}) = A^T \boldsymbol{e} - A^T Y_k^{-1} \boldsymbol{b} = \boldsymbol{e} - A^T Y_k^{-1} \boldsymbol{b},$$

where $\boldsymbol{e} \in \mathbb{R}^n$ has all the entries equal to one, then the method (3) can be considered a special case of a more general scaled steepest descent method

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k D_k \nabla J(\boldsymbol{x}^{(k)}), \tag{4}$$

with $D_k = X_k$ and $\alpha_k = 1$ for any k. Adopting this point of view, we can try to improve the convergence rate of the scheme (4) by applying suitable choices of the parameter α_k, derived by the recent studies on the steplength selections in gradient methods [8, 9, 10, 16]. This idea is implemented within a scaled gradient projection (SGP) algorithm, combining the iteration (4) with a projection on the constraints and with a nonmonotone line–search strategy [5, 7]. The algorithm is especially designed for the deconvolution problem, but it applies to any minimization problem of the form

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{sub. to} \quad & \boldsymbol{x} \in \Omega, \end{aligned} \tag{5}$$

where $\Omega \subset \mathbb{R}^n$ is a closed convex set and f is a continuously differentiable function. We report a convergence result for the proposed algorithm and give crucial details for its practical implementation. The generality of the SGP makes the method very well suited to face, as a special case of (5), the deconvolution problem

$$\begin{aligned} \min \quad & J(\boldsymbol{x}) \\ \text{sub. to} \quad & \boldsymbol{x} \geq 0, \quad \textstyle\sum_{i=1}^n x_i = c, \end{aligned} \tag{6}$$

where the equality constraint represents the energy conservation principle. Thus, the SGP can be regarded as a solver for deblurring problems and evaluated in comparison with the standard EM method and the EM_MATLAB algorithm. To this end, a computational study on a set of astronomical images corrupted by Poisson noise is carried out in the Matlab environment. The numerical results we have obtained show that SGP provides a remarkable computational gain with respect to the EM method and it is competitive with the accelerated algorithm used by Matlab.

The paper is organized as follows. In section 2 the scaled gradient projection approach is presented and global convergence results are reported, while in section 3 the implementation of the method is discussed; the numerical results are described and evaluated in section 4 and, finally, some conclusions are drawn in section 5.

Algorithm 1 Scaled Gradient Projection (SGP) Method

Choose the starting point $\boldsymbol{x}^{(0)} \in \Omega$, set the parameters $\beta, \theta \in (0,1)$, $0 < \alpha_{min} < \alpha_{max}$ and fix a positive integer M.
FOR $k = 0, 1, 2, ...$ DO THE FOLLOWING STEPS:

STEP 1. Choose the parameter $\alpha_k \in [\alpha_{min}, \alpha_{max}]$ and the scaling matrix $D_k \in \mathcal{D}$;
STEP 2. Projection: $\boldsymbol{y}^{(k)} = P_{\Omega, D_k^{-1}}(\boldsymbol{x}^{(k)} - \alpha_k D_k \nabla f(\boldsymbol{x}^{(k)}))$;
STEP 3. Descent direction: $\Delta\boldsymbol{x}^{(k)} = \boldsymbol{y}^{(k)} - \boldsymbol{x}^{(k)}$;
STEP 4. Set $\lambda_k = 1$ and $f_{max} = \max_{0 \le j \le \min(k, M-1)} f(\boldsymbol{x}^{(k-j)})$;
STEP 5. Backtracking loop:
IF $f(\boldsymbol{x}^{(k)} + \lambda_k \Delta\boldsymbol{x}^{(k)}) \le f_{max} + \beta\lambda_k \nabla f(\boldsymbol{x}^{(k)})^T \Delta\boldsymbol{x}^{(k)}$ THEN
go to Step 6;
ELSE
Set $\lambda_k = \theta\lambda_k$ and go to Step 5.
ENDIF
STEP 6. Set $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda_k \Delta\boldsymbol{x}^{(k)}$.

END

2. A scaled gradient projection method

In this section we describe the SGP scheme for the solution of the problem (5) and we report the main convergence results. To this end we recall some basic definitions and notations.
Here and in the following, $\|\cdot\|$ denotes the 2-norm of vectors and matrices, while $\|\cdot\|_D$ indicates the vector norm induced by the symmetric positive definite matrix $D \in \mathbb{R}^{n \times n}$ defined as $\|\boldsymbol{x}\|_D = \sqrt{\boldsymbol{x}^T D \boldsymbol{x}}$. Accordingly, given a closed convex set $\Omega \subset \mathbb{R}^n$, we can define the projection operator associated to D as

$$P_{\Omega, D}(\boldsymbol{x}) = \text{argmin}_{\boldsymbol{y} \in \Omega} \|\boldsymbol{y} - \boldsymbol{x}\|_D. \tag{7}$$

The main SGP steps are given in Algorithm 1.
The only assumptions requested by SGP are that, at each iteration, the parameter α_k is bounded above and below by positive constants and that the scaling matrix D_k belongs to the set $\mathcal{D}$ of the $n \times n$ symmetric positive definite matrices D such that $\|D\| \le L$ and $\|D^{-1}\| \le L$, for a given threshold $L > 1$. At each SGP iteration the vector

$$\boldsymbol{y}^{(k)} = P_{\Omega, D_k^{-1}}(\boldsymbol{x}^{(k)} - \alpha_k D_k \nabla f(\boldsymbol{x}^{(k)}))$$

is defined by combining a scaled steepest descent direction of the same form used in (4) with a projection on Ω. Since the projection is performed by using the projection operator associated to the inverse of the scaling matrix, it is possible to prove that the resulting search direction $\Delta\boldsymbol{x}^{(k)} = \boldsymbol{y}^{(k)} - \boldsymbol{x}^{(k)}$ is a descent direction for the problem (5), that is ${\Delta\boldsymbol{x}^{(k)}}^T \nabla f(\boldsymbol{x}^{(k)}) < 0$ [6, 7]. Thus, we can ensure the global convergence of the whole algorithm computing the new point $\boldsymbol{x}^{(k+1)}$ along $\Delta\boldsymbol{x}^{(k)}$ by means of a nonmonotone line–search procedure, derived by a generalization of the Armijo rule.
A convergence result for SGP can be stated as follows.

THEOREM 2.1 *Let $\{\boldsymbol{x}^{(k)}\}$ be the sequence generated by applying the SGP algorithm to the problem (5) and assume that the level set $\Omega_0 = \{\boldsymbol{x} \in \Omega : f(\boldsymbol{x}) \le f(\boldsymbol{x}^{(0)})\}$ is bounded. Then, every accumulation point $\boldsymbol{x}^* \in \Omega$ of the sequence $\{\boldsymbol{x}^{(k)}\}$ is a constrained stationary point, that is*

$$\nabla f(\boldsymbol{x}^*)^T (\boldsymbol{x} - \boldsymbol{x}^*) \ge 0, \quad \forall\, \boldsymbol{x} \in \Omega.$$

The Theorem 2.1 can be proved by employing standard techniques (see, for example, [6, 7]), and it applies to every constrained problem of the form (5). It is worth to stress that any choice of the steplength α_k in a bounded interval and of the scaling matrix D_k in the set $\mathcal{D}$ are allowed; then, this freedom of choice can be fruitfully exploited for introducing performance improvements.
The special case (6) is a convex problem and, consequently, every stationary point is a minimum point [3, p.194]; furthermore, its feasible region is bounded. Thus, from Theorem 2.1 a stronger convergence result can be formulated.

COROLLARY 2.1 *Let $\boldsymbol{x}^{(k)}$ be the sequence generated by applying the SGP algorithm to the problem* (6)*. Then every accumulation point of the sequence $\{\boldsymbol{x}^{(k)}\}$ is a solution of* (6)*.*

Several reasons make the SGP approach appealing for solving problem (6). First of all it is a very simple and general method. Secondly, due to the special features of the constraints, an appropriate choice of D_k can make the projection operation in step 2. computationally non–expensive. Finally, the iterative scheme can achieve good convergence rate by exploiting effective steplength selection rules for updating α_k. These issues are better explained in the next section.

3. Applying the method for image deblurring

We describe an SGP implementation for solving the image deblurring problem (6). To this end we must show how the projection on the feasible region of (6) is performed and how the scaling matrix D_k and the steplength parameter α_k are updated.

Let $\bar{\Omega}$ be the feasible region of the problem (6): $\bar{\Omega} = \{\boldsymbol{x} \in \mathbb{R}^n \quad | \quad \boldsymbol{x} \geq 0, \ \sum_{i=1}^n x_i = c\}$. When algorithm SGP is applied for solving (6), at each iteration we need to compute $P_{\bar{\Omega},D_k^{-1}}(\boldsymbol{x}^{(k)} - \alpha_k D_k \nabla f(\boldsymbol{x}^{(k)}))$, that is, we must solve the constrained strictly convex quadratic program

$$\begin{aligned} \min \quad & \tfrac{1}{2}\boldsymbol{x}^T D_k^{-1}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{z} \\ \text{sub. to} \quad & \textstyle\sum_{i=1}^N x_i - c = 0, \qquad x_i \geq 0, \quad i = 1, \ldots, N, \end{aligned} \tag{8}$$

where $\boldsymbol{z} = D_k^{-1}\left(\boldsymbol{x}^{(k)} - \alpha_k D_k \nabla f(\boldsymbol{x}^{(k)})\right)$.
From now on we will assume that the scaling matrix D_k is diagonal, $D_k = \operatorname{diag}\left(d_1^{(k)}, \ldots, d_n^{(k)}\right)$, as usually done in scaled gradient projection methods. Under this assumption and due to the special structure of the constraints, the problem (8) can be reformulated as a one-dimensional root-finding problem. In fact, if $\bar{\boldsymbol{x}}$ denotes the solution of (8), then from the KKT first order optimality conditions we know that there exist Lagrange multipliers $\bar{\lambda} \in \mathbb{R}$ and $\bar{\boldsymbol{\alpha}} \in \mathbb{R}^n$ such that

$$D_k^{-1}\bar{\boldsymbol{x}} - \boldsymbol{z} - \bar{\lambda}\boldsymbol{e} - \bar{\boldsymbol{\alpha}} = \boldsymbol{0}, \qquad \bar{\boldsymbol{x}} \geq \boldsymbol{0}, \qquad \bar{\boldsymbol{\alpha}} \geq \boldsymbol{0}, \qquad \bar{\boldsymbol{\alpha}}^T\bar{\boldsymbol{x}} = 0, \qquad \sum_{i=1}^n \bar{x}_i - c = 0.$$

From the first four KKT conditions it is easy to obtain $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{\alpha}}$ as functions of $\bar{\lambda}$:

$$\bar{x}_i(\bar{\lambda}) = \max\left\{0, d_i^{(k)}(z_i + \bar{\lambda})\right\}, \qquad \bar{\alpha}_i(\bar{\lambda}) = \max\left\{0, -(z_i + \bar{\lambda})\right\}, \qquad i = 1, \ldots, n.$$

Thus, in order to solve the KKT system, we must find $\bar{\lambda}$ such that

$$\sum_{i=1}^n \bar{x}_i(\bar{\lambda}) - c = 0, \qquad \text{where} \qquad \bar{x}_i(\bar{\lambda}) = \max\left\{0, d_i^{(k)}(z_i + \bar{\lambda})\right\}. \tag{9}$$

This means that the computation of the projection $P_{\bar{\Omega},D_k^{-1}}(\boldsymbol{x}^{(k)} - \alpha_k D_k \nabla f(\boldsymbol{x}^{(k)}))$ essentially reduces to solve a root-finding problem for a piecewise linear monotonically non-decreasing

function. Efficient linear time algorithms for this problem are available [8, 11, 13]; the SGP implementation we propose exploits the secant-based method suggested in [8].

The choice of the scaling matrix D_k in gradient projection methods must usually aim at two main goals: avoiding to introduce significant computational costs and improving the convergence rate. As previously motivated, a diagonal scaling allows to make the projection in step 2. of SGP a non-excessively expensive task; thus, we will concentrate on updating rules for scaling matrices of the form $D_k = \mathrm{diag}\big(d_1^{(k)}, \ldots, d_n^{(k)}\big)$. A classical choice is to use a diagonal scaling matrix that approximates the inverse of the Hessian matrix $\nabla^2 J(\boldsymbol{x})$; for example by requiring

$$d_i^{(k)} \approx \left(\frac{\partial^2 J(\boldsymbol{x}^{(k)})}{(\partial x_i)^2} \right)^{-1}, \qquad i = 1, \ldots, n.$$

In this case an updating rule for the entries of D_k could be

$$d_i^{(k)} = \min \left\{ L, \max \left\{ \frac{1}{L}, \left(\frac{\partial^2 J(\boldsymbol{x}^{(k)})}{(\partial x_i)^2} \right)^{-1} \right\} \right\}, \qquad i = 1, \ldots, n, \tag{10}$$

where L is an appropriate threshold. With a view to apply SGP to the deblurring problem (6), another appealing choice is suggested by the diagonal scaling used in (4) to rewrite the EM method as a special scaled gradient method; by following this idea we may introduce the updating rule

$$d_i^{(k)} = \min \left\{ L, \max \left\{ \frac{1}{L}, x_i^{(k)} \right\} \right\}, \qquad i = 1, \ldots, n. \tag{11}$$

This means that, when SGP is equipped with the diagonal scaling (11), it could be viewed as a generalization of the EM method, able to exploit non-unitary steplengths and to approach the minimization problem (6) in which the energy conservation is forced. From a computational viewpoint, the updating rule (10) is more expensive than (11), due to the computation of the diagonal entries of the Hessian. With regard to the convergence rate, when SGP applies to the problem (6), the scaling (10) provides improvements not so relevant to balance its additional costs and generally the scaling (11) is preferable. For this reason we report in the experiment section the results obtained by equipping SGP with the updating rule (11), while we refer to [7] for a deeper analysis of the two scaling strategies.

Another important issue for the convergence rate of a gradient method is the choice of the steplength, that has been widely investigated in the last years. Following the original ideas of Barzilai-Borwein (BB) [2], several steplength updating strategies have been devised to accelerate the slow convergence exhibited in most cases by standard gradient methods (see for example [8, 9, 10, 16] and references therein) and many numerical experiments have confirmed the promising improvements involved by these BB-like steplength selections. Thus, with the aim of designing a gradient projection method able to achieve a better convergence rate than the scheme (3), it seems natural to use within SGP a steplength selection that takes into account these recent advances. First of all we must rewrite, in case of a scaled gradient method, the two BB rules usually exploited by the main steplength updating strategies. To this end, we can regard the matrix $B(\alpha_k) = (\alpha_k D_k)^{-1}$ as an approximation of the Hessian $\nabla^2 J(\boldsymbol{x}^{(k)})$ and derive two updating rules for α_k by forcing quasi-Newton properties on $B(\alpha_k)$:

$$\alpha_k^{(1)} = \arg\min_{\alpha_k \in \mathbb{R}} \|B(\alpha_k)\boldsymbol{s}^{(k-1)} - \boldsymbol{z}^{(k-1)}\| \qquad \text{and} \qquad \alpha_k^{(2)} = \arg\min_{\alpha_k \in \mathbb{R}} \|\boldsymbol{s}^{(k-1)} - B(\alpha_k)^{-1}\boldsymbol{z}^{(k-1)}\|,$$

where $\boldsymbol{s}^{(k-1)} = \big(\boldsymbol{x}^{(k)} - \boldsymbol{x}^{(k-1)}\big)$ and $\boldsymbol{z}^{(k-1)} = \big(\nabla J(\boldsymbol{x}^{(k)}) - \nabla J(\boldsymbol{x}^{(k-1)})\big)$.
In this way, the steplengths

$$\alpha_k^{(1)} = \frac{{\boldsymbol{s}^{(k-1)}}^T D_k^{-1} D_k^{-1} \boldsymbol{s}^{(k-1)}}{{\boldsymbol{s}^{(k-1)}}^T D_k^{-1} \boldsymbol{z}^{(k-1)}} \qquad \text{and} \qquad \alpha_k^{(2)} = \frac{{\boldsymbol{s}^{(k-1)}}^T D_k \boldsymbol{z}^{(k-1)}}{{\boldsymbol{z}^{(k-1)}}^T D_k D_k \boldsymbol{z}^{(k-1)}}, \tag{12}$$

are obtained, that reduce to the standard BB rules in case of non-scaled gradient methods, that is when D_k is equal to the identity matrix for all k. At this point, inspired by the steplength alternations successfully implemented in the framework of non-scaled gradient methods [10, 16], we propose a steplength updating rule for SGP which adaptively alternates the values

$$\bar{\alpha}_k = \max\{\alpha_{min}, \min\{\alpha_{max}, \alpha_k^{(1)}\}\} \qquad \text{and} \qquad \hat{\alpha}_k = \max\{\alpha_{min}, \min\{\alpha_{max}, \alpha_k^{(2)}\}\}. \tag{13}$$

The alternation rule we use in the subsequent experiments is described in [7] and consists in a slight modification of the ABB_{min1} strategy proposed in [10].

4. Numerical results

In this section we apply the methods SGP, EM and EM_MATLAB to some deblurring problems arising from a set of astronomical images corrupted by Poisson noise. The behaviour of the three approaches will be compared in terms of number of iterations, reconstruction accuracy and computational time. All the methods are implemented in Matlab 7.0.4 and the experiments are performed on a computer equipped with an 1.8GHz AMD Sempron 3100+. The parameters of SGP are set as follows: $\beta = 10^{-4}$, $\theta = 0.4$, $M = 10$, $\alpha_{min} = 10^{-3}$, $\alpha_{max} = 10^5$; the threshold in (11) is $L = 10^{10}$.

Our test problems are generated by proceeding as in [1]: original 256×256 images are convolved with a simulated *point spread function* (PSF), a background term is added and the results are perturbed with Poisson noise. Several test problems have been considered, with different PSFs and noise levels; however, in all cases the relative behaviours of the methods are very similar and, consequently, we only discuss the results about the two test images reported in Figure 1. The performance of the three methods are described in Table 1 where we report the flux constant c controlling the different noise levels (we recall that, in the case of Poisson noise, the noise level is increasing when the total flux is decreasing), the numbers of iterations required by the three methods (it_opt), the corresponding computational times in seconds (sec) and the relative reconstruction error (err_opt), defined as $\|\boldsymbol{x}^{(k)} - \boldsymbol{x}\|/\|\boldsymbol{x}\|$, $\boldsymbol{x}$ being the image to be reconstructed. For each method, the results refer to the iteration where the minimum of the reconstruction error has been obtained, except for the cases marked with an asterisk in which the maximum number of iterations has been reached. The test images are denoted by Image A and Image B, corresponding to the objects in the upper and lower panels of Figure 1, respectively. For each image, three noise levels are considered; the reconstructions in Figure 1 refer to the noise level corresponding to $c = 4.43 \times 10^9$. Furthermore, in Figure 2 we show the behaviour of the relative reconstruction error as a function of the number of iterations for the three methods.

We remark that the computational time per iteration in SGP and EM_MATLAB is approximately 40% and 30% greater than in EM, respectively. However, SGP and EM_MATLAB exhibit a better convergence rate in comparison with EM, and the time required to obtain a given reconstruction error in these accelerated methods is significantly lower than in EM. In these experiments SGP generally converges faster than EM_MATLAB, even if this feature does not always imply a lower computational time, due to the higher costs per iteration. Concerning the reconstruction accuracy, no significant differences are observed in the three approaches. It is also interesting to remark that, in the case of SGP, the reconstruction error drops to a value close to the minimum in very few iterations and that it remains close to this value for a large number of iterations, so that the choice of the optimal number of iterations does not seem to be critical in the case of real images.

5. Conclusions and future works

We have proposed a scaled gradient projection method, called SGP, for solving the optimization problems arising in the maximum likelihood approach to image deblurring. The main features of

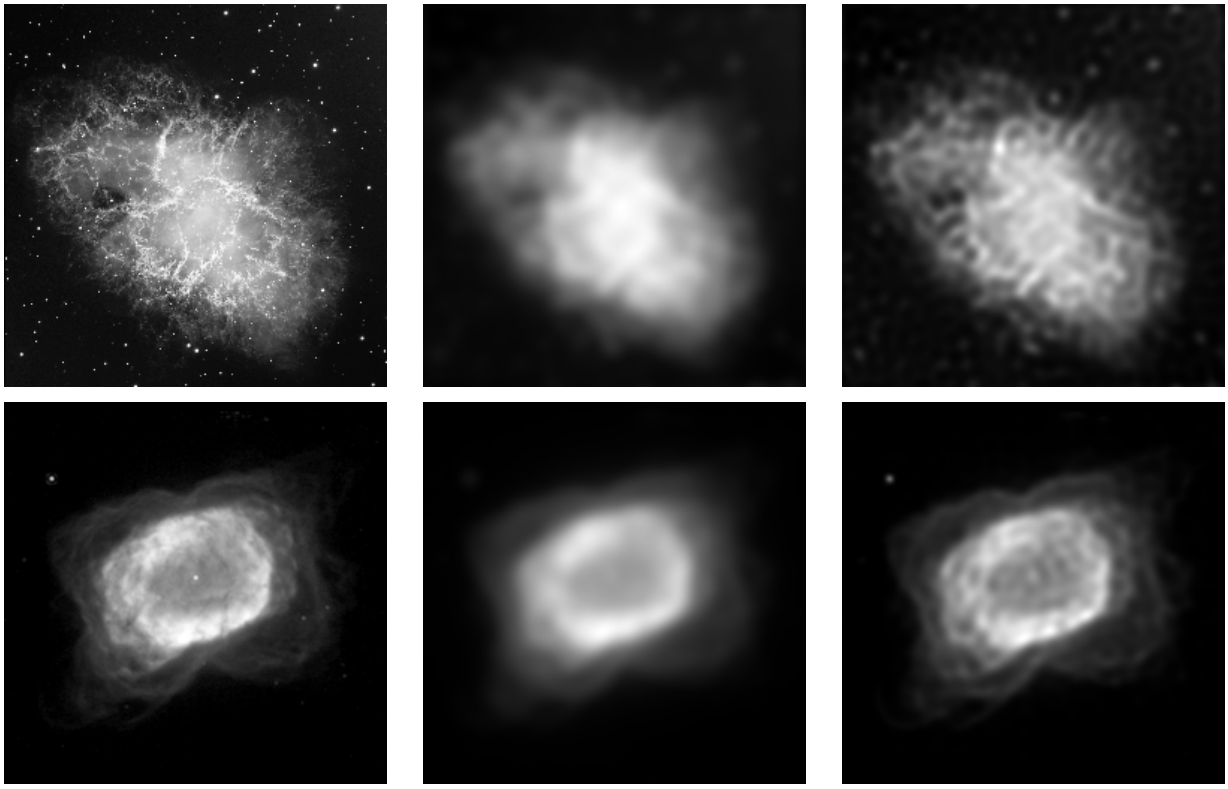

Figure 1. Test images: the objects (left panels), the blurred noisy images (middle panels), the SGP reconstructions (right panels).

Table 1. Behaviour of SGP, EM_MATLAB and EM.

	SGP			EM_MATLAB			EM		
c	it_opt	err_opt	sec	it_opt	err_opt	sec	it_opt	err_opt	sec
Image A									
4.43×10^{9}	339	0.185	63.7	388	0.185	64.9	3500*	0.186	458.9
7.02×10^{8}	108	0.187	20.3	141	0.187	23.0	3500*	0.187	447.4
4.43×10^{7}	20	0.195	4.0	46	0.195	7.7	414	0.194	54.5
Image B									
4.43×10^{9}	270	0.052	53.2	259	0.052	43.6	1500*	0.053	196.1
7.02×10^{8}	134	0.055	26.2	139	0.056	23.2	1500*	0.055	191.8
4.43×10^{7}	37	0.070	7.0	44	0.070	7.3	500	0.069	63.2

this approach are its global convergent properties and its ability to satisfy an energy conservation constraint. A computational study on a set of astronomical images corrupted by Poisson noise shows that the method is much faster than the standard EM method and very well-comparable with the EM_MATLAB algorithm, for which no convergence proof is available.

Future works will regard the evaluation of the proposed algorithm on other optimization problems arising in image deblurring and the comparison with second order schemes such as quasi-Newton and interior point methods. If SGP can compete with these methods, then it

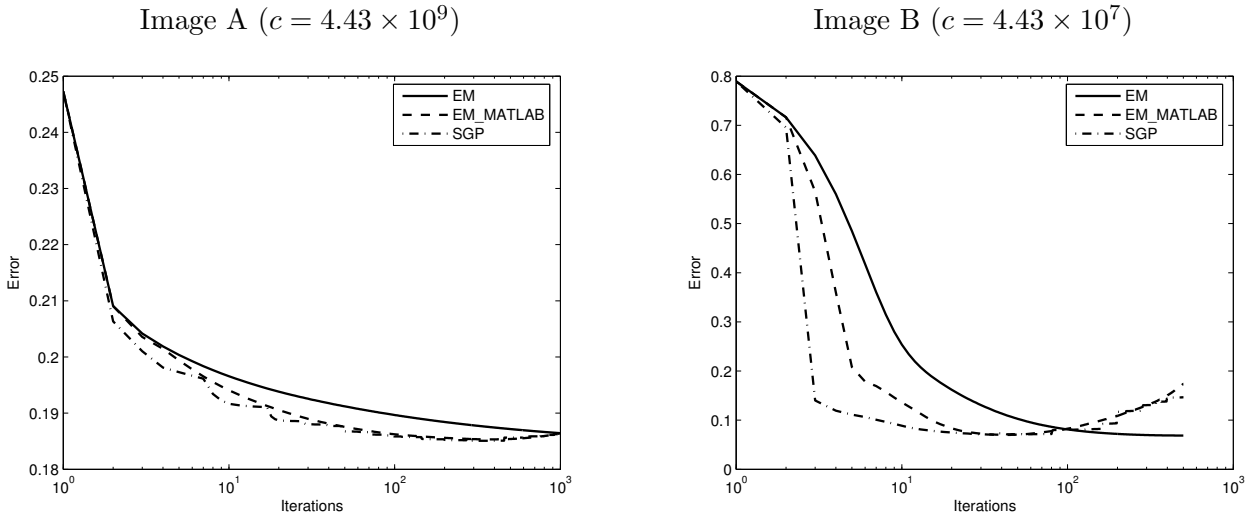

Figure 2. Relative reconstruction error

could provide a very useful and simple approach to iterative image reconstruction.

Acknowledgments

This research is supported by the PRIN2006 project of the Italian Ministry of University and Research *Inverse Problems in Medicine and Astronomy*, grant 2006018748.

References

[1] Anconelli B, Bertero M, Boccacci P, Carbillet M and Lanteri H 2005 Restoration of interferometric images - III. Efficient Richardson-Lucy methods for LINC-NIRVANA data reduction *Astron. Astrophys.* **430** 731–8
[2] Barzilai J and Borwein J M 1988 Two point step size gradient methods *IMA J. Num. Anal.* **8** 141–8
[3] Bertsekas D P 1999 *Nonlinear Programming* (Athena Scientific, 2nd edition)
[4] Biggs D S C and Andrews M 1997 Acceleration of iterative image restoration algorithms *Appl. Opt.* **36** 1766–75
[5] Birgin E Martinez J and Raydan M 2000 Nonmonotone spectral projected gradient methods on convex sets *SIAM J. Optim.* **10** 1196–211
[6] Birgin E Martinez J and Raydan M 2003 Inexact spectral projected gradient methods on convex sets *IMA J. Num. Anal.* **23** 539–59
[7] Bonettini S, Zanella R and Zanni L 2007 A scaled gradient projection method for constrained image deblurring, *Tech. Rep. N.78, Department of Mathematics, University of Modena and Reggio Emilia*
[8] Dai Y H and Fletcher R 2006 New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds, *Math. Prog. A* **106** 403–21
[9] Dai Y H Hager W W Schittkowski K and Zhang H 2006 The cyclic Barzilai-Borwein method for unconstrained optimization, *IMA J. Numer. Anal.* **26** 604–27
[10] Frassoldati G, Zanghirati G and Zanni L 2008 New adaptive stepsize selections in gradient methods *J. Industrial and Management Optim.* **4(2)** (available at <http://www.optimization-online.org>)
[11] Kiwiel K C 2007 On linear time algorithms for the continuous quadratic knapsack problem, *J. Optim. Theory Appl.* **134** 549–54
[12] Lucy L B 1974 An iterative technique for the rectification of observed distributions, *Astron. J.* **79** 745–54
[13] Maculan N *et al.* 2003 An $O(n)$ algorithm for projecting a vector on the intersection of a hyperplane and a box in $\mathbb{R}^n$ *J. Optim. Theory Appl.* **117** 553–74
[14] Richardson W H 1972 Bayesian–based iterative method of image restoration, *J. Opt. Soc. Amer. A* **62** 55–9
[15] Shepp L A and Vardi Y 1982 Maximum likelihood reconstruction for emission tomography *IEEE Trans. Med. Im.* **1** 113–22
[16] Zhou B, Gao L, and Dai Y H 2006 Gradient methods with adaptive step-sizes *Comput. Optim. Appl.* **35** 69–86