

Nonnegative least-squares image deblurring: improved gradient projection approaches

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2010 Inverse Problems 26 025004

(<http://iopscience.iop.org/0266-5611/26/2/025004>)

[The Table of Contents](#) and [more related content](#) is available

Download details:

IP Address: 130.251.61.251

The article was downloaded on 30/12/2009 at 09:51

Please note that [terms and conditions apply](#).

Nonnegative least-squares image deblurring: improved gradient projection approaches

F Benvenuto^{1,2}, R Zanella³, L Zanni³ and M Bertero⁴

¹ Dipartimento di Matematica, Università di Genova, Via Dodecaneso 35, I 16146 Genova, Italy

² Laboratoire Hippolyte Fizeau, Université de Nice Sophia Antipolis, CNRS UMR6525, 06108 Nice Cedex 2, France

³ Dipartimento di Matematica Pura e Applicata, Università di Modena e Reggio Emilia, Via Campi 213/b, I 41100 Modena, Italy

⁴ Dipartimento di Informatica e Scienze dell'Informazione, Università di Genova, Via Dodecaneso 35, I 16146 Genova, Italy

E-mail: bertero@disi.unige.it

Received 25 June 2009, in final form 29 October 2009

Published 18 December 2009

Online at stacks.iop.org/IP/26/025004

Abstract

The least-squares approach to image deblurring leads to an ill-posed problem. The addition of the nonnegativity constraint, when appropriate, does not provide regularization, even if, as far as we know, a thorough investigation of the ill-posedness of the resulting constrained least-squares problem has still to be done. Iterative methods, converging to nonnegative least-squares solutions, have been proposed. Some of them have the 'semi-convergence' property, i.e. early stopping of the iteration provides 'regularized' solutions. In this paper we consider two of these methods: the projected Landweber (PL) method and the iterative image space reconstruction algorithm (ISRA). Even if they work well in many instances, they are not frequently used in practice because, in general, they require a large number of iterations before providing a sensible solution. Therefore, the main purpose of this paper is to refresh these methods by increasing their efficiency. Starting from the remark that PL and ISRA require only the computation of the gradient of the functional, we propose the application to these algorithms of special acceleration techniques that have been recently developed in the area of the gradient methods. In particular, we propose the application of efficient step-length selection rules and line-search strategies. Moreover, remarking that ISRA is a scaled gradient algorithm, we evaluate its behaviour in comparison with a recent scaled gradient projection (SGP) method for image deblurring. Numerical experiments demonstrate that the accelerated methods still exhibit the semi-convergence property, with a considerable gain both in the number of iterations and in the computational time; in particular, SGP appears definitely the most efficient one.

1. Introduction

Image deblurring is a linear inverse problem since it consists in the inversion of a linear and continuous integral operator $H : X \rightarrow X$, where X is, for instance, a Hilbert space of square-integrable functions of two or three variables. In many instances, H is a convolution operator, i.e. $Hx = K * x$, for any $x \in X$, K being an integrable function, the so-called point-spread function (PSF) of the imaging system. A discussion of deconvolution problems is already contained in the book of Tikhonov and Arsenin [32], as well as in more recent books [4, 21, 23, 33].

The least-squares (LS) approach to image deblurring leads to an ill-posed problem. However, the situation may change if one considers a constrained LS problem

$$\begin{aligned} \text{minimize} \quad & J_y(x) = \frac{1}{2} \|Hx - y\|_X^2 \\ \text{subject to} \quad & x \in \Omega, \end{aligned} \tag{1}$$

where Ω is a closed and convex set. For instance, if Ω is compact and H is injective, then there exists a unique solution that depends continuously on the data y [22]. Very frequently, the most natural constraint is nonnegativity and the set of nonnegative functions is closed and convex in L^2 but not compact. If $H(\Omega)$, the image of Ω , is not closed, then the constrained LS problem is still ill-posed: a solution may not exist if a noisy y is not in $H(\Omega)$.

Since this paper is mainly computational, we are interested in the discrete version of the problem, i.e. H is a matrix, x, y are vectors/arrays and X is the ℓ_2 space; furthermore, we assume that the feasible region Ω is the nonnegative orthant. In this case, problem (1) reduces to the following convex optimization problem:

$$\begin{aligned} \text{minimize} \quad & J_y(x) = \frac{1}{2} \|Hx - y\|^2 \\ \text{subject to} \quad & x \geq 0, \end{aligned} \tag{2}$$

where $\|\cdot\|$ denotes the usual ℓ_2 norm of vectors/arrays. Then a solution always exists, and is also unique if the null space of H contains only the null element ($\mathcal{N}(H) = \{0\}$), a property that, in general, is satisfied as a consequence of the approximation errors in the computation of the matrix H . However, if H is ill-conditioned, we expect that this solution is completely deprived of physical meaning.

In [2] arguments are given to suggest that a nonnegative LS solution is a *night-sky reconstruction*, i.e. consists of a set of bright points over a black background. It may be interesting, from a pedagogical point of view, to verify this effect which is a consequence of the noise perturbing the data and visualizes the ill-posedness of the problem. To this purpose we compute the LS solution in the case of a test problem generated from the 256×256 image of the nebula NGC5979 shown in the upper-left panel of figure 1. This image is convolved with the PSF described in section 4 and the result is corrupted with additive white Gaussian noise. The resulting blurred and noisy image (upper-right panel of figure 1) is deconvolved with an iterative algorithm converging to the minimizer of the LS problem (2) (uniqueness holds true because the discrete Fourier transform of the PSF is never zero); the minimizer is shown in the lower-left panel of figure 1. The result is just a night-sky reconstruction, since it is a ‘sparse’ object, consisting of bright spots over a nearly black background. We believe that the distribution of the bright spots depends on the realization of the noise. Moreover, in the lower-right panel of figure 1 we show the normalized residual, as defined in section 4. It looks as a map of correlated noise and this means that the minimizer is a solution of the problem that is not acceptable even from the statistical point of view [29].

Iterative methods have been proposed for the computation of LS solutions: the *projected Landweber* (PL) method [15], in the case of problem (1) with Ω a general closed and convex

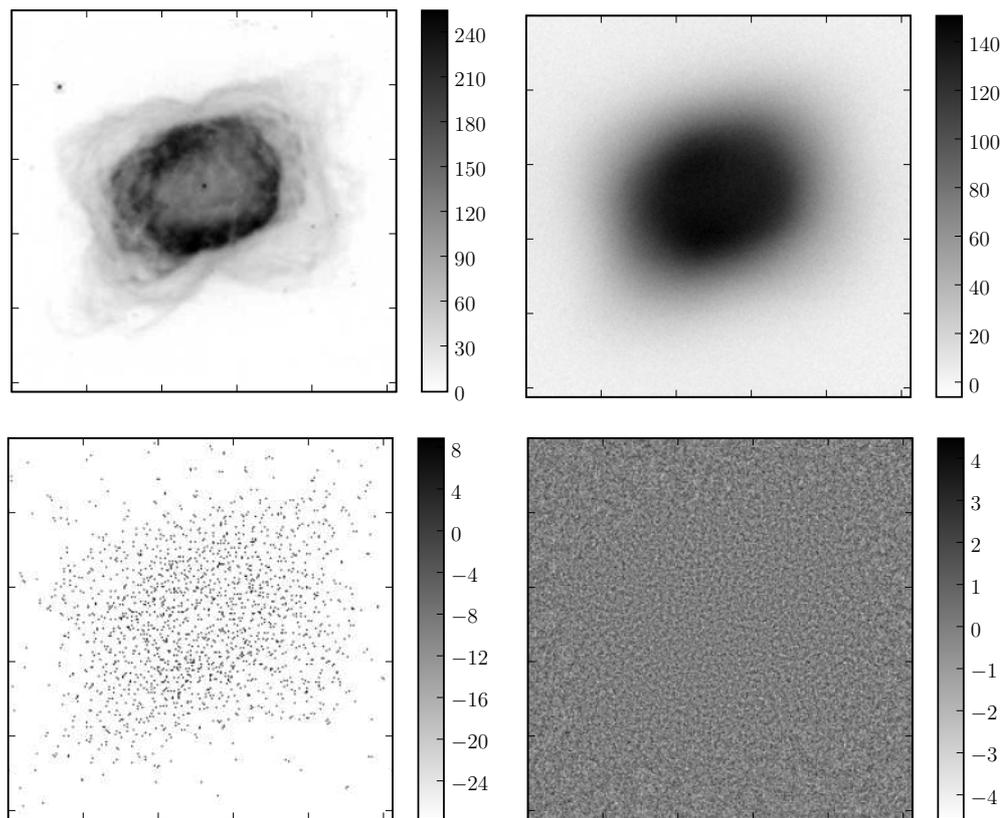


Figure 1. Upper panels: the object (left) and the corresponding blurred and noisy image (right). Lower panels: the reconstruction provided by the nonnegatively constrained minimum of the LS functional (log scale; left) and the corresponding normalized residual (right).

set, and the *iterative space reconstruction algorithm* (ISRA) [13], in the case of problem (2). Numerical experiments on the reconstruction of diffuse objects demonstrate that both methods exhibit the *semi-convergence* property [4, 16, 28]: the results of the iteration first provide better and better approximations of the true object but after a certain point they turn to the worse, due to increased noise propagation. Therefore, regularization is obtained in practice by a suitable stopping of the iterations. These numerical results are not supported by theory since, as far as we know, there is no proof that these iterative methods have regularization properties.

The main drawback of these methods is that their convergence is very slow. This is a general property of the so-called first-order methods, i.e. methods based only on the computation of the gradient. However, rather efficient strategies have been recently proposed for improving the convergence rate of classical gradient methods and the corresponding accelerated algorithms have shown satisfactory performance in many different applications (see, for example, [8, 11, 12, 17, 30, 35]). The main purpose of this work consists in adapting these acceleration techniques to PL and ISRA, and in evaluating, from a computational point of view, their effectiveness on nonnegative LS image deblurring.

In section 2 we discuss the known properties of the two methods and, in particular, we remark that PL is a projected gradient method while ISRA is a scaled gradient method. The

accelerated versions of the two methods are presented in section 3 and numerically evaluated in section 4; finally, some conclusions are given in section 5.

2. The two basic algorithms

Let us denote by x the unknown object, by y the detected image and by H the imaging matrix. We assume the following model of image formation:

$$y = Hx + z, \quad (3)$$

where z is a realization of additive white Gaussian noise with expected value 0 and variance σ^2 . In addition, we assume that the imaging matrix satisfies the following conditions:

$$H_{i,j} \geq 0; \quad \sum_{i \in S} H_{i,j} > 0, \quad \forall j \in R; \quad \sum_{j \in R} H_{i,j} > 0, \quad \forall i \in S, \quad (4)$$

where S and R are the sets of the values of the indices i and j respectively. In other words we assume that each row or column contains at least one non-zero element. We also remark that, as a consequence, we have

$$(H^T H)_{j,j} = \sum_{i \in S} H_{i,j}^2 > 0, \quad \forall j \in R. \quad (5)$$

The maximum likelihood approach based on the above imaging model leads to the minimization problem (2).

2.1. The projected Landweber method

The PL method is investigated in [15] in the general case of problem (1), where X is a Hilbert space. Then if we denote by P_Ω the projection onto Ω , the PL iteration is defined by

$$x^{k+1} = P_\Omega[x^k + \alpha(H^T y - H^T H x^k)], \quad (6)$$

where α is a fixed step-length in the descent direction

$$-\nabla J_y(x) = H^T y - H^T H x. \quad (7)$$

In [15] (theorem 3.2) it is proved that if $y \in H(\Omega)$ and α satisfies the condition

$$0 < \alpha < \frac{2}{\|H\|_2^2}, \quad (8)$$

where $\|H\|_2$ is the spectral norm of H , then, for any initial guess x^0 , the sequence $\{x^k\}$ weakly converges to a solution of problem (1). Moreover, it is conjectured that the convergence should be strong, even if this result is proved only in a number of special cases (theorem 3.3). Unfortunately the conjecture is not true, as demonstrated by a counterexample in [9] (remark 5.12).

In the case of the discrete problem (2), if the sufficient condition for uniqueness is satisfied, then the previous result implies that the PL iteration converges to the constrained LS solution. Even if this solution is deprived of any physical meaning, as already discussed in section 1, the PL method is not useless because, as demonstrated by a large amount of numerical experiments, it exhibits semi-convergence [4]. Obviously the proof of this property and the definition of appropriate stopping rules are open problems.

As in other cases one can use the so-called *discrepancy principle* [16] as a stopping rule: if y is an $n \times n$ image, then the iteration is stopped as soon as

$$\|Hx^k - y\| \leq n\sigma, \quad (9)$$

where σ^2 is the variance of the noise. Again, as far as we know, no result exists proving that this criterion, applied to the projected Landweber iteration, leads to a regularization method for the constrained LS problem.

From the practical point of view, slow convergence is an obstacle to a wide use of this method in spite of the good results provided in many cases.

2.2. The iterative space reconstruction algorithm

ISRA was proposed by Daube-Whiterspoon and Muehlener [13] for reducing the computational cost of the expectation-maximization (EM) algorithm of Shepp and Vardi [31] in emission tomography. The algorithm was presented in a descriptive way, without convergence proof. Subsequently De Pierro [14] proved asymptotic convergence of the iteration to solutions of problem (2), so that ISRA is related to Gaussian rather than to Poisson noise model.

More precisely, given an initial approximation $x^0 > 0$, the sequence $\{x^k\}$, $k = 0, 1, \dots$, provided by ISRA is defined by

$$x^{k+1} = x^k \frac{H^T y}{H^T H x^k}, \quad (10)$$

where H^T is the transpose of H (also called back-projection), and products and quotients of vectors/arrays are intended in the Hadamard sense, i.e. component by component or pixel by pixel.

If $H^T y > 0$, by remarking that the properties (4) imply $Hx > 0$ for any $x > 0$, then it is easy to prove by induction that, for any given $x^0 > 0$, all x^k are strictly positive, and therefore the algorithm is well defined. Under these conditions, as proved in [14], the sequence $\{x^k\}$ is asymptotically convergent to solutions of problem (2), i.e. the limit of each convergent subsequence of the sequence $\{x^k\}$ is a nonnegative LS solution x^* . If the solution is unique, then the sequence is convergent.

We remark that the condition $H^T y > 0$ may not be satisfied because the basic assumption underlying the approach is that the noise corrupting the data is additive, white and Gaussian, with zero expected value. Therefore, y , as well as $H^T y$, can have negative components. A similar situation arises if a constant background is superposed to the image or the expected value of the noise is positive. Indeed these quantities must be subtracted from the data; otherwise the nonnegativity constraint is not active. Also in this case the subtracted data can take negative values.

However, it is possible to avoid this difficulty with a simple modification of the algorithm. Let b be a positive constant such that $H^T y + b > 0$; then, we consider the following modified ISRA:

$$x^{k+1} = x^k \frac{H^T y + b}{H^T H x^k + b}. \quad (11)$$

It is easy to show, by a simple modification of De Pierro's proof, that the new sequence $\{x^k\}$ is also asymptotically convergent to nonnegative LS solutions.

As already remarked in the introduction, numerical experiment demonstrates that the sequence of x^k has the property of semi-convergence, so that regularization can be obtained by early stopping of the iterations. In the absence of ad hoc stopping rules, one can use also in this case the discrepancy principle (9).

An important remark, that will be used in the following, is that both ISRA and modified ISRA are scaled gradient methods with unit step-length. In fact, if we consider the following gradient method:

$$x^{k+1} = x^k - \alpha D_k \nabla J_y(x^k), \quad (12)$$

where the gradient is scaled by a positive definite diagonal matrix, then equation (7) implies that we reobtain equation (11) if we set $\alpha = 1$ and $D_k = x^k / (H^T H x^k + b)$. It is interesting to remark that this scaling, that can be obtained from the computation of the gradient, also mimics the inversion of the Hessian of the problem.

Taking into account the above reformulation of ISRA, one can try to achieve convergence rate improvements by exploiting a variable step-length in combination with a projection step, for preserving the feasibility of the iteration, and with a line-search strategy, for ensuring a suited reduction of the objective function. Based on similar ideas, an efficient *scaled gradient projection* (SGP) method has been recently proposed for accelerating the EM algorithm in the case of images corrupted by Poisson noise [8, 5]. The computational study described in [8] shows that SGP exhibits a much better convergence rate in comparison with EM and a similar reconstruction accuracy. Thus, it is interesting to evaluate the SGP behaviour also on the LS problem (2); we deal with this topic in the following section.

3. Acceleration of the basic algorithms

Improved versions of PL and ISRA can be obtained by exploiting some well-known accelerating strategies widely used in the area of the first-order optimization methods. To this purpose we recall a few definitions taken from Bertsekas [6]; from now on, Ω is the nonnegative orthant and P_+ denotes the projection onto this orthant.

- Given a feasible vector x , i.e. a vector $x \in \Omega$, a *feasible direction* at x is a vector z such that $x + \lambda z \in \Omega$ for sufficiently small $\lambda > 0$. For instance, the negative scaled gradient of equation (12) is a descent direction that is also a feasible direction.
- The *projection arc* associated with x is the set of vectors defined by

$$x(\alpha) = P_+[x - \alpha \nabla J_y(x)], \quad \alpha > 0. \quad (13)$$

This definition can be extended to the case where the gradient is replaced by a scaled gradient with a positive definite scaling matrix D . Then the projection arc is the set of vectors given by

$$x_D(\alpha) = P_+[x - \alpha D \nabla J_y(x)], \quad \alpha > 0. \quad (14)$$

3.1. PL combined with Armijo rule along the projection arc

The most simple way for accelerating PL is to update the current iteration x^k by exploiting a variable step-length obtained from an Armijo rule along the projection arc $x^k(\alpha)$, defined by equation (13) with x replaced by x^k . Therefore, the iteration takes the following form:

$$x^{k+1} = P_+[x^k + \alpha^k (H^T y - H^T H x^k)], \quad (15)$$

with α^k chosen according to the following rule: fix the constants $\bar{\alpha} > 0$ and $\beta, \gamma \in (0, 1)$; then $\alpha^k = \beta^{m_k} \bar{\alpha}$, where m_k is the first nonnegative integer for which

$$J_y(x^k) - J_y(x^k(\beta^{m_k} \bar{\alpha})) \geq \gamma \nabla J_y(x^k)^T (x^k - x^k(\beta^{m_k} \bar{\alpha})), \quad (16)$$

the gradient being given in equation (7). This algorithm will be called the *projection arc* (PA) algorithm. Convergence proof is given, for instance, in [24].

In PA, the search for α^k uses a fixed initial guess $\bar{\alpha}$, the same at each iteration, and, for this reason, the procedure may be inefficient in some cases. To overcome this drawback, the more sophisticated line-search strategy introduced in [27] can be used. At iteration k the previous step-length α^{k-1} is assumed as the initial trial in the search for the suited α^k . If this

value does not satisfy the sufficient decrease condition (16) at iteration k , then the step-length is decreased by multiplying by β , until the condition is satisfied. On the other hand, if the sufficient decrease condition is satisfied by α^{k-1} , then the step-length is increased by dividing by β to reach a value such that the condition is not satisfied; the preceding value is used as the new step-length. We denote this algorithm by PA_1.

3.2. Gradient projection methods with Armijo rule along the feasible direction

A third acceleration approach consists in introducing the feasible direction

$$z^k = P_+[x^k + \alpha^k(H^T y - H^T H x^k)] - x^k, \quad (17)$$

with α^k chosen according to rules that will be discussed in a moment. In [7] it is proved that this is also a descent direction, so that one can update the current iteration x^k by means of Armijo rule along this direction, i.e.

$$x^{k+1} = x^k + \lambda^k z^k, \quad (18)$$

with $\lambda^k \in (0, 1)$ given as follows: fix the constants $\beta, \gamma \in (0, 1)$; then $\lambda^k = \beta^{m_k}$, m_k being the first nonnegative integer for which

$$J_y(x^k) - J_y(x^k + \beta^m z^k) \geq -\gamma \beta^m \nabla J_y(x^k)^T z^k. \quad (19)$$

The crucial point is the choice of the step-lengths α^k . First of all, we observe that the convergence of this method is ensured for any choice of the step-length α^k in a closed interval of positive values (see [7]). Motivated by the original ideas of Barzilai and Borwein (BB) [3], several step-length updating strategies have been devised to accelerate the convergence rate of classical gradient methods and many numerical experiments have confirmed the practical effectiveness of these strategies [10–12, 17–20, 35, 36]. Inspired by the promising results recently obtained in [19, 36], we equip our gradient projection method with a step-length updating rule based on an adaptive alternation of the BB rules.

We start by recalling these rules: by denoting with I the identity matrix and regarding the matrix $B(\alpha^k) = (\alpha^k I)^{-1}$ as an approximation of the Hessian $\nabla^2 J_y(x^k)$, the two BB updating rules for α^k are derived by forcing quasi-Newton properties on $B(\alpha^k)$:

$$\alpha_{\text{BB1}}^k = \arg \min_{\alpha^k \in \mathbb{R}} \|B(\alpha^k) s^{k-1} - t^{k-1}\| \quad (20)$$

and

$$\alpha_{\text{BB2}}^k = \arg \min_{\alpha^k \in \mathbb{R}} \|s^{k-1} - B(\alpha^k)^{-1} t^{k-1}\|, \quad (21)$$

where

$$s^{k-1} = x^k - x^{k-1}, \quad t^{k-1} = \nabla J_y(x^k) - \nabla J_y(x^{k-1}). \quad (22)$$

In this way, the step-lengths

$$\alpha_{\text{BB1}}^k = \frac{s^{k-1T} s^{k-1}}{s^{k-1T} t^{k-1}}, \quad \alpha_{\text{BB2}}^k = \frac{s^{k-1T} t^{k-1}}{t^{k-1T} t^{k-1}}, \quad (23)$$

are obtained.

At this point, following [19, 36], we exploit an adaptive alternation of the values provided by equation (23). More precisely the selection algorithm works as follows. We fix an interval $[\alpha_{\min}, \alpha_{\max}]$, and we choose a value of α^0 in this interval. We also fix $\tau_1 \in (0, 1)$ and a nonnegative integer M_α . Next, for $k = 1, 2, \dots$, if $s^{k-1T} t^{k-1}$ is not positive, then $\alpha^k = \alpha_{\max}$, else we compute

$$\alpha_1^k = \max\{\alpha_{\min}, \min\{\alpha_{\text{BB}1}^k, \alpha_{\max}\}\} \quad (24)$$

$$\alpha_2^k = \max\{\alpha_{\min}, \min\{\alpha_{\text{BB}2}^k, \alpha_{\max}\}\}. \quad (25)$$

If $\alpha_2^k/\alpha_1^k \leq \tau_k$, then we set

$$\alpha^k = \min\{\alpha_2^j, j = \max\{1, k - M_\alpha, \dots, k\}\}; \quad \tau_{k+1} = 0.9\tau_k; \quad (26)$$

else we set

$$\alpha^k = \alpha_1^k; \quad \tau_{k+1} = 1.1\tau_k. \quad (27)$$

Unlike the rules in [19, 36], the above strategy updates the threshold τ_k at each iteration; in our experience, this makes the choice of τ_1 less important for the performance of the method and reduces the drawbacks due to the use of the same step-length rule in too many consecutive iterations.

This algorithm will be called the *gradient projection* (GP) algorithm. Its convergence can be obtained by proceeding as in [7].

3.3. ISRA combined with Armijo rule along the projection arc

As already observed in section 2.2, ISRA can be viewed as a special scaled gradient method that exploits the direction

$$d^k = -\frac{x^k}{H^T H x^k + b} \nabla J_y(x^k). \quad (28)$$

This scaled gradient direction is feasible since, for any $\alpha \in (0, 1]$,

$$x^{k+1} = x^k + \alpha d^k \quad (29)$$

is nonnegative; more precisely, the nonnegativity of x^{k+1} holds true for all α smaller than some $\alpha_{\max}^k \geq 1$, and, in [25], a moderate acceleration of ISRA is obtained by searching for a suited α (applying, for instance, the Armijo rule) in the interval $(0, \alpha_{\max}^k)$.

The extension of the acceleration technique used in the PA algorithms to ISRA is obvious. Given the current iterate x^k and the scaling matrix D_k , the corresponding projection arc is given by equation (14), with x replaced by x^k and D replaced by D_k . Then the iteration of this generalized ISRA is given by

$$x^{k+1} = P_+[x^k + \alpha^k D_k(H^T y - H^T H x^k)], \quad (30)$$

where P_+ is the projection on the nonnegative orthant and α^k is chosen according to the Armijo rule as in equation (16). This algorithm will be called the *scaled projection arc* (SPA) algorithm and the proof of its convergence can be found in [24]. Moreover, as in the case of PL, we can use the line-search approach suggested in [27]; the corresponding algorithm will be called SPA_1.

About the choice of the scaling matrix D_k , it can be obtained from the diagonal scaling derived in equation (12) from equation (11). However, for ensuring properties useful in the convergence proofs (see [6, 8, 24]), the following slightly modified form of the diagonal scaling is used:

$$D_k = \text{diag} \left(\min \left[L, \max \left\{ \frac{1}{L}, \frac{x^k}{H^T H x^k + b} \right\} \right] \right), \quad (31)$$

where $L > 1$ is an appropriate large threshold (in our simulations we take $L = 10^5$). It is obvious that such a modification is irrelevant in practice.

3.4. Scaled gradient projection method with Armijo rule along the feasible direction

Concerning the acceleration approach based on the use of the Armijo rule along the feasible direction, we first remark that, since Ω is the nonnegative orthant, it is not necessary to modify the projection operator as done in [8], where the projection is performed according to the norm induced by D_k^{-1} , i.e. $\|x\|_{D_k^{-1}} = \sqrt{x^T D_k^{-1} x}$. Then the feasible direction is given by

$$z^k = P_+[x^k + \alpha^k D_k(H^T y - H^T H x^k)] - x^k. \quad (32)$$

In [8] it is proved that this is also a descent direction and therefore the current iteration can be updated as in equation (18), with λ^k determined by means of Armijo rule as in equation (19).

About the selection rule for the step-length α^k , the same alternating strategy used in the algorithm GP is considered, but with the modified BB-rules [8]

$$\alpha_{\text{BB1}}^k = \frac{s^{k-1T} D_k^{-1} D_k^{-1} s^{k-1}}{s^{k-1T} D_k^{-1} t^{k-1}}, \quad \alpha_{\text{BB2}}^k = \frac{s^{k-1T} D_k t^{k-1}}{t^{k-1T} D_k D_k t^{k-1}}, \quad (33)$$

derived by forcing the quasi-Newton properties (20) and (21) on the matrix $B(\alpha^k) = (\alpha^k D_k)^{-1}$. The resulting algorithm will be called the *scaled gradient projection* (SGP) algorithm and its convergence is proved in [8].

4. Numerical experiments

We compare the algorithms introduced in the previous sections in the case of image deconvolution by assuming that Ω is the nonnegative orthant. In such a case $H_{i,j} = K(i-j)$, where i, j are multi-indices ranging on the domain of the image, an array $n \times n$. The array $K(i)$ is the point spread function (PSF) of the imaging system and we assume that it satisfies the normalization condition

$$\sum_i K(i) = 1. \quad (34)$$

We consider a few test problems. The PSF used in our numerical experiments is shown in figure 2 (left panel) together with the corresponding modular transfer function (MTF), i.e. the modulus of its (discrete) Fourier transform (right panel). The MTF is visualized using a log scale. The PSF simulations are taken by a ground-based telescope (in the right panel of figure 2 the band of the telescope is evident as well as the out-of-band noise affecting the PSF) and can be downloaded from <http://www.mathcs.emory.edu/~nagy/RestoreTools/index.html>.

We remark that the minimum value of the MTF is not zero, even if it is quite small, about 5.82×10^{-13} . Therefore, the minimum of the corresponding constrained LS problem is unique, even if it is not an acceptable solution (see the introduction and figure 1).

We consider two objects: an image of the nebula Ngc5979, already shown in the upper-left panel of figure 1, and the frequently used spacecraft image, shown in the left panel of figure 3. Both objects have values ranging approximately from 0 to 255, but they have quite different features: the nebula is a diffuse and smooth object, while the spacecraft has sharp edges and a rather complex structure.

Both objects are convolved with the PSF shown in figure 2 and the results are perturbed with additive white Gaussian noise with zero expected value and two different variances: $\sigma^2 = 1$ and $\sigma^2 = 5$. The blurred and noisy images, corresponding to $\sigma^2 = 5$, are shown in figure 1 (upper-right panel) and in figure 3 (right panel).

The six algorithms described in section 3, as well as PL and ISRA, implemented in C on an Intel Quad Core Q9600, are applied to the four test problems indicated above. Since all the

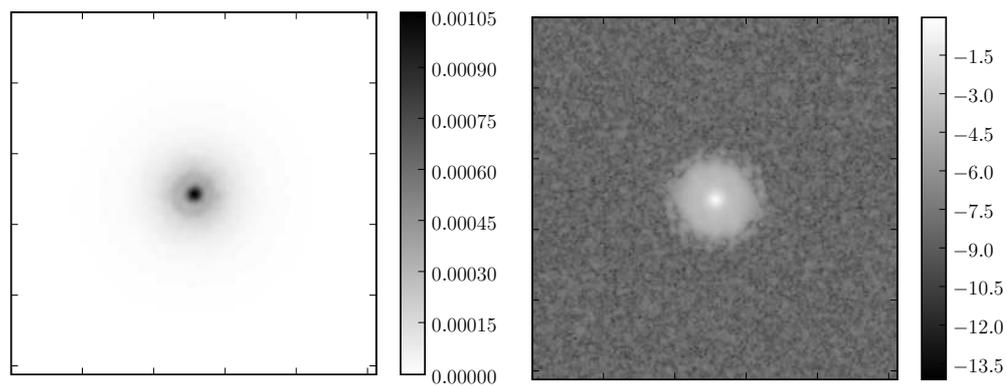


Figure 2. The PSF used in our numerical experiments (left panel) and the corresponding MTF (right panel).

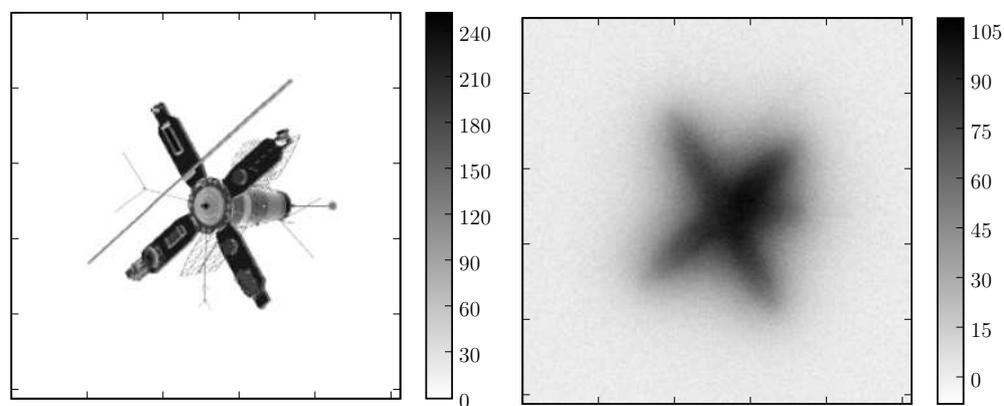


Figure 3. The spacecraft (left panel) and the corresponding blurred and noisy image (right panel).

algorithms converge to the unique LS solution, and have the semi-convergence property, we consider two stopping rules.

The first criterion consists in choosing the value of k corresponding to the minimum of the relative reconstruction error, defined as follows in terms of the ℓ_2 norm:

$$\rho^k = \frac{\|x^k - \bar{x}\|}{\|\bar{x}\|}, \quad (35)$$

where \bar{x} is the original object. Obviously this criterion cannot be used in real life but is frequently used in numerical experiments for its pedagogical value.

The second stopping rule we consider is the discrepancy principle defined in (9) which can be used with real data because, very frequently, information about the variance of the noise is available. Moreover, if \bar{x}^* is a generic solution provided by one of these stopping rules, we verify its statistical significance [29] by computing the normalized residual

$$R^* = \frac{1}{\sigma}(H\bar{x}^* - y). \quad (36)$$

Table 1. Reconstruction of the nebula Ngc5979 in the case $\sigma^2 = 1$. Relative reconstruction error and number of iterations are given for the two stopping rules (minimum and discrepancy).

	Minimum				Discrepancy	
	Relative error	Iteration	Iterations per second	Time (s)	Relative error	Iteration
PL	8.06%	1134	485	2.34 s	8.95%	464
PA	8.05%	777	166	4.67 s	8.93%	318
PA_1	8.04%	211	106	2.00 s	8.60%	124
GP	8.08%	155	228	0.68 s	8.78%	73
ISRA	7.50%	1800	409	4.40 s	8.21%	806
SPA	7.96%	568	178	3.20 s	8.29%	335
SPA_1	8.93%	472	98	4.81 s	9.05%	344
SGP	7.51%	109	163	0.67 s	8.13%	65

Finally, a careful parameter tuning for our simulations suggests the following parameter setting:

- PL: $\alpha = 1.8$;
- PA: $\bar{\alpha} = 6.0$;
- PA_1: $\bar{\alpha} = 3.0$ in the first iteration;
- GP: $\alpha_0 = 1.3$, $\alpha_{\min} = 10^{-3}$, $\alpha_{\max} = 10^5$, $M_\alpha = 2$, $\tau_1 = 0.15$;
- SPA: $\bar{\alpha} = 6.0$, $L = 10^5$;
- SPA_1: $\bar{\alpha} = 3.0$ in the first iteration, $L = 10^5$;
- SGP: $\alpha_0 = 1.3$, $\alpha_{\min} = 10^{-3}$, $\alpha_{\max} = 10^5$, $M_\alpha = 2$, $\tau_1 = 0.15$, $L = 10^5$.

The other parameters used in the line-search strategies are $\beta = 0.4$ and $\gamma = 10^{-4}$. Furthermore, for all the algorithms, x^0 is the constant image with the same flux of the detected image.

Some comments about the selection of these parameters can be useful. We tune the parameters in order to achieve the best convergence rate for each method; in particular, special care is taken in choosing

- $\bar{\alpha}$ in PA and SPA to avoid, at each iteration, too much step-length reductions or a too small step-length,
- α_{\min} and α_{\max} in GP and SGP respectively to avoid excessive limitations of the step-length suggested by the BB rules,
- L in the scaled versions to avoid excessive limitations of the diagonal values of the scaling matrix.

The values that we find convenient for the other parameters are very similar to those usually exploited in the literature and we observe a slight sensitivity of the methods to small changes of these values.

In tables 1 and 2 we report the results obtained in the case of the nebula for the two different noise levels. From these results we can observe that GP and SGP outperform all the other methods both as the number of iterations and as computing time. Moreover, their reconstruction errors are very well comparable with those provided by the other methods. The last two columns of tables 1 and 2 refer to the results obtained with the discrepancy principle. We remark that they provide a smaller number of iterations and a higher reconstruction error, but the corresponding reconstructions are still acceptable.

Table 2. Reconstruction of the nebula Ngc5979 in the case $\sigma^2 = 5$. Relative reconstruction error and number of iterations are given for the two stopping rules.

	Minimum				Discrepancy	
	Relative error	Iteration	Iterations per second	Time (s)	Relative error	Iteration
PL	10.03%	488	440	1.11 s	11.13%	193
PA	10.01%	333	161	2.07 s	11.08%	135
PA_1	9.96%	103	17	6.14 s	10.64%	54
GP	10.02%	79	214	0.37 s	10.88%	41
ISRA	9.36%	743	383	1.94 s	10.24%	317
SPA	9.95%	219	171	1.28 s	10.22%	139
SPA_1	10.87%	206	99	2.08 s	10.99%	161
SGP	9.43%	54	142	0.38 s	10.05%	35

Table 3. Reconstruction of the spacecraft in the case $\sigma^2 = 1$. Relative reconstruction error and number of iterations are given for the two stopping rules.

	Minimum				Discrepancy	
	Relative error	Iteration	Iterations per second	Time (s)	Relative error	Iteration
PL	27.83%	10307	468	22.02 s	30.35%	2355
PA	27.83%	3600	260	13.86 s	30.34%	920
PA_1	27.83%	3166	121	26.12 s	30.34%	964
GP	27.82%	1702	213	8.00 s	30.18%	416
ISRA	29.95%	5248	403	13.01 s	30.21%	3466
SPA	30.23%	1348	182	7.41 s	30.41%	961
SPA_1	29.99%	1215	100	12.17 s	30.12%	916
SGP	29.98%	238	146	1.63 s	30.10%	170

In figure 4 we show the best reconstructions, i.e. those corresponding to the minimum error (35), obtained by means of SGP for the two noise levels, with the corresponding normalized residuals.

In tables 3 and 4 we report the results obtained in the case of the spacecraft for the two different noise levels, while in figure 5 we show the best reconstructions obtained by means of SGP with the corresponding normalized residuals. In this case SGP clearly outperforms the other methods. However, in the case $\sigma^2 = 1$, the normalized residual shows a few artefacts, indicating that this reconstruction presumably is not the best one from the statistical point of view. Similar artefacts are present also in the residuals of the reconstructions provided by the other scaled methods. We also remark that these methods provide reconstructions with an error that is about 2% higher than that of the non-scaled ones. Therefore, for completeness, we show in figure 6 the reconstructions and the residuals obtained by means of GP. A visual inspection seems to confirm that these reconstructions are better than those provided by SGP. Moreover, the residuals do not exhibit the artefacts observed in figure 5. The effect of the scaling we have observed requires a further investigation.

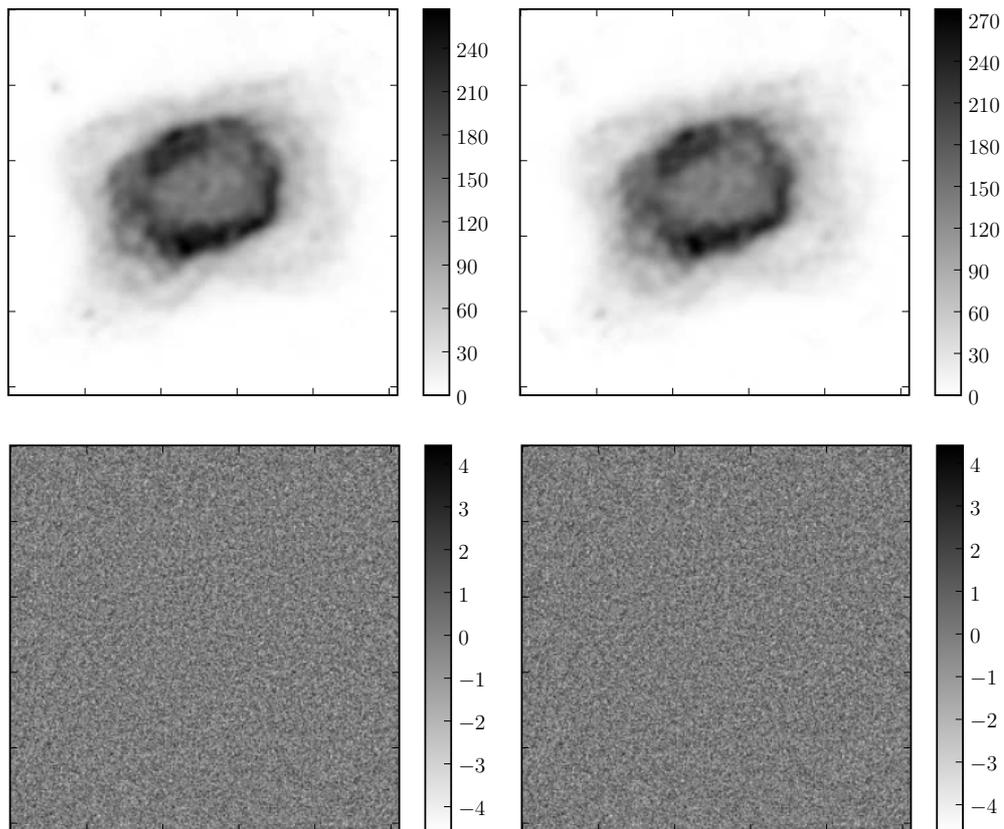


Figure 4. Upper panels: the best reconstruction of the nebula in the case $\sigma^2 = 1$ (left) and in the case $\sigma^2 = 5$ (right). Lower panels: the corresponding normalized residual.

Table 4. Reconstruction of the spacecraft in the case $\sigma^2 = 5$. Relative reconstruction error and number of iterations are given for the two stopping rules.

	Minimum				Discrepancy	
	Relative error	Iteration	Iterations per second	Time (s)	Relative error	Iteration
PL	31.24%	3813	415	9.18 s	34.19%	886
PA	31.24%	1400	246	5.69 s	34.17%	377
PA_1	31.23%	1316	126	10.42 s	34.18%	406
GP	31.21%	806	228	3.54 s	33.86%	99
ISRA	32.84%	2369	419	5.65 s	33.73%	1280
SPA	33.18%	630	174	3.62 s	33.94%	363
SPA_1	32.90%	581	101	5.78 s	33.47%	373
SGP	32.82%	124	146	0.85 s	33.65%	78

Figure 6 clearly shows the limitations inherent in the methods providing regularization through early stopping of the iteration. Indeed, we observe a breaking up of the image into

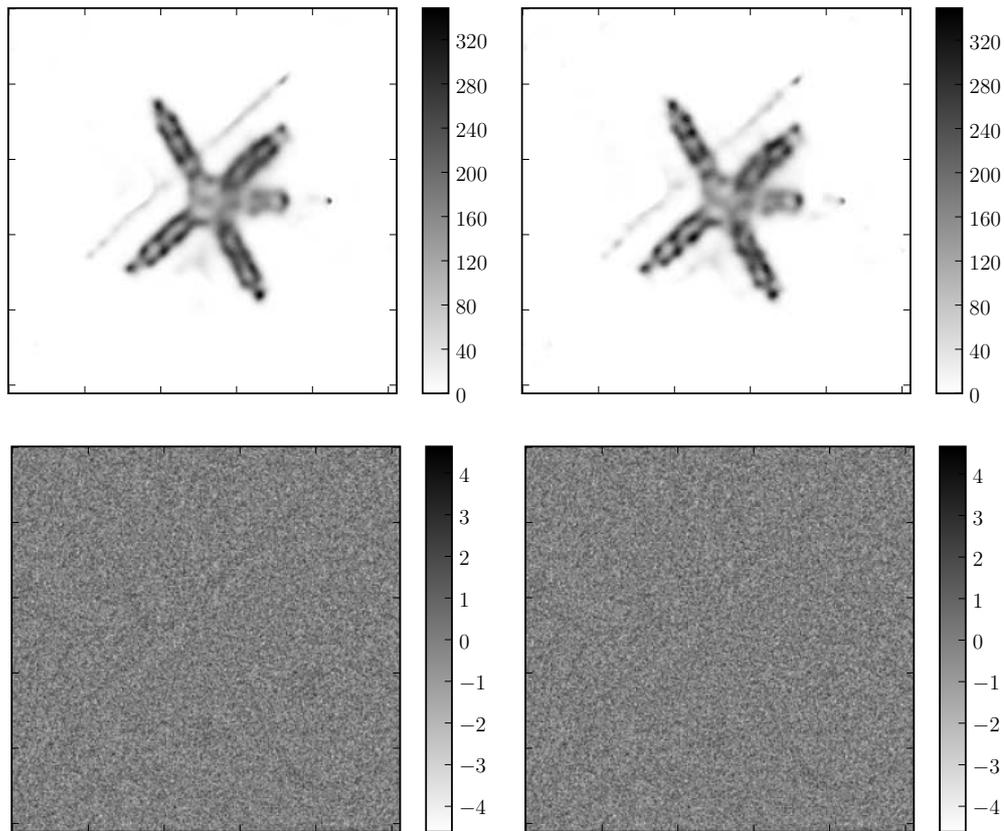


Figure 5. Upper panels: the best SGP reconstruction of the spacecraft in the case $\sigma^2 = 1$ (left) and in the case $\sigma^2 = 5$ (right). Lower panels: the corresponding normalized residual.

blobs along the solar panels of the spacecraft. This effect indicates the need of regularization. Work is in progress for applying SGP to the regularization of the LS approach by means of edge-preserving or other functionals. On the other hand, it seems that regularization is not needed in the case of the nebula; in general, we expect that SGP provides a very fast and accurate reconstruction in the case of diffuse objects.

In order to assess the computational complexity of SGP, we study the reconstruction time in relation to the size of the test images. We start with the 256×256 objects of the previous tests and we expand, by zero padding, their Fourier transform to an array 512×512 . The same procedure is applied to the PSF (remark that, in these simulations, we use an ideal PSF, i.e. an Airy pattern, providing a smaller reconstruction error for a given noise level). But, while the new PSF is already normalized to 1 as the previous one, we multiply the new object by 4 so that its mean pixel value is approximately the same of the original object. As a consequence, if we convolve the new object with the new PSF and we perturb the result with additive Gaussian noise with the same σ^2 , we obtain a noise level very close to that of the smaller images. The same procedure is repeated for obtaining images 1024×1024 and 2048×2048 . In this way, for each object, we have four test images with different sizes and approximately the same noise level.

Next, we determine the optimal number of iterations for the 256×256 images by minimizing the relative reconstruction error (35) and we reconstruct the images with larger

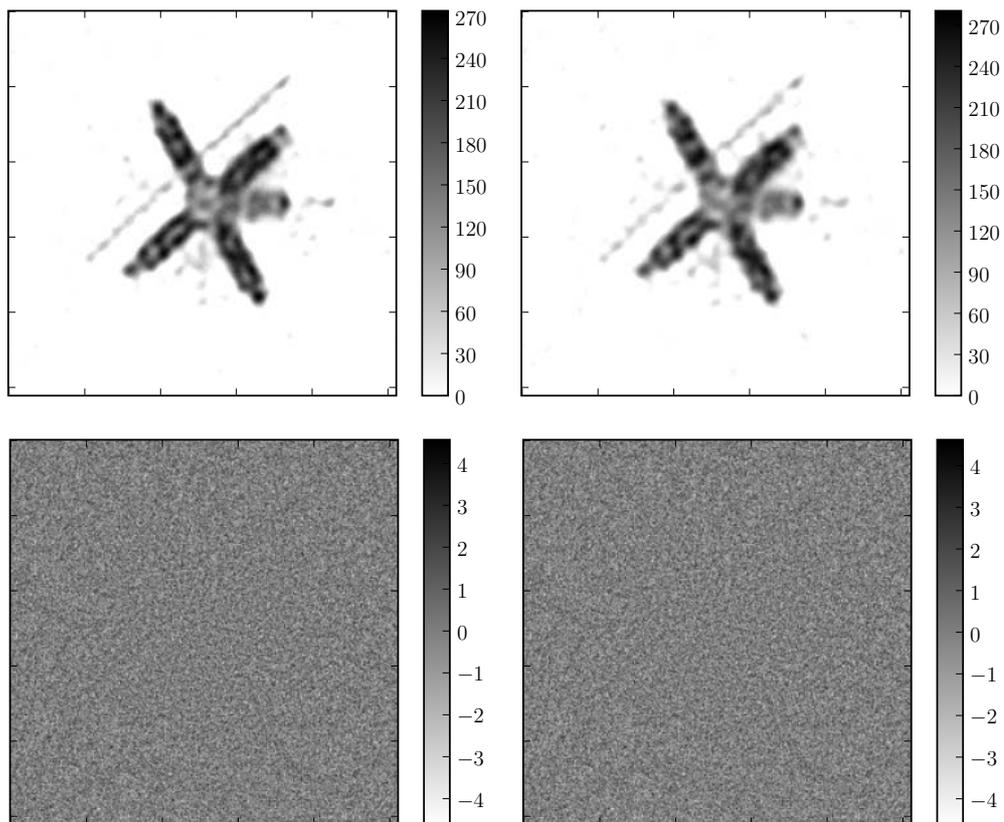


Figure 6. Upper panels: the best GP reconstruction of the spacecraft in the case $\sigma^2 = 1$ (left) and in the case $\sigma^2 = 5$ (right). Lower panels: the corresponding normalized residual.

size using this number: therefore, the computational time of each reconstruction depends only on the computational time of a single iteration. The results are reported in table 5. We remark a slight decrease in the reconstruction error when the size of the images increases, even if it does not correspond to a significant improvement of the reconstruction. By comparing the computational times for the different sizes it seems that the SGP complexity is essentially related to the complexity of the FFT algorithm.

5. Concluding remarks

In this paper we show that recent strategies of step-length selection, possibly combined with a suitable scaling of the gradient, provide very efficient iterative methods for nonnegative least-squares image deblurring. The proposed methods can be essentially considered as accelerations of well-known methods such as the projected Landweber and ISRA. The semi-convergence property of these algorithms is preserved by the accelerated versions, so that regularization can be obtained by early stopping of the iterations. We consider two stopping rules: the minimum of the relative mean square error and the discrepancy principle. Even if the latter is not theoretically justified in the case of iterative methods for nonnegative LS solutions, we find that it provides reasonable results since it allows us to use a smaller number of iterations than the criterion of minimum relative mean square error (remark, however, that

Table 5. Behaviour of the SGP algorithm applied to test images with different sizes. The same number of iterations is used for a given test object and a given noise level. An ideal psf is used in these simulations.

Dimension	Nebula			
	$\sigma^2 = 1$, iterations = 58		$\sigma^2 = 5$, iterations = 12	
	Relative error	Time (s)	Relative error	Time (s)
256 ²	6.33%	0.37	7.12%	0.08
512 ²	6.20%	1.90	6.84%	0.40
1024 ²	6.20%	8.86	6.98%	1.86
2048 ²	6.18%	36.90	6.78%	7.78

Dimension	Spacecraft			
	$\sigma^2 = 1$, iterations = 379		$\sigma^2 = 5$, iterations = 58	
	Relative error	Time (s)	Relative error	Time (s)
256 ²	26.18%	2.29	28.55%	0.36
512 ²	25.83%	12.44	27.42%	1.86
1024 ²	25.55%	59.74	27.46%	8.57
2048 ²	25.51%	250.20	26.83%	36.51

this can be used only in the case of simulations) with a still acceptable reconstruction error. Therefore, it could be used in the case of real data.

Among the considered accelerated methods, SGP shows the best convergence rate; it largely outperforms the other methods in the case of the spacecraft, but it is not able (as all the scaled methods) to reach the same reconstruction error of the non-scaled ones. This is a point that is still to be understood. However, the fast convergence of SGP could be used in the case of regularized problems. Work is in progress in this direction, based on the scaling suggested by the split-gradient method (SGM) [26]. SGP, combined with SGM, has been already successfully applied to a regularized problem arising in the edge-preserving removal of Poisson noise [34]. Since, in the case of regularized problems, one is interested in the minimizer of the functional and not in a semi-convergence property, one can expect that SGP can provide a very efficient and very simple approach.

We conclude by observing that we also performed some comparisons with the *modified residual norm steepest descent* (MRNSD) algorithm and its preconditioned version (PMRNSD) [1]. With respect to GP and SGP, MRNSD exhibits a very poor convergence rate in the case of the nebula, while it behaves similarly to GP in the case of the spacecraft. PMRNSD is much faster and requires a very small number of iterations, but the computational cost per iteration is rather high. For instance, in a comparison based on MATLAB codes, PMRNSD requires approximately the same number of iterations as SGP for the reconstruction of the nebula, but the computational time is about five times higher. In the case of the spacecraft, the computational time of the two algorithms is comparable for $\sigma^2 = 1$ but that of PMRNSD is higher by a factor of 2 for $\sigma^2 = 5$. Moreover, the minimal reconstruction errors provided by PMRNSD are always higher than those provided by SGP.

Finally, in [29] a modification of the conjugate gradient (CG) method, able to impose nonnegativity, is briefly described. The authors claim that a quite small number of iterations is required for obtaining a reliable solution. A preliminary computational study, based on a CG implementation derived from the suggestions given in [29], indicates that SGP has a

convergence rate well comparable with that of the CG method and sometimes is preferable in terms of computational time, thanks to its lower cost per iteration.

Acknowledgments

This work was partially supported by Ministero Università e Ricerca, Italy, grant 2006018748, and by Università Italo-Francese. We thank the referees for their valuable comments.

References

- [1] Bardsley J and Nagy J 2006 Covariance-preconditioned iterative methods for nonnegatively constrained astronomical imaging *SIAM J. Matrix Anal. Appl.* **27** 1184–98
- [2] Barrett H H and Meyers K J 2003 *Foundations of Image Science* (New York: Wiley) pp 1047–8
- [3] Barzilai J and Borwein J M 1988 Two point step size gradient methods *IMA J. Numer. Anal.* **8** 141–8
- [4] Bertero M and Boccacci P 1998 *Introduction to Inverse Problems in Imaging* (Bristol: Institute of Physics Publishing)
- [5] Bertero M, Lanteri H and Zanni L 2008 Iterative image reconstruction: a point of view *Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT) (CRM series)* vol 7, ed Y Censor, M Jiang and A K Louis (Pisa: Edizioni della Normale) pp 37–63
- [6] Bertsekas D P 2003 *Nonlinear Programming* 2nd edn (Belmont: Athena Scientific)
- [7] Birgin E G, Martinez J M and Raydan M 2003 Inexact spectral projected gradient methods on convex sets *IMA J. Numer. Anal.* **23** 539–59
- [8] Bonettini S, Zanella R and Zanni L 2009 A scaled gradient projection method for constrained image deblurring *Inverse Problems* **25** 015002
- [9] Combettes P L and Wajs V R 2005 Signal recovery by proximal forward-backward splitting *Multiscale Model. Simul.* **4** 1168–200
- [10] Dai Y H and Fletcher R 2005 On the asymptotic behaviour of some new gradient methods *Math. Program.* **103** 541–59
- [11] 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
- [12] Dai Y H and Fletcher R 2006 New algorithms for singly linearly constrained quadratic programming problems subject to lower and upper bounds *Math. Program.* **106** 403–21
- [13] Daube-Witherspoon M E and Muehlener G 1986 An iterative image space reconstruction algorithm suitable for volume ECT *IEEE Trans. Med. Imaging* **5** 61–6
- [14] De Pierro A R 1987 On the convergence of the iterative image space reconstruction algorithm for volume ECT *IEEE Trans. Med. Imaging* **6** 174–75
- [15] Eicke B 1992 Iteration methods for convexly constrained ill-posed problems in Hilbert spaces *Numer. Funct. Anal. Optim.* **13** 413–29
- [16] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)
- [17] Figueiredo M A T, Nowak R D and Wright S J 2007 Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems *IEEE J. Sel. Top. Signal Process* **1** 586–97
- [18] Fletcher R 2001 On the Barzilai–Borwein method *Technical Report NA/207* Department of Mathematics, University of Dundee, Dundee, UK
- [19] Frassoldati G, Zanghirati G and Zanni L 2008 New adaptive stepsize selections in gradient methods *J. Ind. Manage. Optim.* **4** 299–312
- [20] Friedlander A, Martínez J M, Molina B and Raydan M 1999 Gradient method with retards and generalizations *SIAM J. Numer. Anal.* **36** 275–89
- [21] Hansen P C, Nagy J G and O’Leary D P 2006 *Deblurring Images: Matrices, Spectra and Filtering* (Philadelphia: SIAM)
- [22] Ivanov V K 1961 On linear problems that are not well-posed *Sov. Math. Dokl.* **3** 981–83
- [23] Jansson P A 1997 *Deconvolution of Images and Spectra* (New York: Academic)
- [24] Kelley C T 1999 *Iterative Methods for Optimization* (Philadelphia: SIAM)
- [25] Lantèri H, Roche M, Cuevas O and Aime C 2001 A general method to devise maximum-likelihood signal restoration multiplicative algorithms with nonnegativity constraints *Signal Process.* **81** 945–74
- [26] Lantèri H, Roche M and Aime C 2002 Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms *Inverse Problems* **18** 1397–419

- [27] Lin C-J 2007 Projected gradient methods for non-negative matrix factorization *Neural Comput.* **19** 2756–99
- [28] Natterer F 1986 *The Mathematics of Computerized Tomography* (New York: Wiley) (reprinted by SIAM, 2001)
- [29] Puetter R C, Gosnell T R and Yahil A 2005 Digital image reconstruction; deblurring and denoising *Annu. Rev. Astron. Astrophys.* **43** 139–94
- [30] Serafini T, Zanghirati G and Zanni L 2005 Gradient projection methods for quadratic programs and applications in training support vector machines *Optim. Methods Softw.* **20** 343–78
- [31] Shepp L A and Vardi Y 1982 Maximum likelihood reconstruction for emission tomography *Trans. Med. Imaging* **1** 113–22
- [32] Tikhonov A N and Arsenin V Y 1977 *Solution of ill-posed problems* (New York: Wiley) (Russian edition in 1974)
- [33] Vogel C R 2002 *Computational Methods for Inverse Problems* (Philadelphia: SIAM)
- [34] Zanella R, Boccacci P, Zanni L and Bertero M 2009 Efficient gradient projection methods for edge-preserving removal of Poisson noise *Inverse Problems* **25** 045010
- [35] Zanni L 2006 An improved gradient projection-based decomposition technique for support vector machines *Comput. Manage. Sci.* **3** 131–45
- [36] Zhou B, Gao L and Dai Y H 2006 Gradient methods with adaptive step-sizes *Comput. Optim. Appl.* **35** 69–86