

A Scaled Gradient Projection Method for Constrained Image Deblurring

S Bonettini¹, R Zanella² and L Zanni²

¹ Dipartimento di Matematica, Università di Ferrara, Polo Scientifico Tecnologico, Blocco B, Via Saragat 1, I-44100 Ferrara, Italy

² Dipartimento di Matematica, Università di Modena e Reggio Emilia, Via Campi 213/B, I-41100 Modena, Italy

E-mail: bntslv@unife.it, riccardo.zanella@unimore.it,
luca.zanni@unimore.it

Abstract. A class of scaled gradient projection methods for optimization problems with simple constraints is considered. These iterative algorithms can be useful in variational approaches to image deblurring that lead to minimize convex nonlinear functions subject to nonnegativity constraints and, in some cases, to an additional flux conservation constraint. A special gradient projection method is introduced that exploits effective scaling strategies and steplength updating rules, appropriately designed for improving the convergence rate. We give convergence results for this scheme and we evaluate its effectiveness by means of an extensive computational study on the minimization problems arising from the maximum likelihood approach to image deblurring. Comparisons with the standard expectation maximization algorithm and with other iterative regularization schemes are also reported to show the computational gain provided by the proposed method.

AMS classification scheme numbers: 65K10, 65F22, 68U10

Submitted to: *Inverse Problems*

Keywords Image deblurring, deconvolution methods, gradient projection methods, large scale optimization.

1. Introduction

In image deblurring problems, image formation is modeled as a Fredholm integral equation of the first kind which, after discretization, results in a system of linear equations. By representing a two dimensional image $\mathcal{X} \in \mathbb{R}^{n \times n}$ as a vector $\mathbf{x} = (x_1, \dots, x_N)^T \in \mathbb{R}^N$, $N = n^2$, in which the entries of \mathcal{X} are stacked column by column, the above system can be stated as

$$A\mathbf{x} = \mathbf{b}, \quad (1)$$

where A is an $N \times N$ matrix representing the physical effect of the imaging system, \mathbf{x} is the image to be reconstructed and $\mathbf{b} \in \mathbb{R}^N$ is the sum of two terms: $\mathbf{b} = \mathbf{g} + \boldsymbol{\eta}$, $\mathbf{g} \in \mathbb{R}^N$ being the blurred image that would have been recorded in absence of noise and $\boldsymbol{\eta} \in \mathbb{R}^N$ denoting the noise affecting the image acquisition [1]. The image restoration problem is then to obtain an approximation of \mathbf{x} , knowing A and \mathbf{b} . Since the system (1) is given by the discretization of an ill-posed problem [1], the matrix A could be very ill-conditioned and a trivial approach that looks for the solution of (1) is in general not successful. Thus, alternative strategies must be exploited, that often consist in iterative schemes motivated by different approaches, ranging from methods for the solution of linear equations to methods for solving variational problems [2].

In this work we propose a scaled gradient projection method suited for the constrained minimization problems arising from several variational approaches to image restoration. Examples of such minimization problems are the following:

$$\begin{aligned} \min \quad & J(\mathbf{x}) \\ \text{sub. to} \quad & \mathbf{x} \geq \mathbf{0}, \end{aligned} \quad (2)$$

or

$$\begin{aligned} \min \quad & J(\mathbf{x}) \\ \text{sub. to} \quad & \mathbf{x} \geq \mathbf{0} \\ & \sum_{i=1}^N x_i = c, \end{aligned} \quad (3)$$

where $J(\mathbf{x})$ is a continuously differentiable convex function measuring the difference between reconstructed and measured data and, possibly, containing a penalty term expressing additional information on the solution, while the constraints force the nonnegativity of the solution and, in case of the problem (3), the so-called flux conservation property. Gradient projection type methods seem appealing approaches for these problems for two main reasons. Firstly, the special structure of the constraints makes the projection of a vector on the feasible region a non-excessively expensive operation: this is obviously the case of the problem (2), but also of the problem (3), in which the projection can be performed by linear-time algorithms [3, 4, 5, 6, 7]. Secondly, the recent advances on the steplength selection in gradient methods [8, 4, 9, 10] allow to largely improve the convergence rate of these schemes, without introducing significant additional costs. Thus, new gradient projection methods can nowadays be designed that, thanks to the low computational cost per iteration and the good convergence

rate, may represent a valid alternative to other gradient-based iterative approaches widely used in image restoration [11, 12, 13, 14, 15]. The main feature of the gradient projection method introduced in this paper consists in the combination of non-expensive diagonally scaled gradient directions with steplength selection rules specially designed for these directions. Moreover, global convergence properties are ensured by exploiting a nonmonotone line-search strategy along the feasible direction [16, 17]. Scaled gradient directions are also exploited by other popular algorithms for image restoration; see, for example, the projected Newton methods described in [2, 18, 19, 20]. However, these schemes are substantially different from our approach since they require inner linear solvers to compute the non-diagonally scaled gradient direction, do not consider steplength selection strategies and use line-search along the projection arc instead of along the feasible direction [21, p. 226].

The effectiveness of the proposed scheme is evaluated by solving the minimization problems arising from the maximum likelihood approach to the deconvolution of images corrupted by Poisson noise, that is when the objective function in (2) and (3) is the Kullback-Leibler divergence of the blurred image $A\mathbf{x}$ from the observed noisy image \mathbf{b} [1, 12, 15, 22]. Comparisons with the well known Expectation Maximization (EM) method [15], the accelerated EM version proposed in [23] and the Weighted Modified Residual Norm Steepest Descent (WMRNSD) algorithm introduced in [24] are reported to assess the reconstruction accuracy and the computational gain provided by the new method.

The paper is organized as follows: in section 2 the scaled gradient projection approach is presented and its global convergence theory is developed, while in section 3 crucial details for a practical implementation of the method are given. An extensive numerical experimentation on astronomical test images is presented in section 4. Our conclusions and future developments are discussed in section 5.

2. The algorithm and its convergence

We introduce a Scaled Gradient Projection (SGP) method for solving constrained minimization problems of the general form

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{sub. to} \quad & \mathbf{x} \in \Omega, \end{aligned} \tag{4}$$

where $\Omega \subset \mathbb{R}^N$ is a closed convex set and $f : \Omega \rightarrow \mathbb{R}$ is a continuously differentiable function. Obviously, the deblurring problems (2) and (3) can be considered as special cases of this formulation.

Before to state the SGP algorithm we recall some basic properties of the projection operator.

2.1. Definitions and basic properties

Throughout the paper, the 2-norm of vectors and matrices is denoted by $\|\cdot\|$ while $\|\cdot\|_D$ indicates the vector norm associated to a symmetric positive definite matrix D : $\|\mathbf{x}\|_D = \sqrt{\mathbf{x}^T D \mathbf{x}}$.

Given the optimization problem (4), we recall that $\mathbf{x}_* \in \Omega$ is a stationary point of f over Ω if

$$-\nabla f(\mathbf{x}_*)^T (\mathbf{y} - \mathbf{x}_*) \leq 0, \quad \forall \mathbf{y} \in \Omega.$$

When Ω is convex, the stationarity condition can be stated also as $-\nabla f(\mathbf{x}_*)^T \mathbf{w} \leq 0$ for any \mathbf{w} in the tangent cone of Ω at \mathbf{x}_* [21, p. 336].

Let $\Omega \subset \mathbb{R}^N$ be a closed convex set and D be a symmetric positive definite $N \times N$ matrix, we define the projection operator $P_{\Omega, D} : \mathbb{R}^N \rightarrow \Omega$ as

$$P_{\Omega, D}(\mathbf{x}) \equiv \arg \min_{\mathbf{y} \in \Omega} \|\mathbf{y} - \mathbf{x}\|_D = \arg \min_{\mathbf{y} \in \Omega} \left(\phi(\mathbf{y}) \equiv \frac{1}{2} \mathbf{y}^T D \mathbf{y} - \mathbf{y}^T D \mathbf{x} \right). \quad (5)$$

We observe that, given the set Ω and the point \mathbf{x} , the operator $P_{\Omega, D}(\mathbf{x})$ is a continuous function with respect to the elements of the matrix D . Furthermore, from the definition of stationary point and the strict convexity of the function ϕ introduced in (5), we have that $P_{\Omega, D}(\mathbf{x})$ is defined also by

$$(P_{\Omega, D}(\mathbf{x}) - \mathbf{x})^T D (P_{\Omega, D}(\mathbf{x}) - \mathbf{y}) \leq 0, \quad \forall \mathbf{y} \in \Omega. \quad (6)$$

Let $\mathcal{D}_L \subset \mathbb{R}^{N \times N}$ be the compact set of the symmetric positive definite $N \times N$ matrices such that $\|D\| \leq L$ and $\|D^{-1}\| \leq L$, for a given threshold $L > 1$. The next lemmas state two basic properties related to the projection operator: a Lipschitz continuity condition on $P_{\Omega, D}$ and a characterization for the stationary points of the problem (4) [25].

Lemma 2.1 *If $D \in \mathcal{D}_L$, then*

$$\|P_{\Omega, D}(\mathbf{x}) - P_{\Omega, D}(\mathbf{z})\| \leq L^2 \|\mathbf{x} - \mathbf{z}\| \quad (7)$$

for any $\mathbf{x}, \mathbf{z} \in \mathbb{R}^N$.

Proof. By applying the condition (6) we get

$$\begin{aligned} (P_{\Omega, D}(\mathbf{x}) - \mathbf{x})^T D (P_{\Omega, D}(\mathbf{x}) - P_{\Omega, D}(\mathbf{z})) &\leq 0 \\ (P_{\Omega, D}(\mathbf{z}) - \mathbf{z})^T D (P_{\Omega, D}(\mathbf{z}) - P_{\Omega, D}(\mathbf{x})) &\leq 0 \end{aligned}$$

and, by adding the two inequalities,

$$((P_{\Omega, D}(\mathbf{x}) - \mathbf{x}) - (P_{\Omega, D}(\mathbf{z}) - \mathbf{z}))^T D (P_{\Omega, D}(\mathbf{x}) - P_{\Omega, D}(\mathbf{z})) \leq 0,$$

that is

$$\|P_{\Omega, D}(\mathbf{x}) - P_{\Omega, D}(\mathbf{z})\|_D^2 \leq (P_{\Omega, D}(\mathbf{x}) - P_{\Omega, D}(\mathbf{z}))^T D (\mathbf{x} - \mathbf{z}). \quad (8)$$

If σ_{min} denotes the minimum eigenvalue of the matrix D , for the left hand side of the previous inequality we have

$$\begin{aligned} \|P_{\Omega,D}(\mathbf{x}) - P_{\Omega,D}(\mathbf{z})\|_D^2 &\geq \sigma_{min} \|P_{\Omega,D}(\mathbf{x}) - P_{\Omega,D}(\mathbf{z})\|^2 \\ &= \frac{1}{\|D^{-1}\|} \|P_{\Omega,D}(\mathbf{x}) - P_{\Omega,D}(\mathbf{z})\|^2 \\ &\geq \frac{1}{L} \|P_{\Omega,D}(\mathbf{x}) - P_{\Omega,D}(\mathbf{z})\|^2 \end{aligned}$$

and from (8) we obtain

$$\begin{aligned} \frac{1}{L} \|P_{\Omega,D}(\mathbf{x}) - P_{\Omega,D}(\mathbf{z})\|^2 &\leq \|P_{\Omega,D}(\mathbf{x}) - P_{\Omega,D}(\mathbf{z})\|_D^2 \\ &\leq (P_{\Omega,D}(\mathbf{x}) - P_{\Omega,D}(\mathbf{z}))^T D (\mathbf{x} - \mathbf{z}) \\ &\leq \|P_{\Omega,D}(\mathbf{x}) - P_{\Omega,D}(\mathbf{z})\| \|D\| \|\mathbf{x} - \mathbf{z}\| \\ &\leq L \|P_{\Omega,D}(\mathbf{x}) - P_{\Omega,D}(\mathbf{z})\| \|\mathbf{x} - \mathbf{z}\| \end{aligned}$$

which yields (7). \square

Lemma 2.2 *A vector $\mathbf{x}_* \in \Omega$ is a stationary point of the problem (4) if and only if $\mathbf{x}_* = P_{\Omega,D^{-1}}(\mathbf{x}_* - \alpha D \nabla f(\mathbf{x}_*))$ for any positive scalar α and for any symmetric positive definite matrix D .*

Proof. Let $\alpha \in \mathbb{R}^+$ and let D be a symmetric positive definite matrix. Assume that $\mathbf{x}_* = P_{\Omega,D^{-1}}(\mathbf{x}_* - \alpha D \nabla f(\mathbf{x}_*))$. From (6) we obtain

$$(\mathbf{x}_* - \mathbf{x}_* + \alpha D \nabla f(\mathbf{x}_*))^T D^{-1} (\mathbf{x}_* - \mathbf{x}) \leq 0, \quad \forall \mathbf{x} \in \Omega,$$

which implies the stationarity condition

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

Conversely, let us assume that $\mathbf{x}_* \in \Omega$ is a stationary point of (4), and suppose that $\bar{\mathbf{x}} = P_{\Omega,D^{-1}}(\mathbf{x}_* - \alpha D \nabla f(\mathbf{x}_*))$, with $\bar{\mathbf{x}} \neq \mathbf{x}_*$. Then, from (6) we can write

$$(\bar{\mathbf{x}} - \mathbf{x}_* + \alpha D \nabla f(\mathbf{x}_*))^T D^{-1} (\bar{\mathbf{x}} - \mathbf{x}_*) \leq 0,$$

that is

$$\|\bar{\mathbf{x}} - \mathbf{x}_*\|_{D^{-1}}^2 + \alpha \nabla f(\mathbf{x}_*)^T (\bar{\mathbf{x}} - \mathbf{x}_*) \leq 0.$$

The previous inequality yields

$$\nabla f(\mathbf{x}_*)^T (\mathbf{x}_* - \bar{\mathbf{x}}) \geq \frac{\|\bar{\mathbf{x}} - \mathbf{x}_*\|_{D^{-1}}^2}{\alpha} > 0,$$

which gives a contradiction with the stationarity assumption on \mathbf{x}_* . \square

Algorithm SGP (Scaled Gradient Projection Method)

Choose the starting point $\mathbf{x}^{(0)} \in \Omega$, set the parameters $\beta, \theta \in (0, 1)$, $0 < \alpha_{min} < \alpha_{max}$ and fix a positive integer M .

FOR $k = 0, 1, 2, \dots$ DO THE FOLLOWING STEPS:

STEP 1. Choose the parameter $\alpha_k \in [\alpha_{min}, \alpha_{max}]$ and the scaling matrix $D_k \in \mathcal{D}_L$;

STEP 2. Projection: $\mathbf{y}^{(k)} = P_{\Omega, D_k^{-1}}(\mathbf{x}^{(k)} - \alpha_k D_k \nabla f(\mathbf{x}^{(k)}))$;

If $\mathbf{y}^{(k)} = \mathbf{x}^{(k)}$ then stop, declaring that $\mathbf{x}^{(k)}$ is a stationary point;

STEP 3. Descent direction: $\mathbf{d}^{(k)} = \mathbf{y}^{(k)} - \mathbf{x}^{(k)}$;

STEP 4. Set $\lambda_k = 1$ and $f_{max} = \max_{0 \leq j \leq \min(k, M-1)} f(\mathbf{x}^{(k-j)})$;

STEP 5. Backtracking loop:

IF $f(\mathbf{x}^{(k)} + \lambda_k \mathbf{d}^{(k)}) \leq f_{max} + \beta \lambda_k \nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)}$ THEN

go to Step 6;

ELSE

set $\lambda_k = \theta \lambda_k$ and go to Step 5;

ENDIF

STEP 6. Set $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda_k \mathbf{d}^{(k)}$.

END

2.2. The SGP method

The Lemma 2.2 shows the effect of the projection operator $P_{\Omega, D^{-1}}$ on the points $(\mathbf{x}_* - \alpha D \nabla f(\mathbf{x}_*))$, $\alpha > 0$, when \mathbf{x}_* is a stationary point of (4). In the case $\bar{\mathbf{x}} \in \Omega$ is a nonstationary point, $P_{\Omega, D^{-1}}(\bar{\mathbf{x}} - \alpha D \nabla f(\bar{\mathbf{x}}))$ can be exploited to generate a descent direction for the function f in $\bar{\mathbf{x}}$. This idea serves as the basis for the method described in Algorithm SGP.

Before to discuss the convergence properties of the method, some considerations about its main steps can be useful.

First of all, it is worth stressing that any choice of the steplength α_k in a closed interval and of the scaling matrix D_k in the compact set \mathcal{D}_L is permitted. This is very important from a practical point of view since it allows one to make the updating rules of α_k and D_k problem related and oriented at optimizing the performance. Refer to the next section for a study on the selection strategies of these parameters in the case of the deblurring problems (2) and (3).

If the projection performed in step 2 returns a vector $\mathbf{y}^{(k)}$ equal to $\mathbf{x}^{(k)}$, then Lemma 2.2 implies that $\mathbf{x}^{(k)}$ is a stationary point and the algorithm stops. When $\mathbf{y}^{(k)} \neq \mathbf{x}^{(k)}$, it is possible to prove that \mathbf{d}_k is a descent direction for f in $\mathbf{x}^{(k)}$ (see the next lemma) and the backtracking loop in step 5 terminates with a finite number of runs; thus the algorithm is well defined.

The nonmonotone line-search strategy implemented in step 5 ensures that $f(\mathbf{x}^{(k+1)})$ is lower than the maximum of the objective function on the last M iterations [17]; of course, if $M = 1$ then the strategy reduces to the standard monotone Armijo rule [21].

Finally, we remark the relation between SGP and the projected gradient method proposed in [26]. We observe that

$$\begin{aligned} \mathbf{d}^{(k)} &= \mathbf{y}^{(k)} - \mathbf{x}^{(k)} = P_{\Omega, D_k^{-1}}(\mathbf{x}^{(k)} - \alpha_k D_k \nabla f(\mathbf{x}^{(k)})) - \mathbf{x}^{(k)} \\ &= \left(\arg \min_{\mathbf{y} \in \Omega} \frac{1}{2} \mathbf{y}^T D_k^{-1} \mathbf{y} - \mathbf{y}^T D_k^{-1} (\mathbf{x}^{(k)} - \alpha_k D_k \nabla f(\mathbf{x}^{(k)})) \right) - \mathbf{x}^{(k)} \end{aligned}$$

and, by introducing a new variable \mathbf{d} such that $\mathbf{y} = \mathbf{x}^{(k)} + \mathbf{d}$, we may rewrite $\mathbf{d}^{(k)}$ in an equivalent form:

$$\mathbf{d}^{(k)} = \arg \min_{\mathbf{x}^{(k)} + \mathbf{d} \in \Omega} \frac{1}{2} \mathbf{d}^T B_k \mathbf{d} + \nabla f(\mathbf{x}^{(k)})^T \mathbf{d}, \quad (9)$$

where $B_k = \frac{D_k^{-1}}{\alpha_k}$. Since the general method proposed in [26] allows to define the search direction as in (9) (or even by means of an inexact solution of the problem (9)), the SGP algorithm could be seen as an alternative formulation of that scheme in which the steplength and the scaling matrix are managed separately, with the aim to better emphasize their contribution and to simplify the design of effective updating rules for these parameters. This means also that the convergence results obtained in [26] could be adapted to our method; however, for the sake of completeness, we develop a slightly different convergence analysis, appropriately designed for SGP.

2.3. A convergence analysis for SGP

In this subsection we will focus on the case in which the algorithm generates an infinite sequence of iterates, denoted by $\{\mathbf{x}^{(k)}\}$. The main SGP convergence result is stated in Theorem 2.1, whose proof is based on some crucial properties that we report in the next lemmas.

The first two lemmas are concerned with the descent condition and the boundedness of the directions $\mathbf{d}^{(k)}$, respectively.

Lemma 2.3 *Assume that $\mathbf{d}^{(k)} \neq \mathbf{0}$. Then, $\mathbf{d}^{(k)}$ is a descent direction for the function f at $\mathbf{x}^{(k)}$, that is, $\nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} < 0$.*

Proof. From the inequality (6) with $\mathbf{x} = \mathbf{x}^{(k)} - \alpha_k D_k \nabla f(\mathbf{x}^{(k)})$, $D = D_k^{-1}$ and $\mathbf{y} = \mathbf{x}^{(k)}$, it follows that

$$(\mathbf{d}^{(k)} + \alpha_k D_k \nabla f(\mathbf{x}^{(k)}))^T D_k^{-1} \mathbf{d}^{(k)} \leq 0,$$

and then

$$\nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} \leq -\frac{\mathbf{d}^{(k)T} D_k^{-1} \mathbf{d}^{(k)}}{\alpha_k} < 0. \quad (10)$$

□

Lemma 2.4 *If the sequence $\{\mathbf{x}^{(k)}\}$ is bounded, then also the sequence $\{\mathbf{d}^{(k)}\}$ is bounded.*

Proof. From the definition of $\mathbf{d}^{(k)}$ and (7) we have that, for any k ,

$$\begin{aligned}\|\mathbf{d}^{(k)}\| &= \|P_{\Omega, D_k^{-1}}(\mathbf{x}^{(k)} - \alpha_k D_k \nabla f(\mathbf{x}^{(k)})) - \mathbf{x}^{(k)}\| \\ &= \|P_{\Omega, D_k^{-1}}(\mathbf{x}^{(k)} - \alpha_k D_k \nabla f(\mathbf{x}^{(k)})) - P_{\Omega, D_k^{-1}}(\mathbf{x}^{(k)})\| \\ &\leq L^2 \|\alpha_k D_k \nabla f(\mathbf{x}^{(k)})\| \leq \alpha_{max} L^3 \|\nabla f(\mathbf{x}^{(k)})\|.\end{aligned}$$

Let $\bar{\Omega} \subset \Omega$ be a closed and bounded set that contains the iterates $\mathbf{x}^{(k)}$. Since ∇f is a continuous function on Ω , then it is bounded in $\bar{\Omega}$ and thus $\{\mathbf{d}^{(k)}\}$ is bounded. \square

Now we prove two properties of the accumulation points of the sequence generated by SGP.

Lemma 2.5 *Assume that the subsequence $\{\mathbf{x}^{(k)}\}_{k \in K}$, $K \subset \mathbb{N}$, is converging to a point $\mathbf{x}_* \in \Omega$. Then, \mathbf{x}_* is a stationary point of (4) if and only if*

$$\lim_{k \in K} \nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} = 0.$$

Proof. Let \mathbf{x}_* be a stationary point of (4); this means that $\nabla f(\mathbf{x}_*)^T \mathbf{d} \geq 0$ for any vector \mathbf{d} such that $\mathbf{x}_* + \mathbf{d} \in \Omega$. Suppose that $\{\nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)}\}$ does not tend to 0 for $k \in K$. In this case, taking into account Lemma 2.3, we know that there exists $\varepsilon > 0$ and an infinite set $K_1 \subset K$ such that

$$\nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} \leq -\varepsilon < 0, \quad \forall k \in K_1.$$

By the compactness of the interval $[\alpha_{min}, \alpha_{max}]$ and of the set \mathcal{D}_L , we can extract a set of indices $K_2 \subset K_1$ such that $\alpha_k \rightarrow \alpha_*$, $\alpha_* \in [\alpha_{min}, \alpha_{max}]$, and $D_k \rightarrow D_*$, $D_* \in \mathcal{D}_L$, for $k \in K_2$; hence, by continuity, we can write $\lim_{k \in K_2} \mathbf{d}^{(k)} = \mathbf{d}_*$, where

$$\mathbf{d}_* = P_{\Omega, D_*^{-1}}(\mathbf{x}_* - \alpha_* D_* \nabla f(\mathbf{x}_*)) - \mathbf{x}_*. \quad (11)$$

Thus,

$$\lim_{k \in K_2} \nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} = \nabla f(\mathbf{x}_*)^T \mathbf{d}_* \leq -\varepsilon < 0. \quad (12)$$

Since from by the definition (11) we have that $\mathbf{x}_* + \mathbf{d}_*$ belongs to Ω , the inequality (12) contradicts the stationarity assumption on \mathbf{x}_* .

On the other hand, let us assume that $\lim_{k \in K} \nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} = 0$. Suppose by contradiction that \mathbf{x}_* is not a stationary point. Let $K_3 \subset K$ be a set of indices such that $\alpha_k \rightarrow \alpha_*$ and $D_k \rightarrow D_*$, when k diverges, $k \in K_3$; we have $\lim_{k \in K_3} \mathbf{d}^{(k)} = (P_{\Omega, D_*^{-1}}(\mathbf{x}_* - \alpha_* D_* \nabla f(\mathbf{x}_*)) - \mathbf{x}_*)$. Furthermore, from Lemma 2.2 there exists $\delta > 0$ such that $\|P_{\Omega, D_*^{-1}}(\mathbf{x}_* - \alpha_* D_* \nabla f(\mathbf{x}_*)) - \mathbf{x}_*\|^2 = \delta$. By exploiting (10), we can write, for a sufficiently large $\bar{k} \in K_3$,

$$\nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} \leq -\frac{\mathbf{d}^{(k)T} D_k^{-1} \mathbf{d}^{(k)}}{\alpha_k} \leq -\frac{\delta}{2\alpha_{max} L} < 0, \quad \forall k \geq \bar{k}, \quad k \in K_3.$$

This contradicts the assumption $\lim_{k \in K} \nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} = 0$ and then \mathbf{x}_* must be a stationary point. \square

Lemma 2.6 *Let $\mathbf{x}_* \in \Omega$ be an accumulation point of the sequence $\{\mathbf{x}^{(k)}\}$ such that $\lim_{k \in K} \mathbf{x}^{(k)} = \mathbf{x}_*$, for some $K \subset \mathbb{N}$. If \mathbf{x}_* is a stationary point of (4), then \mathbf{x}_* is an accumulation point also for the sequence $\{\mathbf{x}^{(k+r)}\}_{k \in K}$ for any $r \in \mathbb{N}$. Furthermore,*

$$\lim_{k \in K} \|\mathbf{d}^{(k+r)}\| = 0, \quad \forall r \in \mathbb{N}.$$

Proof. From Lemma 2.5 we have that $\lim_{k \in K} \nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} = 0$ and, from (10), we obtain that $\lim_{k \in K} \|\mathbf{d}^{(k)}\| = 0$. Thus $\lim_{k \in K} \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| = 0$, and this implies that \mathbf{x}_* is an accumulation point also for the sequence $\{\mathbf{x}^{(k+1)}\}_{k \in K}$. Recalling again Lemma 2.5, we obtain that $\lim_{k \in K} \nabla f(\mathbf{x}^{(k+1)})^T \mathbf{d}^{(k+1)} = 0$; for the same reasons as before, we conclude that $\lim_{k \in K} \|\mathbf{d}^{(k+1)}\| = 0$. Hence, the statement of the lemma follows by induction. \square

At this point we may state a convergence result for SGP.

Theorem 2.1 *Assume that the level set $\Omega_0 = \{\mathbf{x} \in \Omega : f(\mathbf{x}) \leq f(\mathbf{x}^{(0)})\}$ is bounded. Every accumulation point of the sequence $\{\mathbf{x}^{(k)}\}$ generated by the SGP algorithm is a stationary point of (4).*

Proof. Since every iterate $\mathbf{x}^{(k)}$ lies in Ω_0 , the sequence $\{\mathbf{x}^{(k)}\}$ is bounded and has at least one accumulation point. Let $\mathbf{x}_* \in \Omega$ be such that $\lim_{k \in K} \mathbf{x}^{(k)} = \mathbf{x}_*$ for a set of indices $K \subset \mathbb{N}$. Let us consider separately the two cases

- a. $\inf_{k \in K} \lambda_k = 0$;
- b. $\inf_{k \in K} \lambda_k = \rho > 0$.

Case a.

Let $K_1 \subset K$ be a set of indices such that $\lim_{k \in K_1} \lambda_k = 0$. This implies that, for $k \in K_1$, k sufficiently large, the backtracking rule fails to be satisfied at least once. Thus, at the penultimate step of the backtracking loop, we have

$$f(\mathbf{x}^{(k)} + \frac{\lambda_k}{\theta} \mathbf{d}^{(k)}) > f(\mathbf{x}^{(k)}) + \beta \frac{\lambda_k}{\theta} \nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)},$$

hence

$$\frac{f(\mathbf{x}^{(k)} + \frac{\lambda_k}{\theta} \mathbf{d}^{(k)}) - f(\mathbf{x}^{(k)})}{\frac{\lambda_k}{\theta}} > \beta \nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)}. \quad (13)$$

By the mean value theorem, we have that there exists a scalar $t_k \in [0, \frac{\lambda_k}{\theta}]$ such that the left hand side of (13) is equal to $\nabla f(\mathbf{x}^{(k)} + t_k \mathbf{d}^{(k)})^T \mathbf{d}^{(k)}$. Thus, the inequality (13) becomes

$$\nabla f(\mathbf{x}^{(k)} + t_k \mathbf{d}^{(k)})^T \mathbf{d}^{(k)} > \beta \nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)}. \quad (14)$$

Since α_k and D_k are bounded, it is possible to find a set of indices $K_2 \subset K_1$ such that $\lim_{k \in K_2} \alpha_k = \alpha_*$ and $\lim_{k \in K_2} D_k = D_*$. Thus the sequence $\{\mathbf{d}^{(k)}\}_{k \in K_2}$ converges to the

vector $\mathbf{d}_* = (P_{\Omega, D_*^{-1}}(\mathbf{x}_* - \alpha_* D_* \nabla f(\mathbf{x}_*)) - \mathbf{x}_*)$ and, furthermore, $t_k \mathbf{d}^{(k)} \rightarrow 0$ when k diverges, $k \in K_2$. Taking limits in (14) as $k \rightarrow \infty$, $k \in K_2$, we obtain

$$(1 - \beta) \nabla f(\mathbf{x}_*)^T \mathbf{d}_* \geq 0.$$

Since $(1 - \beta) > 0$ and $\nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} < 0$ for all k , then we necessarily have $\lim_{k \in K_2} \nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} = \nabla f(\mathbf{x}_*)^T \mathbf{d}_* = 0$. Then, by Lemma 2.5, we conclude that \mathbf{x}_* is a stationary point.

Case b.

Let us define the point $\mathbf{x}^{(\ell(k))}$ as the point such that

$$f(\mathbf{x}^{(\ell(k))}) = f_{max} = \max_{0 \leq j \leq \min(k, M-1)} f(\mathbf{x}^{(k-j)}).$$

Then, for $k > M - 1$, $k \in \mathbb{N}$, the following condition holds:

$$f(\mathbf{x}^{(\ell(k))}) \leq f(\mathbf{x}^{(\ell(k)-1)}) + \beta \lambda_{\ell(k)-1} \nabla f(\mathbf{x}^{(\ell(k)-1)})^T \mathbf{d}^{(\ell(k)-1)}. \quad (15)$$

Since the iterates $\mathbf{x}^{(k)}$, $k \in \mathbb{N}$ belong to a bounded set, the monotone non-increasing sequence $\{f(\mathbf{x}^{(\ell(k))})\}$ admits a finite limit $\mathcal{L} \in \mathbb{R}$ for $k \in K$. Let $K_3 \subset K$ be a set of indices such that $\lim_{k \in K_3} \lambda_{\ell(k)-1} = \rho_1 \geq \rho > 0$ and $\lim_{k \in K_3} \nabla f(\mathbf{x}^{(\ell(k)-1)})^T \mathbf{d}^{(\ell(k)-1)}$ exists (recall that, from Lemma 2.4, the sequence $\{\mathbf{d}^{(k)}\}_{k \in \mathbb{N}}$ is bounded); taking limits on (15) for $k \in K_3$ we obtain

$$\mathcal{L} \leq \mathcal{L} + \beta \rho_1 \lim_{k \in K_3} \nabla f(\mathbf{x}^{(\ell(k)-1)})^T \mathbf{d}^{(\ell(k)-1)},$$

that is

$$\lim_{k \in K_3} \nabla f(\mathbf{x}^{(\ell(k)-1)})^T \mathbf{d}^{(\ell(k)-1)} \geq 0.$$

Recalling that $\nabla f(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} < 0$ for any k , the previous inequality implies that

$$\lim_{k \in K_3} \nabla f(\mathbf{x}^{(\ell(k)-1)})^T \mathbf{d}^{(\ell(k)-1)} = 0. \quad (16)$$

Then, by the Lemma 2.5, (16) implies that every accumulation point of the sequence $\{\mathbf{x}^{(\ell(k)-1)}\}_{k \in K_3}$ is a stationary point of (4).

Let us prove that the point \mathbf{x}_* is an accumulation point of $\{\mathbf{x}^{(\ell(k)-1)}\}_{k \in K_3}$.

The definition of $\mathbf{x}^{(\ell(k))}$ implies that $k - M + 1 \leq \ell(k) \leq k$. Thus we can write

$$\|\mathbf{x}^{(k)} - \mathbf{x}^{(\ell(k)-1)}\| \leq \sum_{j=0}^{k-\ell(k)} \lambda_{\ell(k)-1+j} \|\mathbf{d}^{(\ell(k)-1+j)}\|, \quad k \in K. \quad (17)$$

Let $K_4 \subset K_3$ be a subset of indices such that the sequence $\{\mathbf{x}^{(\ell(k)-1)}\}_{k \in K_4}$ converges to an accumulation point $\bar{\mathbf{x}} \in \Omega$. Recalling that, from (16) and Lemma 2.5, $\bar{\mathbf{x}}$ is a stationary point of (4), we can apply Lemma 2.6 to obtain that $\lim_{k \in K_4} \|\mathbf{d}^{(\ell(k)-1+j)}\| = 0$ for any $j \in \mathbb{N}$. By using (17) we conclude that

$$\lim_{k \in K_4} \|\mathbf{x}^{(k)} - \mathbf{x}^{(\ell(k)-1)}\| = 0. \quad (18)$$

Since

$$\|\mathbf{x}_* - \mathbf{x}^{(\ell(k)-1)}\| \leq \|\mathbf{x}^{(k)} - \mathbf{x}^{(\ell(k)-1)}\| + \|\mathbf{x}^{(k)} - \mathbf{x}_*\|$$

and $\lim_{k \in K} \mathbf{x}^{(k)} = \mathbf{x}_*$, then (18) implies that \mathbf{x}_* is an accumulation point also for the sequence $\{\mathbf{x}^{(\ell(k)-1)}\}_{k \in K_3}$. Hence, we conclude that \mathbf{x}_* is a stationary point of (4). \square

3. The SGP method for image deblurring

In this section we describe SGP implementations for solving special constrained minimization problems arising in image deblurring. We consider the maximum likelihood approach to the deconvolution of images corrupted by Poisson noise [1, 12, 15]. It is well known that in this approach an estimate of the unknown image is obtained by approximating the solutions of minimization problems of the form (2) or (3), in which, following the notation used in the introduction, the objective function is the Kullback-Leibler divergence of the blurred image $A\mathbf{x}$ from the observed noisy image \mathbf{b} . In particular, we deal with models that take into account also the existence of a constant background radiation $bg > 0$ and, consequently, we must look for approximate solutions of (2) or (3) when $J(\mathbf{x})$ is the Kullback-Leibler divergence of $(A\mathbf{x} + bg)$ from \mathbf{b} :

$$\begin{aligned} J(\mathbf{x}) &= D_{KL}(A\mathbf{x} + bg, \mathbf{b}) = \\ &= \sum_{i=1}^N \left(\sum_{j=1}^N A_{ij}x_j + bg - b_i - b_i \log \frac{\sum_{j=1}^N A_{ij}x_j + bg}{b_i} \right). \end{aligned} \quad (19)$$

We remark that, due to the ill-posedness of the image restoration problem, when $J(\mathbf{x})$ is defined as in (19), one is not interested in computing the exact solution of (2) or (3), because this does not provide a sensible estimate of the unknown image; for this reason, the iterative minimization methods are usually exploited to obtain acceptable (regularized) solutions by early stopping. A different approach to image restoration consists in minimizing special regularized functionals given by the sum of a fit-to-data function and a penalty term; in this case the trade-off between data fidelity and stability is controlled by means of a regularization parameter and a suited image restoration is obtained by an exact solution of the minimization problem. Even if SGP can be used in this context, we do not consider such approaches here and we refer to [19, 20, 27] for examples of minimization methods applied to these problems.

Some comments about the objective function (19) are necessary for the subsequent discussions. The gradient and the hessian of $J(\mathbf{x})$ can be written as

$$\nabla J(\mathbf{x}) = A^T \mathbf{e} - A^T Y^{-1} \mathbf{b} \quad (20)$$

$$\nabla^2 J(\mathbf{x}) = A^T B Y^{-2} A, \quad (21)$$

where $\mathbf{e} \in \mathbb{R}^N$ is a vector whose components are all equal to one, $Y = \text{diag}(A\mathbf{x} + bg)$ is a diagonal matrix with the entries of $(A\mathbf{x} + bg)$ on the main diagonal and $B = \text{diag}(\mathbf{b})$. The matrix A can be considered with nonnegative entries, generally dense and such that $A^T \mathbf{e} = \mathbf{e}$ and $\sum_j A_{ij} > 0$, $\forall i$; moreover, we may assume that periodic boundary conditions are imposed for the discretization of the Fredholm integral equation that models the image formation process, so that the matrix A is block-circulant with circulant blocks. Other boundary conditions can be used, but the crucial point is that matrix-vector products can be done quickly. For example, one can replace periodic boundary conditions with zero, reflexive or anti-reflexive boundary conditions. In all of these cases the products $A\mathbf{x}$ can be performed with $\mathcal{O}(N \log N)$ complexity, by employing the Fast Fourier Transform (FFT) [28], or a fast trigonometric transform

such as the Discrete Cosine Transform (DCT) or Discrete Sine Transform (DST) [29]. Concerning the components b_i of the observed image, we remark that they are nonnegative; this implies that the hessian matrix (21) is positive semidefinite in any point of the nonnegative orthant and the problems (2) and (3) are convex minimization problems. Thus, when $J(\mathbf{x})$ is defined as in (19), each accumulation point of the sequence generated by applying SGP to (2) or (3) is a solution of the problem (we recall that the assumption of Theorem 2.1 is trivially satisfied for these problems). Concerning the computational cost of SGP in this particular application, we observe that the heaviest tasks in each iteration are two matrix-vector products useful to compute $\nabla J(\mathbf{x}^{(k)})$ in Step 2 and $J(\mathbf{x}^{(k)} + \lambda_k \mathbf{d}^{(k)})$ in Step 5; that is, two (two-dimensional) FFT/IFFT pairs ($\mathcal{O}(N \log N)$ complexity). It is possible to implement the other significant operations in each iteration in such a way that their costs are no larger than those of the above products. To this end, in the following subsections we show how the projection on the feasible region of (3) can be performed and how the scaling matrix D_k and the steplength parameter α_k can be updated.

3.1. Compute the projection

From now on we will assume that the scaling matrix D_k is diagonal, $D_k = \text{diag}(d_1^{(k)}, d_2^{(k)}, \dots, d_N^{(k)})$, as usually done in scaled gradient projection methods; some examples of diagonal scaling matrices will be given in the next subsection.

Under this assumption, the projections arising from the application of SGP to the problem (2) are trivial operations and do not require special investigations.

Let $\bar{\Omega}$ be the feasible region of the minimization problem (3):

$$\bar{\Omega} = \left\{ \mathbf{x} \in \mathbb{R}^N \mid \mathbf{x} \geq \mathbf{0}, \sum_{i=1}^N x_i = c \right\}. \quad (22)$$

When algorithm SGP is applied for solving (3), at each iteration we need to compute $P_{\bar{\Omega}, D_k^{-1}}(\mathbf{x}^{(k)} - \alpha_k D_k \nabla f(\mathbf{x}^{(k)}))$, that is, we must solve the constrained strictly convex quadratic program

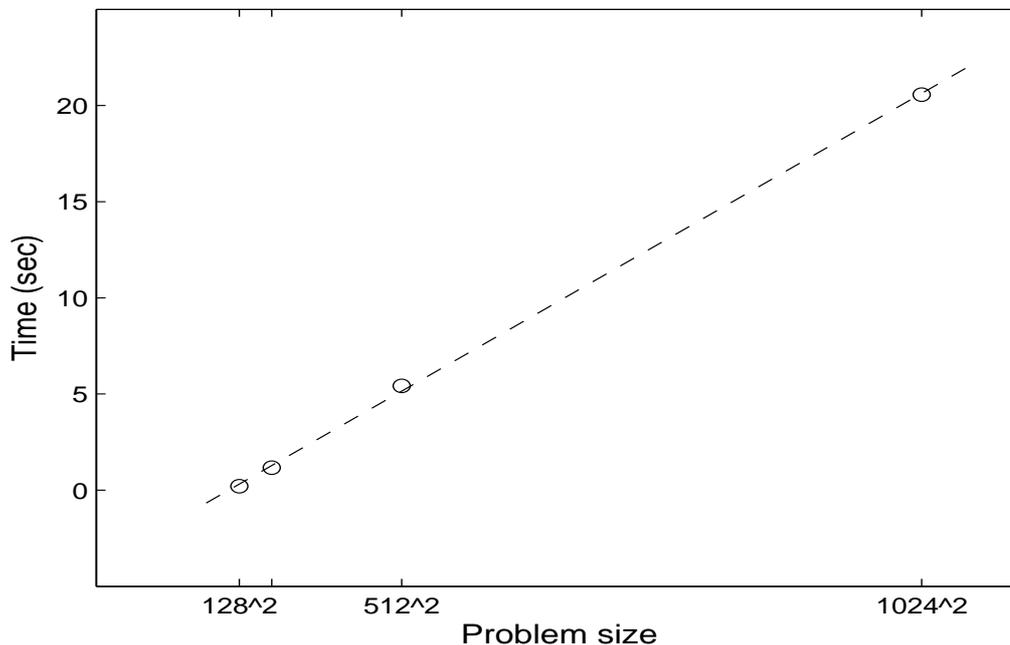
$$\begin{aligned} \min \quad & \frac{1}{2} \mathbf{x}^T D_k^{-1} \mathbf{x} - \mathbf{x}^T \mathbf{z} \\ \text{sub. to} \quad & \sum_{i=1}^N x_i - c = 0, \\ & x_i \geq 0, \quad i = 1, \dots, N, \end{aligned} \quad (23)$$

where $\mathbf{z} = D_k^{-1}(\mathbf{x}^{(k)} - \alpha_k D_k \nabla f(\mathbf{x}^{(k)}))$.

Due to the special structure of the constraints, the problem (23) can be reformulated as a one-dimensional root-finding problem.

In fact, if $\bar{\mathbf{x}}$ denotes the solution of (23), then from the KKT first order optimality conditions we know that there exist Lagrange multipliers $\bar{\lambda} \in \mathbb{R}$ and $\bar{\boldsymbol{\mu}} \in \mathbb{R}^N$ such that

$$\begin{aligned} D_k^{-1} \bar{\mathbf{x}} - \mathbf{z} - \bar{\lambda} \mathbf{e} - \bar{\boldsymbol{\mu}} &= \mathbf{0} \\ \bar{\mathbf{x}} &\geq \mathbf{0} \\ \bar{\boldsymbol{\mu}} &\geq \mathbf{0} \\ \bar{\boldsymbol{\mu}}^T \bar{\mathbf{x}} &= 0 \\ \sum_{i=1}^N \bar{x}_i - c &= 0 \end{aligned}$$

Figure 1. Time in seconds to perform 100 projections.

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

$$\bar{x}_i(\bar{\lambda}) = \max \left\{ 0, d_i^{(k)}(z_i + \bar{\lambda}) \right\}, \quad \bar{\mu}_i(\bar{\lambda}) = \max \left\{ 0, -(z_i + \bar{\lambda}) \right\}, \quad i = 1, \dots, N.$$

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

$$\sum_{i=1}^N \bar{x}_i(\bar{\lambda}) - c = 0. \quad (24)$$

This means that the computation of the projection $P_{\bar{\Omega}, D_k^{-1}}(\mathbf{x}^{(k)} - \alpha_k D_k \nabla f(\mathbf{x}^{(k)}))$ essentially reduces to solve a root-finding problem for a piecewise linear monotonically non-decreasing function. Specialized linear time algorithms for this root-finding problem can be derived from the wide literature available for the continuous quadratic knapsack problem [3, 4, 5, 6, 7]. In the SGP implementation proposed in this work we face the problem (24) by the secant-based method suggested in [4] that has shown very good performance also within the gradient projection methods used for the quadratic programs arising in training the learning methodology Support Vector Machines [30, 31]. To emphasize the scaling properties of the projection algorithm [4] we show in Figure 1 the computational time it requires with respect to the size of the problem. The projection problems arise from the application of SGP to the reconstruction of images of different sizes; for each size, the time to perform one hundred projections is reported. The experiments are carried out on an AMD Opteron Dual Core 2.4 GHz processor using Matlab 7.5.0

From Figure 1 we may observe the linear scaling of the secant-based method [4] and the low computational cost of each projection: for example, in case of $N = 1024^2$ a projection is performed in approximately 0.2 seconds.

3.2. Update the scaling matrix

The choice of the scaling matrix D_k in SGP must aim at two main goals: avoiding to introduce significant computational costs and improving the convergence rate. As previously motivated, a diagonal scaling allows one to make the projection in step 2 of SGP a non-excessively expensive task; thus, we will concentrate on such kind of scaling matrices. A classical choice is to use a scaling matrix $D_k = \text{diag}(d_1^{(k)}, d_2^{(k)}, \dots, d_N^{(k)})$ that approximates the inverse of the Hessian matrix $\nabla^2 J(\mathbf{x})$; for example by requiring

$$d_i^{(k)} \approx \left(\frac{\partial^2 J(\mathbf{x}^{(k)})}{(\partial x_i)^2} \right)^{-1}, \quad i = 1, \dots, N.$$

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

$$d_i^{(k)} = \min \left\{ L, \max \left\{ \frac{1}{L}, \left(\frac{\partial^2 J(\mathbf{x}^{(k)})}{(\partial x_i)^2} \right)^{-1} \right\} \right\}, \quad i = 1, \dots, N, \quad (25)$$

where L is an appropriate threshold. Another appealing choice is suggested by the diagonal scaling used to rewrite the EM method as a special scaled gradient method for minimizing $J(\mathbf{x})$ [15]:

$$\bar{\mathbf{x}}^{(k+1)} = X_k A^T Y_k^{-1} \mathbf{b} = \bar{\mathbf{x}}^{(k)} - X_k \nabla J(\bar{\mathbf{x}}^{(k)}),$$

where $X_k = \text{diag}(\bar{\mathbf{x}}^{(k)})$ and $Y_k = \text{diag}(A\bar{\mathbf{x}}^{(k)} + \mathbf{b}g)$ (refer to section 4.3 for more details on the EM method and a comparison with SGP). By following this idea we may introduce the updating rule

$$d_i^{(k)} = \min \left\{ L, \max \left\{ \frac{1}{L}, x_i^{(k)} \right\} \right\}, \quad i = 1, \dots, N. \quad (26)$$

From a computational viewpoint, the updating rule (25) is more expensive than (26), due to the computation of the diagonal entries of the Hessian (see (21)). With regard to the effect of the scaling on the SGP convergence rate, we will show the behaviour of the above updating rules on several test problems in the next section. Now we go into details about another crucial issue for the convergence rate of a gradient method: the choice of the steplength.

3.3. Update the steplength

Steplength selection rules in gradient methods have received an increasing interest in the last years from both the theoretical and the practical point of view. On one hand, following the original ideas of Barzilai and Borwein (BB) [32], several steplength updating strategies have been devised to accelerate the slow convergence exhibited in most cases by standard gradient methods, and a lot of effort has been put into explaining

the effects of these strategies [33, 8, 4, 34, 9, 35, 10]. On the other hand, numerical experiments on randomly generated, library and real-life test problems have confirmed the remarkable convergence rate improvements involved by some BB-like steplength selections [8, 4, 36, 9, 37, 30, 10]. Thus, it seems natural to equip SGP with a steplength selection that takes into account of the recent advances on the BB-like updating rules.

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 $\nabla^2 J(\mathbf{x}^{(k)})$ and derive two updating rules for α_k by forcing quasi-Newton properties on $B(\alpha_k)$:

$$\alpha_k^{\text{BB1}} = \arg \min_{\alpha_k \in \mathbb{R}} \|B(\alpha_k) \mathbf{s}^{(k-1)} - \mathbf{z}^{(k-1)}\| \quad (27)$$

and

$$\alpha_k^{\text{BB2}} = \arg \min_{\alpha_k \in \mathbb{R}} \|\mathbf{s}^{(k-1)} - B(\alpha_k)^{-1} \mathbf{z}^{(k-1)}\|, \quad (28)$$

where $\mathbf{s}^{(k-1)} = (\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)})$ and $\mathbf{z}^{(k-1)} = (\nabla J(\mathbf{x}^{(k)}) - \nabla J(\mathbf{x}^{(k-1)}))$.

In this way, the steplengths

$$\alpha_k^{(1)} = \frac{\mathbf{s}^{(k-1)T} D_k^{-1} D_k^{-1} \mathbf{s}^{(k-1)}}{\mathbf{s}^{(k-1)T} D_k^{-1} \mathbf{z}^{(k-1)}} \quad (29)$$

and

$$\alpha_k^{(2)} = \frac{\mathbf{s}^{(k-1)T} D_k \mathbf{z}^{(k-1)}}{\mathbf{z}^{(k-1)T} D_k D_k \mathbf{z}^{(k-1)}} \quad (30)$$

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 :

$$\alpha_k^{\text{BB1}} = \frac{\mathbf{s}^{(k-1)T} \mathbf{s}^{(k-1)}}{\mathbf{s}^{(k-1)T} \mathbf{z}^{(k-1)}}, \quad \alpha_k^{\text{BB2}} = \frac{\mathbf{s}^{(k-1)T} \mathbf{z}^{(k-1)}}{\mathbf{z}^{(k-1)T} \mathbf{z}^{(k-1)}}. \quad (31)$$

At this point, inspired by the steplength alternations successfully implemented in the framework of non-scaled gradient methods [9, 10], we propose a steplength updating rule for SGP which adaptively alternates the values provided by (29) and (30). The details of the SGP steplength selection are given in Algorithm SS. This rule decides the alternation between two different selection strategies by means of the variable threshold τ_k instead of a constant parameter as done in [9] and [10]. This trick makes the choice of τ_0 less important for the SGP performance and, in our experience, seems able to avoid the drawbacks due to the use of the same steplength rule in too many consecutive iterations. Finally, we remark that a deeper analysis about the SGP steplength selections should be worthwhile; in fact, at least to our knowledge, steplength rules for scaled gradient methods are not well investigated in literature and the generalization of the standard BB-like selections to this context is an interesting open problem. However, such an analysis is beyond the aims of this work and we will limit to show in the next section the effectiveness of the proposed strategy in comparison with some widely used steplength selections.

Algorithm SS (SGP Steplength Selection)

```

IF  $k = 0$ 
    set  $\alpha_0 \in [\alpha_{min}, \alpha_{max}]$ ,  $\tau_1 \in (0, 1)$  and a nonnegative integer  $M_\alpha$ ;
ELSE
    IF  $\mathbf{s}^{(k-1)T} D_k^{-1} \mathbf{z}^{(k-1)} \leq 0$  THEN
         $\alpha_k^{(1)} = \alpha_{max}$ ;
    ELSE
         $\alpha_k^{(1)} = \max \left\{ \alpha_{min}, \min \left\{ \frac{\mathbf{s}^{(k-1)T} D_k^{-1} D_k^{-1} \mathbf{s}^{(k-1)}}{\mathbf{s}^{(k-1)T} D_k^{-1} \mathbf{z}^{(k-1)}}, \alpha_{max} \right\} \right\}$ ;
    ENDIF
    IF  $\mathbf{s}^{(k-1)T} D_k \mathbf{z}^{(k-1)} \leq 0$  THEN
         $\alpha_k^{(2)} = \alpha_{max}$ ;
    ELSE
         $\alpha_k^{(2)} = \max \left\{ \alpha_{min}, \min \left\{ \frac{\mathbf{s}^{(k-1)T} D_k \mathbf{z}^{(k-1)}}{\mathbf{z}^{(k-1)T} D_k D_k \mathbf{z}^{(k-1)}}, \alpha_{max} \right\} \right\}$ ;
    ENDIF
    IF  $\alpha_k^{(2)} / \alpha_k^{(1)} \leq \tau_k$  THEN
         $\alpha_k = \min \left\{ \alpha_j^{(2)}, j = \max \{1, k - M_\alpha\}, \dots, k \right\}$ ;     $\tau_{k+1} = \tau_k * 0.9$ ;
    ELSE
         $\alpha_k = \alpha_k^{(1)}$ ;     $\tau_{k+1} = \tau_k * 1.1$ ;
    ENDIF
ENDIF

```

4. Numerical experiments

The experiments of this section aim to show the practical usefulness of SGP for image restoration problems. Firstly, in order to determine an SGP setting suited for this application, we carefully evaluate the updating rules for the scaling matrix and the steplength parameter proposed in the previous section. Secondly, we compare SGP with the standard EM method [15], the accelerated EM version described in [23] and the WMRNSD algorithm proposed in [24]. All the methods are implemented in Matlab 7.5.0 and the experiments are performed on a computer equipped with a processor AMD Opteron Dual Core 2.4 GHz.

4.1. Test problems and performance measures

The considered methods are tested on several optimization problems of the form (2) and (3), with the objective function $J(\mathbf{x})$ defined as in (19), corresponding to the deblurring problems arising from a set of astronomical images corrupted by Poisson noise. The test problems are generated by convolving the original 256×256 images shown in Figure 2 and denoted by the letters A, B, C, with a *point spread function* (PSF). Then, a constant background term is added and the resulting images are perturbed with Poisson

noise. From the original images, several blurred noisy images have been generated, with

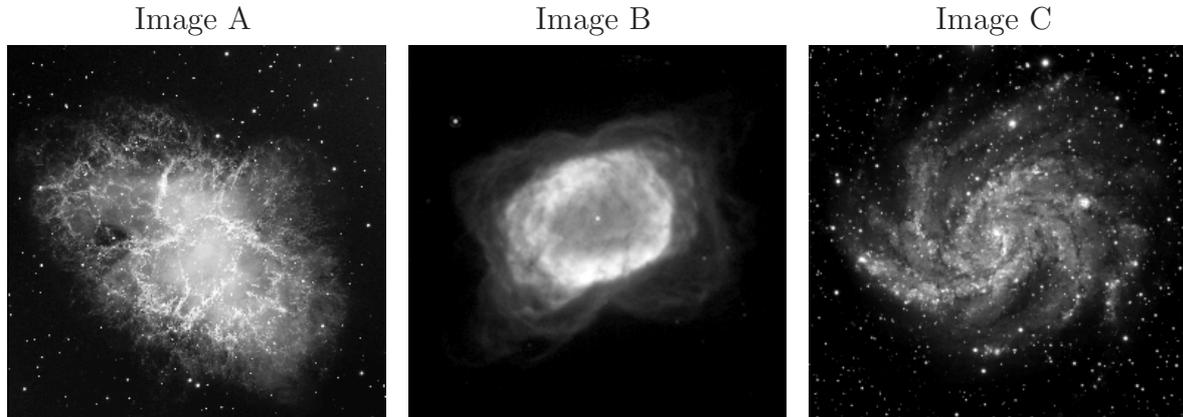


Figure 2. Original images.

different PSFs and noise levels; however, the relative behaviour of the methods on the corresponding deblurring problems are very similar and, consequently, we only discuss the results corresponding to an ideal PSF and three levels of noise. In particular, by following [38], our PSF is defined as

$$2 \left(\frac{J_1(R)}{R} \right)^2,$$

where J_1 is the Bessel function of the first kind and $R = \sqrt{x^2 + y^2}$, $x, y \in \mathbb{R}$. The PSF is computed on a 256×256 grid defined by considering uniformly spaced values in $[-36.4113, 36.4113]$; the resulting matrix is then normalized such that $A^T \mathbf{e} = \mathbf{e}$. Since we are considering the case of Poisson noise, the different noise levels are obtained by changing the total flux (total number of counts) of the original image: the noise level is increasing when the total flux is decreasing. For each of the considered images, we show in Figure 3 the three blurred noisy images used in these experiments: the total flux is 4.43×10^9 for the images in the left panels, 7.02×10^8 for the images in the middle panels and 4.43×10^7 for the images in the right panels. In all the experiments the background level is $bg = 6.76 \times 10^3$.

In order to evaluate the performance of the deblurring methods, we measure for each method the relative reconstruction error, defined as $\|\mathbf{x}^{(k)} - \mathbf{x}\|/\|\mathbf{x}\|$, where \mathbf{x} is the image to be reconstructed and $\mathbf{x}^{(k)}$ is the reconstruction after k iterations; then we report the minimum relative reconstruction error (err_opt), the number of iterations (it_opt) and the computational time in seconds (sec) required to provide the minimum error. The cases where a method reaches the prefixed maximum number of iterations will be marked with an asterisk and the relative error and the time corresponding to that number of iterations will be given.

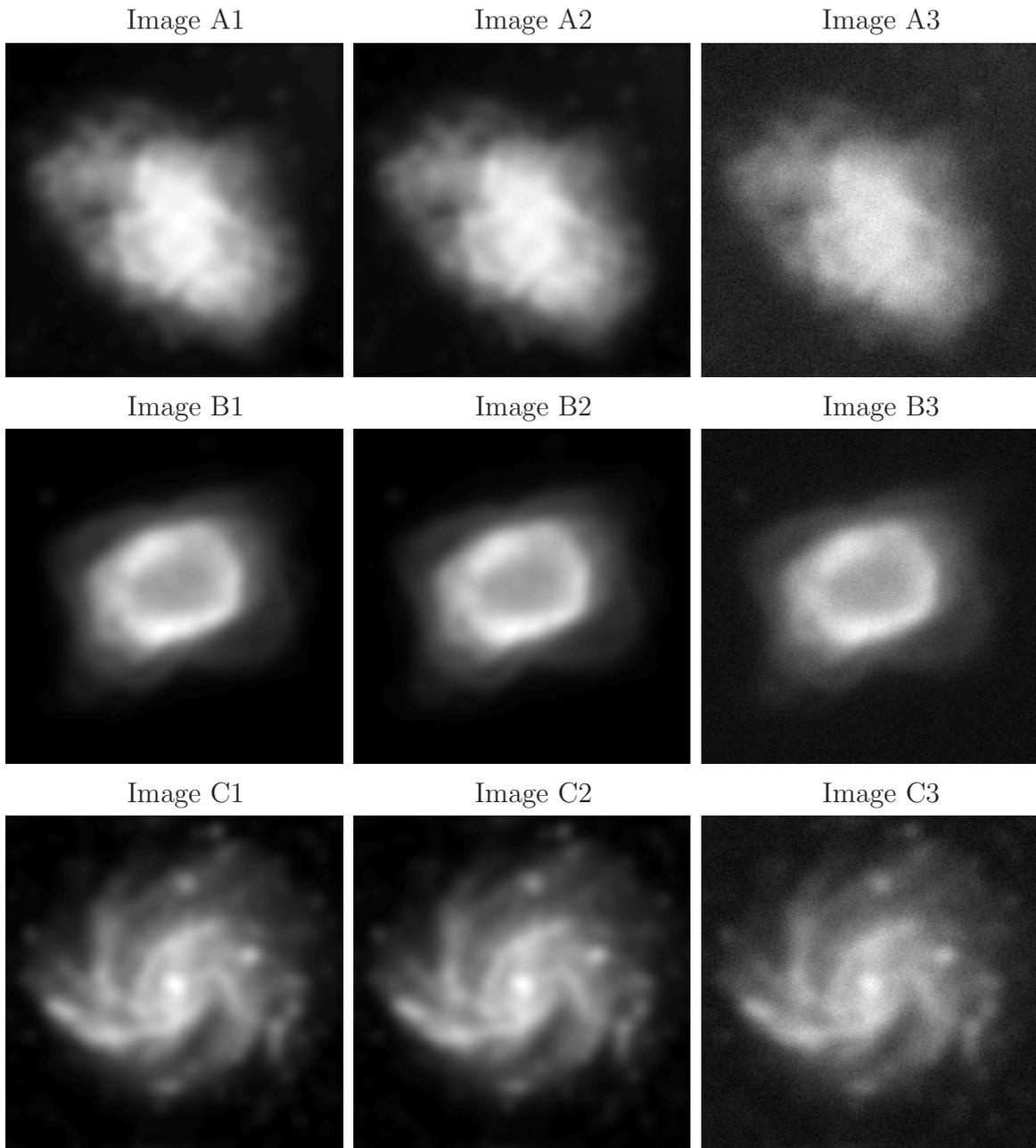


Figure 3. Blurred noisy images. The left, middle and right panels refer to low, medium and high noise levels, respectively.

4.2. Scaling matrix and steplength parameter in SGP

We study the SPG behaviour for different choices of the scaling matrix and of the steplength rule. We test three scaling matrices: $D_k = I$, where I denotes the identity matrix, D_k selected as in (25) and D_k defined as in (26). In the last two cases we use $L = 10^{10}$. Concerning the steplength, we evaluate the following updating rules:

- SGP-BB1: $\alpha_k = \alpha_k^{(1)}$, where $\alpha_k^{(1)}$ is defined as in Algorithm SS;
- SGP-BB2: $\alpha_k = \alpha_k^{(2)}$, where $\alpha_k^{(2)}$ is defined as in Algorithm SS;

- SGP-ABB: α_k defined by alternating $\alpha_k^{(1)}$ and $\alpha_k^{(2)}$ as in the ABB method described in [10] (that is, by setting in Algorithm SS $M_\alpha = 0$ and $\tau_k = 0.15$, for any k);
- SGP-SS: α_k defined by the Algorithm SS, with $\tau_1 = 0.5$ and $M_\alpha = 2$.

In all the above selections we set $\alpha_{min} = 10^{-10}$, $\alpha_{max} = 10^5$ and $\alpha_0 = 1.3$, except for SGP-BB1 where we found convenient to start with a smaller steplength: $\alpha_0 = 10^{-2}$.

The above scaling matrices and steplength rules are examined in both the monotone and the nonmonotone version of SGP; the line-search parameter are: $\theta = 0.4$, $\beta = 10^{-4}$, $M = 1$ in the monotone SGP and $M = 10$ in the nonmonotone SGP. In every experiment the pixels of the starting image $\mathbf{x}^{(0)}$ are set as follows: $x_i^{(0)} = c/N$, $i = 1, \dots, N$, where $c = \sum_{i=1}^N (b_i - bg)$ denotes the right hand side of the equality constraint in the problem (3).

The numerical results obtained by solving the problem (2) are reported in Tables 1 and 3 for the nonmonotone and monotone versions, respectively. The behaviour on the problem (3) is described in Table 2 for the nonmonotone SGP and in Table 4 for the monotone SGP. The main conclusion that can be drawn from these experiments is that the updating rule (26) for the scaling matrix D_k combined with the steplength selection suggested in Algorithm SS gives generally the best performance in terms of computational time and a reconstruction error comparable with those provided by the other choices of D_k and α_k . This special version of SGP is able to achieve a very good convergence rate and benefits from a non-expensive updating of the matrix D_k . Interesting results in terms of number of iterations are observed also when D_k is updated by the rule (25); nevertheless, the additional costs introduced by this rule seem to imply significant reductions of the overall performance. Concerning the steplength, the selections exploited in SGP-ABB and SGP-SS provide the better results, confirming the effectiveness of the strategies based on adaptive alternations of the BB-like rules [9, 10]. In particular, the alternation implemented in Algorithm SS yields remarkable improvements when D_k is updated as in (26). For shortness, in the following we will denote by SGP the scheme that uses Algorithm SS for defining the steplength and exploits the rule (26) for updating the scaling matrix.

For the sake of completeness, we show in Figure 4 the SGP relative reconstruction error as a function of the number of iterations. In each panel of the figure the errors obtained by applying SGP to the problems (2) and (3) are reported (denoted by SGP-(2) and SGP-(3), respectively); for each test image, both the nonmonotone (the left panels) and the monotone (the right panels) version of SGP is considered. No significant differences are observed between the reconstruction errors corresponding to the two problems. In particular, in both cases the error drops to a value close to the minimum in very few iterations and it remains close to this value for a large number of iterations; this suggests that the choice of the optimal number of iterations does not seem to be critical in the case of real images. From a computational point of view, it is worth recalling that, when SGP applies to the problem (3), each iteration is slightly more expensive due to the more complicated projection operation. If we compare the

Table 1. Nonmonotone SGP on the deblurring problem (2).

	$D_k = I$			D_k as in (25)			D_k as in (26)		
	it_opt	err_opt	sec	it_opt	err_opt	sec	it_opt	err_opt	sec
Test problem: Image A1									
SGP-BB1	2443	0.1865	147.43	607	0.1862	61.90	748	0.1850	53.08
SGP-BB2	898	0.1865	46.75	736	0.1862	69.29	678	0.1851	38.24
SGP-ABB	979	0.1866	49.92	443	0.1862	32.54	624	0.1851	36.94
SGP-SS	624	0.1866	33.91	346	0.1862	27.03	380	0.1851	23.40
Test problem: Image B2									
SGP-BB1	1101	0.0586	66.34	307	0.0571	30.76	550	0.0544	48.66
SGP-BB2	1107	0.0586	59.05	380	0.0571	34.58	229	0.0542	12.83
SGP-ABB	1084	0.0586	55.57	367	0.0571	27.01	205	0.0547	12.24
SGP-SS	1086	0.0586	60.70	292	0.0571	23.06	157	0.0542	9.82

monotone and the nonmonotone version of SGP, almost identical behaviours in terms of reconstruction error are observed. Thus, taking into account that the two SGP versions exhibit similar convergence rate and that the nonmonotone line-search strategy requires less function evaluations with respect to the monotone version, the nonmonotone SGP seems preferable (see also the numerical results in Tables 1-4).

Table 2. Nonmonotone SGP on the deblurring problem (3).

	$D_k = I$			D_k as in (25)			D_k as in (26)		
	it_opt	err_opt	sec	it_opt	err_opt	sec	it_opt	err_opt	sec
Test problem: Image A1									
SGP-BB1	4000*	0.1871	266.97	794	0.1861	91.93	1178	0.1850	102.92
SGP-BB2	4000*	0.1873	241.27	673	0.1861	76.45	666	0.1850	44.12
SGP-ABB	4000*	0.1873	266.38	338	0.1861	28.16	399	0.1851	29.27
SGP-SS	4000*	0.1868	262.98	385	0.1862	33.38	336	0.1851	22.37
Test problem: Image B2									
SGP-BB1	4000*	0.0613	631.13	352	0.0577	42.22	363	0.0635	35.32
SGP-BB2	4000*	0.0588	348.76	372	0.0578	42.38	214	0.0542	14.17
SGP-ABB	4000*	0.0608	590.31	360	0.0577	34.49	180	0.0547	13.65
SGP-SS	4000*	0.0587	281.35	353	0.0577	34.95	163	0.0544	11.37

Table 3. Monotone SGP on the deblurring problem (2).

	$D_k = I$			D_k as in (25)			D_k as in (26)		
	it_opt	err_opt	sec	it_opt	err_opt	sec	it_opt	err_opt	sec
Test problem: Image A1									
SGP-BB1	1592	0.1865	103.36	527	0.1862	44.01	748	0.1850	49.94
SGP-BB2	985	0.1866	55.99	659	0.1863	50.48	605	0.1850	35.35
SGP-ABB	1125	0.1866	60.17	659	0.1862	49.84	538	0.1851	33.24
SGP-SS	568	0.1866	37.40	418	0.1862	33.61	388	0.1851	28.85
Test problem: Image B2									
SGP-BB1	1475	0.0586	90.54	528	0.0571	43.20	550	0.0542	49.72
SGP-BB2	1087	0.0586	60.55	400	0.0582	32.49	190	0.0543	11.71
SGP-ABB	1072	0.0586	57.11	327	0.0571	26.47	191	0.0548	12.54
SGP-SS	1110	0.0586	72.74	386	0.0571	31.56	137	0.0543	9.27

Table 4. Monotone SGP on the deblurring problem (3).

	$D_k = I$			D_k as in (25)			D_k as in (26)		
	it_opt	err_opt	sec	it_opt	err_opt	sec	it_opt	err_opt	sec
Test problem: Image A1									
SGP-BB1	4000*	0.1882	569.74	450	0.1861	46.99	778	0.1851	58.47
SGP-BB2	4000*	0.1886	647.82	573	0.1861	48.53	573	0.1850	38.41
SGP-ABB	4000*	0.1884	637.89	571	0.1861	46.98	511	0.1851	35.18
SGP-SS	4000*	0.1874	557.53	324	0.1861	27.75	420	0.1851	38.26
Test problem: Image B2									
SGP-BB1	4000*	0.0638	627.51	400	0.0577	42.55	633	0.0599	83.25
SGP-BB2	4000*	0.0659	651.00	355	0.0578	32.69	228	0.0541	17.71
SGP-ABB	4000*	0.0631	648.92	313	0.0577	28.57	243	0.0546	17.47
SGP-SS	4000*	0.0598	648.81	320	0.0577	30.23	145	0.0543	13.91

4.3. Comparisons with other methods

To better evaluate the SGP behaviour in image deblurring, we report some comparisons with other widely used iterative regularization methods.

We first consider the EM method [15], also known as Richardson-Lucy method [14, 13]. Starting from a positive initial image $\mathbf{x}^{(0)}$, the EM algorithm looks for a minimum of

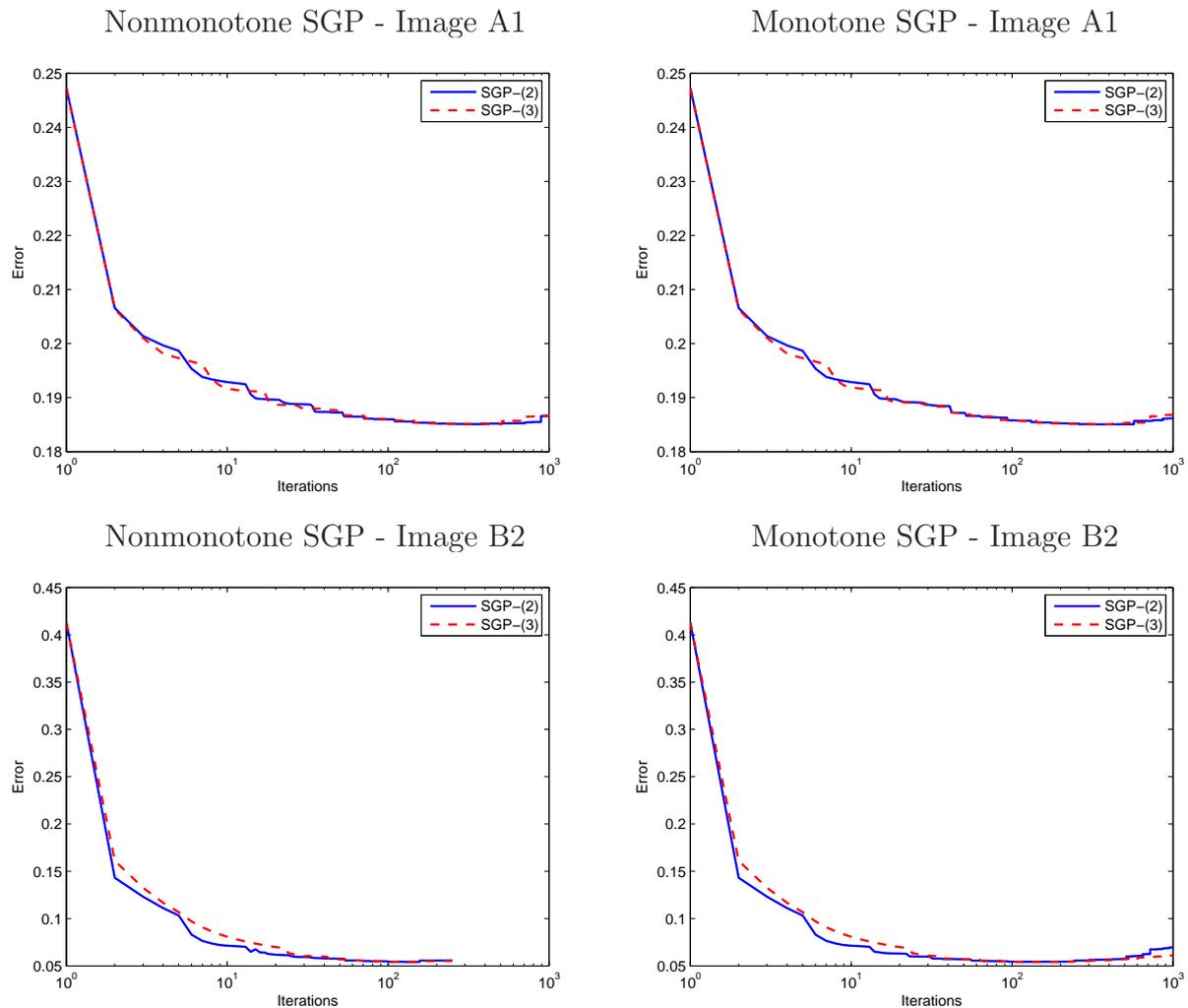


Figure 4. SGP relative reconstruction error.

the functional (19) by exploiting the iteration

$$\mathbf{x}^{(k+1)} = X_k A^T Y_k^{-1} \mathbf{b}. \quad (32)$$

where $X_k = \text{diag}(\mathbf{x}^{(k)})$ and $Y_k = \text{diag}(A\mathbf{x}^{(k)} + b\mathbf{g})$. The nonnegativity of A , $b\mathbf{g}$ and \mathbf{b} guarantees that all the iterates remain nonnegative and, under the assumption $b\mathbf{g} = 0$, useful to ensure the flux conservation property, the convergence to a solution of (2) has been proved by several authors [39, 40, 41, 42, 43] (see also [44] for a convergence analysis derived by a proximal point interpretation of the EM algorithm). In the general case $b\mathbf{g} \neq 0$, at least to our knowledge, the EM convergence is not proved. The EM method is attractive because of its low computational cost, consisting in $\mathcal{O}(N \log N)$ operations per iteration (needed to perform the two matrix-vector products involving the matrix A). This nice feature made the method one of the most popular approaches for astronomical and medical image restoration problems. However, the main drawback of the method is the slow convergence, that, in many cases, leads to the desired approximation of the solution in a too large time. With regard to the comparison between SGP and EM, it

is interesting to observe that the iteration (32) can be written as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k D_k \nabla J(\mathbf{x}^{(k)}), \quad (33)$$

with

$$D_k = X_k, \quad \alpha_k = 1, \quad \forall k.$$

This means that EM can be interpreted as a scaled steepest descent method with a special scaling matrix and a constant steplength [21, 45]. On the other hand, we have seen that SGP uses a similar scaling matrix, exploits variable steplengths to improve the convergence rate and can handle in a natural way a flux conservation constraint. Thus, we may consider SGP a generalization of the EM method.

An interesting accelerated EM version is provided in [23]: it exploits a vector extrapolation to determine the point on which the EM iteration is applied. The extrapolation, consisting in a shift along the direction given by the difference between the current iteration and the previous iteration, introduces a little computational overhead and, consequently, the cost per iteration of the method is only slightly larger than in EM. In the following of the paper we will denote by EM_MATLAB the implementation of this algorithm available in the *deconvlucy* function of the *Image Processing* MATLAB toolbox.

Another algorithm able to show superior convergence rate when compared to EM is the WMRNSD method recently proposed in [24]. This algorithm can be viewed as a steepest descent method applied to the minimization problem arising from a constrained weighted least squares approach to the image restoration problem; the nonnegativity constraint is satisfied by modifying the standard line-search that minimizes the residual norm. The main tasks per iteration consist in two matrix-vector products, as for the other methods. In our computational study we consider the WMRNSD version tested in [24], whose MATLAB implementation is available at the web page <http://web.math.umt.edu/bardsley/codes.html>.

In Table 5 we show the numerical results obtained by solving some deblurring problems with the above iterative image reconstruction algorithms: the nonmonotone SGP applied to the problem (2) (SGP-(2)) and (3) (SGP-(3)), the standard EM, the EM_MATLAB and the WMRNSD. The methods start from the same image ($x_i^{(0)} = c/N$, $i = 1, \dots, N$) and all the parameters of the nonmonotone SGP are set as described in the previous section. Also in this table, the numbers of iterations (it_opt) refer to the iteration where the minimum relative reconstruction error is obtained; an asterisk is used to mark the cases where this minimum is not reached within the prefixed maximum number of iterations. In all the experiments, SGP largely outperforms EM in the number of iterations and in the computational time, even if the time per iterations exhibited by SGP-(2) and SGP-(3) is approximately 40% and 70% greater than in EM, respectively. Concerning the optimal reconstruction error, no significant differences are observed between SGP and EM. To better compare the behaviour of the two methods, we plot in Figure 5 the reconstruction error as a function of the number of iterations for the test problems corresponding to the blurred images C1 and C3. The reconstructed

Table 5. Comparison among SGP, EM, EM_MATLAB and WMRNSD

	it_opt	err_opt	sec	it_opt	err_opt	sec	it_opt	err_opt	sec
	Image A1			Image A2			Image A3		
SGP-(2)	380	0.1851	23.40	103	0.1866	6.37	21	0.1944	1.32
SGP-(3)	336	0.1851	22.37	108	0.1865	8.44	20	0.1947	1.55
EM	10000*	0.1852	433.25	4047	0.1865	177.14	414	0.1942	17.90
EM_MATLAB	388	0.1853	18.32	141	0.1868	7.19	46	0.1947	2.40
WMRNSD	10000*	0.1853	573.31	2904	0.1866	163.61	50	0.1942	2.72
	Image B1			Image B2			Image B3		
SGP-(2)	251	0.0513	15.67	157	0.0542	9.82	26	0.0689	1.61
SGP-(3)	274	0.0514	19.12	163	0.0544	11.37	35	0.0688	2.57
EM	10000*	0.0512	413.20	4185	0.0541	183.49	500	0.0687	21.69
EM_MATLAB	259	0.0516	12.30	139	0.0559	7.83	44	0.0700	2.64
WMRNSD	10000*	0.0521	571.69	2928	0.0548	166.43	80	0.0690	4.26
	Image C1			Image C2			Image C3		
SGP-(2)	736	0.2922	45.89	374	0.2953	23.36	41	0.3127	2.51
SGP-(3)	1125	0.2923	80.13	272	0.2949	18.99	45	0.3125	3.26
EM	10000*	0.2929	431.77	10000*	0.2945	430.14	1459	0.3110	61.60
EM_MATLAB	811	0.2952	38.30	280	0.2977	14.60	68	0.3138	3.75
WMRNSD	10000*	0.2939	582.13	10000*	0.2948	580.76	168	0.3115	9.09

images corresponding to the minimum errors are shown in Figure 6 for the blurred images A2, B2 and C2. From the above results we may conclude that SGP seems able to provide the same reconstruction accuracy given by EM but with a remarkable computational gain.

We complete the experiments by comparing SGP also with the accelerated algorithm EM_MATLAB and with the WMRNSD method. The numerical results in Table 5 show that, even if the cost per iteration in SGP is higher than in EM_MATLAB, the two approaches are very well comparable and sometimes SGP-(2) can be a valid alternative. We point out that, as far as we know, no convergence proof of the Biggs-Andrews algorithm implemented by EM_MATLAB is available.

Finally, we can observe that SGP largely outperforms the WMRNSD method in terms

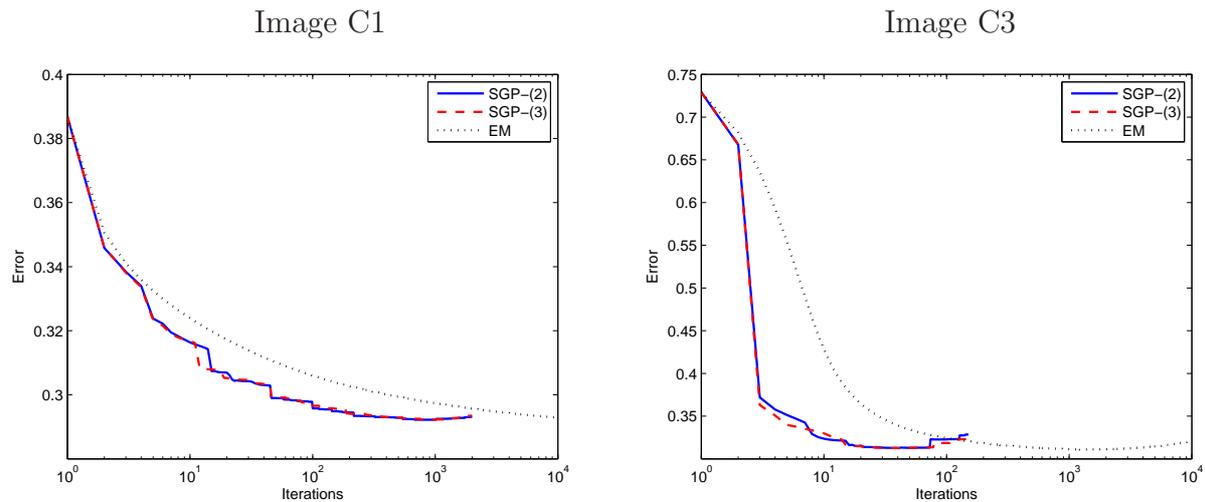


Figure 5. SGP and EM relative reconstruction error.

of both iterations number and computational time.

5. Conclusions

We have proposed a scaled gradient projection method, called SGP, for solving optimization problems with simple constraints. The main features of SGP are its global convergence properties and the use of efficient updating rules for the steplength parameter and for the scaling matrix able to improve the convergence rate. A wide computational study on the minimization problems arising from the maximum likelihood approaches to image deblurring shows that SGP can provide the same reconstruction accuracy of the standard EM method with much less iterations and, consequently, with a remarkable computational gain. Furthermore, it exhibits the ability to effectively handle a linear equality constraint in addition to the nonnegativity or box constraints. The SGP effectiveness is also confirmed by the comparisons with other popular iterative regularization methods.

Future works will regard the evaluation of the proposed algorithm on different optimization problems arising in image deblurring and the comparison with quasi-Newton and interior point methods. If SGP can compete with these methods, then it could provide a very useful and simple approach to iterative image reconstruction.

Acknowledgments

The authors are thankful to the anonymous referees for their useful comments and suggestions. This research is supported by the PRIN2006 project of the Italian Ministry of University and Research *Inverse Problems in Medicine and Astronomy*, grant 2006018748.

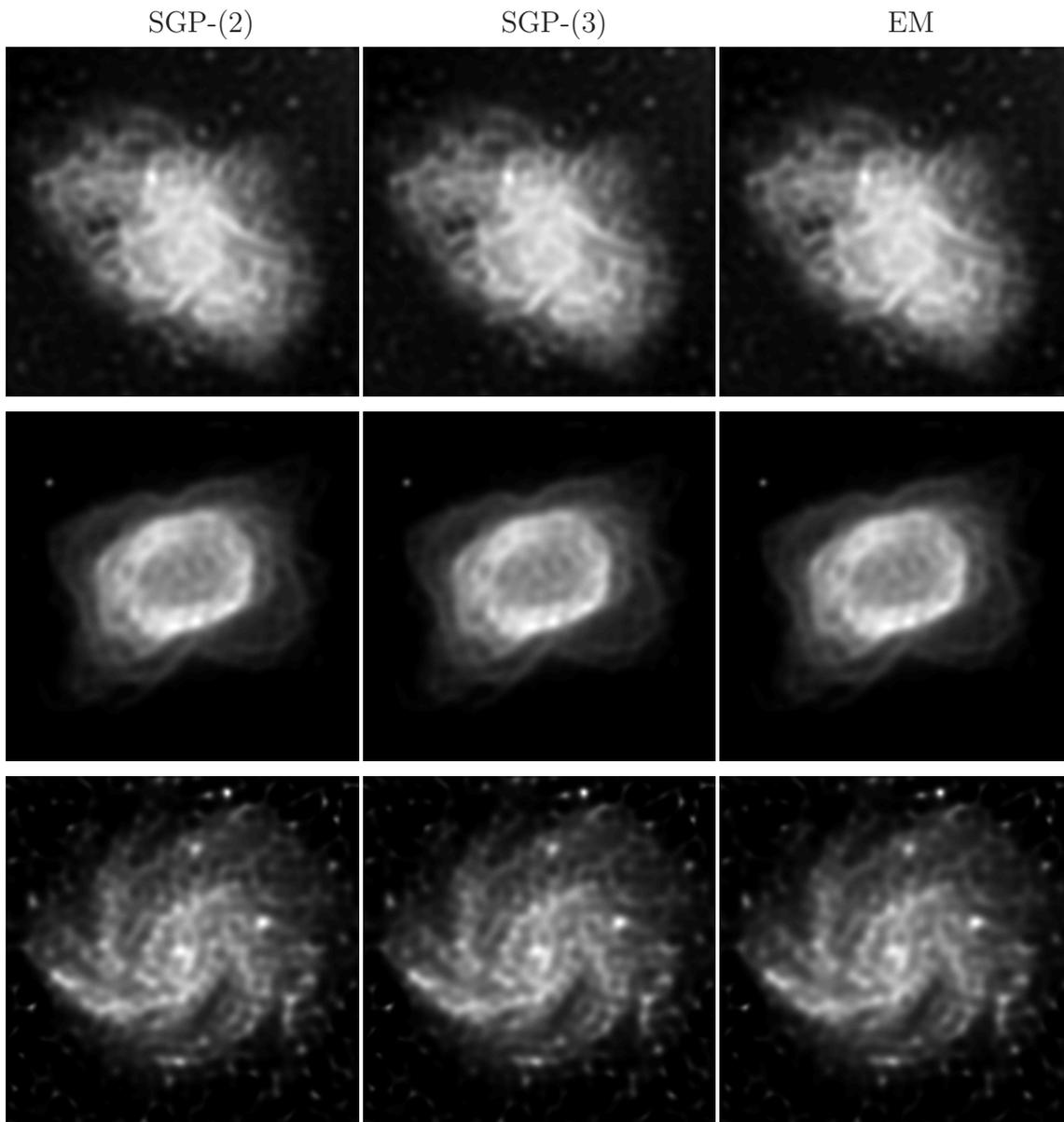


Figure 6. Reconstruction obtained with the SGP-(2), the SGP-(3) and the EM methods for the blurred images A2, B2, C2 (from the top to the bottom of the figure).

References

- [1] Bertero M and Boccacci P 1998 *Introduction to Inverse Problems in Imaging* (Bristol: Institute of Physics Publishing)
- [2] Vogel C R 2002 *Computational Methods for Inverse Problems* (Frontiers Appl. Math., Philadelphia: SIAM)
- [3] Brucker P 1984 An $O(n)$ algorithm for quadratic knapsack problems *Oper. Res. Lett.* **3(3)** 163–166
- [4] Dai Y H and Fletcher R 2006 New algorithms for singly linearly constrained quadratic programming problems subject to lower and upper bounds *Math. Programming* **106(3)** 403–421
- [5] Kiwiel K C 2008 Breakpoint searching algorithms for the continuous quadratic knapsack problem *Math. Programming* **112** 473–491
- [6] Maculan N, Santiago C P, Macambira E M, and Jardim M H C 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(3)** 553–574
- [7] Pardalos P M and Kovoor N 1990 An algorithm for a singly constrained class of quadratic programs subject to upper and lower bounds *Math. Programming* **46** 321–328
- [8] 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** 604–627
- [9] Frassoldati G, Zanghirati G and Zanni L 2008 New adaptive stepsize selections in gradient methods *J. Industrial and Management Optim.* **4(2)** 299–312
- [10] Zhou B, Gao L and Dai Y H 2006 Gradient methods with adaptive step-sizes *Comput. Optim. Appl.* **35(1)** 69–86
- [11] Daube-Witherspoon M E and Muehllehner G 1986 An iterative image space reconstruction algorithm suitable for volume ect *IEEE Trans. Med. Imaging* **5** 61–66
- [12] Lanteri H, Roche M and Aime C 2002 Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms *Inverse Problems* **18** 1397–1419
- [13] Lucy L B 1974 An iterative technique for the rectification of observed distributions *Astronom. J.* **79** 745–754
- [14] Richardson W H 1972 Bayesian-based iterative method of image restoration *J. Opt. Soc. Amer. A* **62** 55–59
- [15] Shepp L A and Vardi Y 1982 Maximum likelihood reconstruction for emission tomography *IEEE Trans. Med. Imaging* **1** 113–122
- [16] Birgin E G, Martinez J M, and Raydan M 2000 Nonmonotone spectral projected gradient methods on convex sets *SIAM J. Optim.* **10** 1196–1211
- [17] Grippo L, Lampariello F, and Lucidi S 1986 A nonmonotone line-search technique for Newton’s method *SIAM J. Numer. Anal.* **23** 707–716
- [18] Kelley C T 1999 *Iterative Methods for Optimization* (Frontiers Appl. Math., Philadelphia: SIAM)
- [19] Bardsley J and Vogel C 2003 Nonnegatively constrained convex programming methods for image reconstruction *SIAM J. Sci. Comput.* **25** 1326–1343
- [20] Landi G and Loli Piccolomini E 2008 A projected Newton-CG method for nonnegative astronomical image deblurring *Numer. Alg.* **48(4)** 279–300
- [21] Bertsekas D P 1999 *Nonlinear Programming* (Athena Scientific, 2nd edition)
- [22] Csiszär I 1991 Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems *Ann. Stat.* **19** 2032–2066
- [23] Biggs D S C and Andrews M 1997 Acceleration of iterative image restoration algorithms *Appl. Opt.* **36** 1766–1775
- [24] Bardsley J and Nagy J 2006 Covariance-preconditioned iterative methods for nonnegatively constrained astronomical imaging *SIAM J. Matr. Anal. Appl.* **27** 1184–1198
- [25] Johansson B, Elfving T, Kozlov V, Censor Y, Forssén P E and Granlund G 2006 The application of an oblique-projected Landweber method to a model of supervised learning *Math. Comp. Modelling* **43** 892–909
- [26] Birgin E G, Martinez J M and Raydan M 2003 Inexact spectral projected gradient methods on convex sets *IMA J. Numer. Anal.* **23** 539–559
- [27] Hanke M, Nagy J and Vogel C 2000 Quasi-Newton approach to nonnegative image restoration *Lin. Alg. and Appl.* **316** 223–236
- [28] Davis P J 1979 *Circulant Matrices* (John Wiley & Sons)
- [29] Strang G 1999 The discrete cosine transform *SIAM Review* **41(1)** 135–147
- [30] Zanni L 2006 An improved gradient projection-based decomposition technique for support vector machines *Comput. Management Sci.* **3** 131–145
- [31] Zanni L, Serafini T and Zanghirati G 2006 Parallel software for training large scale support vector machines on multiprocessor systems *J. Mach. Learn. Res.* **7** 1467–1492
- [32] Barzilai J and Borwein J M 1988 Two point step size gradient methods *IMA J. Numer. Anal.* **8** 141–148

- [33] Dai Y H and Fletcher R 2005 On the asymptotic behaviour of some new gradient methods *Math. Programming* **103(3)** 541–559
- [34] Fletcher R 2001 *On the Barzilai-Borwein method* Technical Report NA/207, Department of Mathematics, University of Dundee, Dundee, UK
- [35] Friedlander A, Martínez J M, Molina B and Raydan M 1999 Gradient method with retards and generalizations *SIAM J. Numer. Anal.* **36** 275–289
- [36] 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. Selected Topics in Signal Process.* **1** 586–597
- [37] Serafini T, Zanghirati G and Zanni L 2005 Gradient projection methods for quadratic programs and applications in training support vector machines *Optim. Meth. Soft.* **20(2–3)** 343–378
- [38] Anconelli B, Bertero M, Boccacci P, Carbillet C and Lanteri H 2005 Restoration of interferometric images - iii. Efficient Richardson-Lucy methods for LINC-NIRVANA data reduction *Astron. Astrophys.* **430** 731–738
- [39] Iusem A N 1991 Convergence analysis for a multiplicatively relaxed EM algorithm *Math. Meth. Appl. Sci.* **14** 573–593
- [40] Iusem A N 1992 A short convergence proof of the EM algorithm for a specific Poisson model *REBRAPE* **6** 57–67
- [41] Lange K and Carson R 1984 EM reconstruction algorithms for emission and transmission tomography *J. Comp. Assisted Tomography* **8** 306–316
- [42] Mülthei H N and Schorr B 1989 On properties of the iterative maximum likelihood reconstruction method *Math. Meth. Appl. Sci.* **11** 331–342
- [43] Vardi Y, Shepp L A and Kaufman L 1985 A statistical model for positron emission tomography *J. Amer. Statist. Soc.* **80(389)** 8–37
- [44] Tseng P. 2004 An analysis of the EM algorithm and entropy-like proximal point methods *Math. Oper. Res.* **29** 27–44
- [45] Kaufman K 1987 Implementing and accelerating the EM algorithm for positron emission tomography *IEEE Trans. Med. Imaging* **6** 37–51