Accelerated gradient methods for the x-ray imaging of solar flares

S Bonettini$^1$ and M Prato$^2$

$^1$ Dipartimento di Matematica e Informatica, Università di Ferrara, Via Saragat 1, I-44122 Ferrara, Italy
$^2$ Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio Emilia, Via Campi 213/b, I-41125 Modena, Italy

E-mail: silvia.bonettini@unife.it and marco.prato@unimore.it

Received 22 November 2013, revised 27 January 2014
Accepted for publication 18 February 2014
Published 17 April 2014

Abstract
In this paper we present new optimization strategies for the reconstruction of x-ray images of solar flares by means of the data collected by the Reuven Ramaty high energy solar spectroscopic imager. The imaging concept of the satellite is based on rotating modulation collimator instruments, which allow the use of both Fourier imaging approaches and reconstruction techniques based on the straightforward inversion of the modulated count profiles. Although in the last decade, greater attention has been devoted to the former strategies due to their very limited computational cost, here we consider the latter model and investigate the effectiveness of different accelerated gradient methods for the solution of the corresponding constrained minimization problem. Moreover, regularization is introduced through either an early stopping of the iterative procedure, or a Tikhonov term added to the discrepancy function by means of a discrepancy principle accounting for the Poisson nature of the noise affecting the data.

Keywords: gradient projection methods, regularization, discrepancy principle, astronomical imaging, solar flares

(Some figures may appear in colour only in the online journal)

1. Introduction
For any imaging problem, the wavelength of the radiation emitted by the object under investigation is the crucial parameter which addresses the hardware design of the instrument devoted to collect that radiation. While a wide range of wavelengths can be handled by means of optical systems, for ultraviolet, x-ray and γ-ray emission, different strategies are mandatory,
involving collimators, masks and/or grids \cite{1, 2}. An example of a similar imaging system is the one provided by the Reuven Ramaty high energy solar spectroscopic imager (RHESSI), launched from Cape Canaveral on 5 February 2002 with the aim to recover temporally, spatially and spectrally resolved x-ray images of solar eruptions \cite{3, 4}. The RHESSI spacecraft is the last descendant of a generation of collimator-based satellites (e.g. HINOTORI, YOHKOH/HXT, HEIDI) that originated in 1980 from the hard x-ray imaging spectrometer, carried in the Solar Maximum Mission. It observes x-ray emission from the entire solar disc through a set of nine rotating modulation collimators (RMCs) \cite{5}, each one being a pair of co-aligned identical grids which allow only a fraction of the incoming radiation to be collected by the corresponding detector. Since each collimator is made up of grids with different pitches, the signals provided by the nine detectors are characterized by different resolutions and signal-to-noise (SNR) ratios.

Thanks to the RMCs hardware, two strategies are available to reconstruct the desired image:

(i) the straightforward inversion of the modulated count profiles provided by RHESSI, or
(ii) the construction (through a fitting procedure of the count profiles) of a set of amplitude and phase pairs of the detected radiation, also known in optical interferometry by the term visibilities, corresponding to spatial frequencies depending on the angular resolutions of each RMC pair (see the appendices of \cite{5, 6}), followed by the application of a Fourier-based inversion technique.

The very first algorithms designed for the RHESSI imaging followed the first strategy, and included CLEAN, Pixon and maximum entropy (see \cite{5} and references therein). However, in the last ten years, the visibility approach became more attractive, since the use of the fast Fourier transform (FFT) led to a computational cost of the reconstruction algorithms of the order of a few seconds. Examples of techniques developed according to this strategy are given by the visibility-based maximum entropy method \cite{7}, the uv-smooth algorithm \cite{8} and the space-D approach \cite{9, 10}. The price to pay when using the visibilities is that the input of the reconstruction algorithms is not the raw data provided by the spacecraft, but the result of a fitting preprocessing step that introduces systematic errors which increase as the count noise increases.

In this paper, we follow the idea proposed in \cite{11}, in which the RHESSI imaging problem from count profiles has been approached by means of the expectation maximization (EM) algorithm. In particular, numerical simulations showed that EM is able to equal the accuracy in the reconstructions of Pixon with, in general, one fourth of the computational time. The aim of this paper is to further improve the computational cost without losing accuracy through several accelerated versions of EM. In fact, since the iterations of EM converge to a minimizer of the Kullback–Leibler (KL) divergence, also known as the Csiszár I-divergence \cite{12}, one can reformulate the RHESSI imaging problem as the non-negative minimization of the KL divergence and address its solution by means of a very powerful constrained optimization scheme. Here we adopt some recently proposed schemes as the scaled gradient projection (SGP) algorithm \cite{13}, the gradient projection method with extrapolation (GPE) \cite{14} and the affine scaling interior point cyclic Barzilai–Borwein (AS_CBB) approach \cite{15}.

A further advantage of these methods is that they can be easily extended to allow for the addition of regularization terms to the KL divergence. Due to the smooth and extended geometries which are typically observed in RHESSI images, here we do not consider edge-preserving or sparsity-inducing priors, but we analyze the effectiveness of a Tikhonov regularization term \cite{16}. As for the choice of the stopping rule for the iterations (in the case of KL minimization) and the regularization parameter (in the case of KL + Tikhonov
minimization), we adopt the discrepancy principle for Poisson data proposed by Bertero et al [17]. In particular, we propose to include in the discrepancy principle a simple procedure which gives a better estimate of the expected value of the counts for high noise levels. Our numerical experiments show that the accelerated methods are able to provide a substantial gain in the number of iterations needed to provide the regularized solution, and with lower reconstruction errors. Moreover, the presence of the Tikhonov regularization seems not to provide any substantial improvement in the reconstructions, thus the time-consuming step of calculating the optimal regularization parameter seems avoidable.

The plan of the paper is as follows: in section 2 the RHESSI imaging problem from the modulated count profiles is formulated. In section 3 the main features of the accelerated gradient methods are briefly described, while in section 4 we analyze possible ways to introduce regularization and recover a meaningful image. Section 5 is devoted to our numerical simulations and the discussion of the results we obtained. Finally, some comments and conclusions are provided in section 6.

2. The RHESSI imaging problem

Let $D$ be the domain of the unknown image, and let $f(x, y)$ be the intensity of the radiation as a function of the source position $(x_s, y_s) \in D$. For the $j$th subcollimator ($j = 1, \ldots, 9$), let us denote by $P_j(x_s, y_s; t)$ the fraction of the incoming radiation from position $(x_s, y_s) \in D$ which will reach the detector at time $t$ (called transmission probability). Two examples of transmission probability maps for subcollimators 2 and 4 are shown in figure 1. It follows that the detected flux at time $t$ is proportional to the whole emission modulated by the corresponding transmission probability:

$$ C_j(t) = K_j \int_D f(x_s, y_s) P_j(x_s, y_s; t) \, dx_s \, dy_s, \quad j = 1, \ldots, 9, \quad (1) $$

where $K_j$ is some known constant parameter related to the system features (detector area, efficiency...). For simplicity we assume that the field of view coincides with the domain $D$, which is equivalent to assuming that there is no radiation from points outside the map.

Equation (1) is a Fredholm integral equation of the first kind whose kernel is the transmission probability $P_j(x_s, y_s; t)$, which, unfortunately, is not space invariant.
In a more realistic frame, each RHESSI collimator actually provides a set of discrete data \( c^{(j)}_k = (c^{(j)}_1, \ldots, c^{(j)}_{N_j})^T \) which can be modeled as the integral of (1) in a certain number of time intervals:

\[
c^{(j)}_k = \int_{t_{k-1}}^{t_k} C^{(j)}(t) \, dt, \quad k = 1, \ldots, N_j, \quad j = 1, \ldots, 9. \tag{2}
\]

The RHESSI imaging problem consists of finding an approximation of the flux distribution \( f(x, y) \) given the data \( c^{(j)}_k \), the transmission probabilities \( P^{(j)}(x_j, y_j; t) \) and the parameters \( K^{(j)} \). We remark that such an inverse problem is ill-posed in the sense of Hadamard [18], due to the boundedness of the kernels \( P^{(j)} \). In fact, if we denote with \( |\mathcal{D}| \) the area of the domain \( \mathcal{D} \) and with \( \mathcal{P} \) the maximum of \( P^{(j)}(x_j, y_j; t) \) over \( (x_j, y_j) \in \mathcal{D}, t \in [t_{k-1}, t_k] \) and \( j = 1, \ldots, 9 \), from the inequality:

\[
\int_{t_{k-1}}^{t_k} \int_{\mathcal{D}} P^{(j)}(x_j, y_j; t)^2 \, dx_j \, dy_j \, dt \leq \mathcal{P}^2 |\mathcal{D}| (t_k - t_{k-1})
\]

it follows that each kernel \( P^{(j)}(x_j, y_j; t) \) belongs to \( L^2(\mathcal{D}, [t_{k-1}, t_k]) \), which means that the operator mapping \( f(x, y) \) into \( C^{(j)}(t) \) is Hilbert–Schmidt [19]. Since the Hilbert–Schmidt condition is sufficient for an operator to be compact, we have the ill-posedness of the inverse problem [20].

In order to define a discrete counterpart of (2), we assume \( t_k - t_{k-1} = \Delta t \) and we define gridpoints \( (x_h, y_\ell), h, \ell = 1, \ldots, n \) on \( \mathcal{D} \), with uniform spacing \( \Delta x, \Delta y \) along the two dimensions. Then, by applying the rectangular rule on (2) we obtain the following approximation:

\[
c^{(j)}_k \approx K^{(j)} \Delta t \Delta x \Delta y \sum_{h,\ell=1}^n f(x_h, y_\ell) P^{(j)}(x_h, y_\ell; t_k).
\]

If we define the vector \( f = (f_1, \ldots, f_n)^T \) by a lexicographic reordering of \( f_{h\ell} \equiv f(x_h, y_\ell) \Delta x \Delta y \), we can write the discrete formulation of the RHESSI imaging problem as follows:

\[
c = Pf
\]

where

\[
c = \begin{pmatrix} c^{(1)} \\ \vdots \\ c^{(9)} \end{pmatrix}; \quad P = \begin{pmatrix} P^{(1)} \\ \vdots \\ P^{(9)} \end{pmatrix}
\]

and \( P^{(j)} \) is a \( N_j \times n^2 \) matrix whose components are given by \( K^{(j)} P^{(j)}(x_h, y_\ell; t_k) \Delta t \) in the suitable order. Problem (3) is again ill-posed since in general its solution does not exist or is not unique, according to the thickness of the discretization of the observation time in time bins (i.e. the number \( N_j \), which can be chosen by the user). This pathology is typically eluded by computing the generalized solution of (3), which for discrete problems guarantees the well-posedness of the procedure. However, it is well-known that the discretization of a continuous problem which lacks continuity of the solution in the data leads to a highly ill-conditioned problem [21], which needs to be addressed by means of some regularization approach.

Indeed, we must also take into account that the data suffers from the loss of information due to measurement errors which occur in the acquisition process, yielding a perturbation of the data which is usually defined as noise. The noise is a statistical feature in the acquisition process, and, since our data corresponds to photon counts, it is reasonable to assume that it obeys Poisson statistics. More precisely, the actual detected data \( c^{(j)}_k \) should be considered as...
a realization of a Poisson random variable, whose expected value and variance are given by
the noise-free data:
\[ c^{(j)}_k \propto \text{Poisson} \left( \int_{t_{k-1}}^{t_k} C^{(j)}(t) \, dt \right), \quad k = 1, \ldots, N_j, \quad j = 1, \ldots, 9. \]

Therefore maximizing the likelihood (or, equivalently, minimizing the negative logarithm of
the likelihood) of measuring the count profiles \( c \) given a flux distribution \( f \) yields to the
minimization of the KL divergence of the vector \( P f \) from the vector \( c \), defined by:
\[ D_{KL}(c, P f) = \sum_{j=1}^{9} \sum_{k=1}^{N_j} c^{(j)}_k \ln \left( \frac{c^{(j)}_k}{(P f)^{k}_j} \right) + (P f)^{k}_j - c^{(j)}_k. \]

Since the entries of the flux distribution \( f \) must be non-negative, the RHESSI imaging problem
can be formulated as the following constrained minimization problem:
\[ \min_{f \geq 0} J(f) := D_{KL}(c, P f). \] (4)

In a recent paper [11], the authors propose to solve (4) with the well-known EM algorithm,
defined by:
\[ f^{(k+1)} = f^{(k)} \mathbf{P} f^{(k)} \mathbf{P} \left( \frac{c}{f^{(k)}} \right), \quad k = 1, 2, \ldots, \]
where \( \mathbf{1} \) denotes a constant array of ones and \( \cdot, / \) are the componentwise product and quotient,
respectively. In particular, they show that the method is able to provide reconstructions of the
same accuracy level of the Pixon algorithm, which is generally considered as the most reliable
technique in providing accurate image photometry [22]. The key result with the use of EM
is that it is considerably faster than Pixon, with an average gain of a factor of 4 (potentially
growing up to a factor of 20 if EM is employed with the highly optimized interactive data
language (IDL) routine available in the Solar SoftWare (SSW) to compute the transmission
probabilities—see [5]). In this work we want to go one step further and address the solution
of (4) by means of recently proposed accelerations of EM. In the following sections, we recall
the main features of these schemes and discuss the regularization issue, which is a crucial
point to contrast the corruptive effect of the noise on the reconstructed image.

3. Optimization methods

In this section, we focus on some optimization methods which are able to address the
minimization problem (4). In recent literature, a number of methods have been proposed
for non-negatively constrained minimization and also for specific image restoration problems
involving the KL divergence. Among them, we mention the first order methods in [13, 15]; the
second order Newton-like methods [23, 24]; the especially tailored versions of the alternating
direction method of multipliers (ADMM) proposed in [25, 26]; and the gradient projection with
extrapolation methods [27, 28], which have proved to be optimal in the sense that the decrease
of the objective function with respect to its minimum value is quadratic with the iteration
number. To devise the appropriate approaches for the count-based image reconstruction
problem (4), its special features must be taken into account. In particular, one should remember
that:

- the matrix \( P \) and the Hessian \( \nabla^2 J(f) \) are very large, dense and have no special structures;
- the Lipschitz constant of \( \nabla J(f) \) is not available.
Algorithm 1. Scaled gradient projection (SGP) method.

Choose the starting point \( f^{(0)} \geq 0 \) and set the parameters \( \beta, \theta \in (0, 1), 0 < \alpha_{\text{min}} < \alpha_{\text{max}} \).

For \( k = 0, 1, 2, \ldots \), do the following steps:

Step 1. Choose the parameter \( \alpha_k \in [\alpha_{\text{min}}, \alpha_{\text{max}}] \) and the scaling matrix \( D_k \in \mathcal{D} \).

Step 2. Projection:
\[
y^{(k)} = \mathcal{P}_+ (f^{(k)} - \alpha_k D_k \nabla J (f^{(k)})).
\]

Step 3. Descent direction:
\[
d^{(k)} = y^{(k)} - f^{(k)}.
\]

Step 4. Set \( \lambda_k = 1 \).

Step 5. Backtracking loop:
- let \( J_{\text{new}} = J(f^{(k)} + \lambda_k d^{(k)}) \).
- If \( J_{\text{new}} \leq J(f^{(k)}) + \beta \lambda_k \nabla J (f^{(k)})^T d^{(k)} \), then go to step 6;
- Else set \( \lambda_k = \theta \lambda_k \) and go to step 5.

Step 6. Set \( f^{(k+1)} = f^{(k)} + \lambda_k d^{(k)} \).

Indeed, unlike the deconvolution case, here the computation of matrix–vector products involving \( P \) and \( P^T \) is very expensive. Moreover, there is no effective way to solve a linear system related to the Hessian \( \nabla^2 J(f) \) or the matrix \( I + P^T P \). This makes the ADMM methods [25, 26] and the Newton-like methods [23, 24] very impractical. On the other hand, the explicit knowledge of the Lipschitz constant of \( \nabla J(f) \) is needed for the implementation of the Nesterov algorithm [28].

For the reasons above, we restrict the choice of the optimization method for solving (4) to gradient-based methods. In particular, we compare the SGP method [13], the AS_CBB method [15] and the GPE [14, 27], which can be implemented also when the Lipschitz constant of \( \nabla J(f) \) is unknown.

It is worth stressing that the image reconstruction problem underlying (4) is ill-posed and an appropriate regularization is needed, as explained in detail in section 4. Thus the optimization methods should also be evaluated, focusing on their behavior in the regularization context.

3.1. The SGP algorithm

The SGP algorithm is a gradient projection method proposed by Bonettini et al [13] to solve a minimization problem of a differentiable function over a convex set. The scheme of the SGP method in the case of non-negative constraints is given in algorithm 1, where \( \mathcal{D} \) is the set of the \( n^2 \times n^2 \) diagonal positive definite matrices, whose diagonal elements have values between \( L_1 \) and \( L_2 \), for given thresholds \( 0 < L_1 < L_2 \). We do not discuss here the choice of all the several parameters of SGP, which can be found, for example, in [29]. We only describe the three main features of SGP, which are the scaling matrix multiplying the gradient, the steplength parameter and the projection on the feasible set.

- \( D_k \) is a diagonal scaling matrix defined as
\[
D_k = \text{diag} \left( \min \left( L_2, \max \left( L_1, \frac{f^{(k)}}{P^T P} \right) \right) \right).
\]

\((L_1, L_2)\) being given constants estimated from min/max values of one EM iteration. It has been shown that scaling the gradient might lead to a more faithful recovery of a regularized solution [30, 31].
• The steplength $\alpha_k$ is selected by alternating the following generalized BB rules [13, 32]:

$$\alpha_k(D_k)^{\text{BB1}} = \frac{(s^{(k-1)})^T D_k^{-1} D_k^{-1} (s^{(k-1)})}{(s^{(k-1)})^T D_k^{-1} (z^{(k-1)})},$$

$$\alpha_k(D_k)^{\text{BB2}} = \frac{(s^{(k-1)})^T D_k (z^{(k-1)})}{(z^{(k-1)})^T D_k (z^{(k-1)})},$$

where $s^{(k-1)} = f^{(k)} - f^{(k-1)}$ and $z^{(k-1)} = \nabla J(f^{(k)}) - \nabla J(f^{(k-1)})$, and introducing suitable upper and lower bounds. The alternation of the BB rules allows a faster convergence with respect to gradient methods exploiting only one of the two rules (see e.g. [33]);

Theoretical convergence properties of SGP. The convergence properties of SGP are stated in [13] for general, possibly nonconvex problems with a convex constraint set. In particular, it is proved that any limit point of the sequence $f^{(k)}$ is a stationary point for the constrained problem. As a consequence, for convex problems such as (4), any limit point of $f^{(k)}$ is a minimum point. If, in addition, the objective function is strictly convex with bounded level sets, then $f^{(k)}$ converges to the unique minimum point. A stronger result for the convex case can be found in [34], where the convergence of the whole sequence is proved for the nonscaled method ($D_k = I$) with the only assumption that a solution exists. As far as we know, a theoretical convergence rate estimate for the gradient projection method with variable scaling matrix and linesearch along the descent direction has not been well investigated in the literature.

3.2. The AS_CBB method

The AS_CBB method proposed by Hager et al in [15] shares with SGP the scaling idea and the exploitation of the BB rules. Indeed, AS_CBB can be framed in the scheme of algorithm 1 in which the computation of the search direction in steps 2 and 3 is substituted by $d^{(k)} = -D_k \nabla J(f^{(k)})$, where:

• the diagonal entries of the diagonal scaling matrix $D_k$ are defined by:

$$(D_k)_{ii} = \frac{f_i^{(k)}}{\alpha_k f_i^{(k)} + [\nabla J(f^{(k)})]^+]},$$

where $[.]^+$ denotes the positive part of a real number, i.e. $[x]^+ = \max(x, 0)$. In this case, the diagonal entries of the scaling matrix are not bounded away from zero.

• The parameter $\alpha_k$ in (6) is computed by a cyclic update of the first BB rule as $\alpha_k = \alpha_{q+c}(I)^{\text{BB1}}$ in (5) when $k = q \cdot c + r$ for a given, fixed cycle length $c > 0$ and $q = 1, 2, \ldots, r = 0, 1, \ldots, c - 1$ (in the numerical experiments of section 5 we set $c = 3$).

Actually, the scaling matrix (6) implies that, when the algorithm is initialized by a point with positive entries, then also the point $f^{(k)} - D_k \nabla J(f^{(k)})$ is still strictly positive for any $k$.

Theoretical convergence properties of AS_CBB. The AS_CBB method applies to box constrained problems, and convergence can be stated if the objective function is strongly convex and has bounded level sets. Moreover, in [15] the authors also prove the local $R$-linear convergence of the sequence $f^{(k)}$ to the unique minimum point.
Algorithm 2. GPE.

Choose the starting point $f^{(0)} \ge 0$ and the parameter $\theta \in (0, 1)$; set $f^{(-1)} = 0$.

For $k = 0, 1, 2, \ldots$ do the following steps:

Step 1. Extrapolation:

$$y^{(k)} = f^{(k)} + \beta_k (f^{(k)} - f^{(k-1)}).$$

Step 2. Choose an initial estimate of the Lipschitz constant $\alpha_k$.

Step 3. Backtracking loop:

- Compute $f(\alpha_k) = P_+ (f^{(k)} - \alpha_k \nabla J(f^{(k)}))$.

  If

  $$J(f(\alpha_k)) \le J(f^{(k)}) + \alpha_k \nabla J(y^{(k)})^T (f(\alpha_k) - y^{(k)}) + \frac{1}{2\alpha_k} \|f(\alpha_k) - y^{(k)}\|^2$$

  then

  go to step 4;

  else

  set $\alpha_k = \theta \alpha_k$ and go to step 3.

  endif

Step 4. Set $f^{(k+1)} = f(\alpha_k)$.

End

3.3. The GPE method

While SGP and AS_CBB exploit gradient scaling and the BB rules to improve the convergence speed, the GPE method uses a completely different acceleration strategy which dates back to Polyak [35] and consists of adding an extrapolation term to the gradient direction. Several recently proposed methods, for example the very popular Nesterov’s method [28] and the fast iterative shrinkage-thresholding algorithm (FISTA) [27], include such an extrapolation step. Here we consider the general scheme described in [14, section 6.9.2] applied to (4), with an automatic estimate of the Lipschitz constant of $\nabla J(f)$ (see algorithm 2).

Possible choices for the extrapolation parameter $\beta_k$ are:

$$\beta_k = \begin{cases} 0 & \text{if } k = 0 \\ \frac{k-1}{k+2} & \text{if } k > 0 \end{cases}$$

(7)

(see [14, p 397]), or also

$$\beta_k = \frac{t_k - 1}{t_{k+1}}, \quad \text{where } t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$

(8)

with $t_0 = 1$, which corresponds to the FISTA algorithm [27].

Theoretical convergence properties of GPE. The GPE method is specific for convex problems, and its more interesting property is the estimate:

$$J(f^{(k)}) - J^* = O \left( \frac{1}{k^2} \right),$$

where $J^*$ is the minimum of $J$, which can be proved for both the choices (7)–(8). It is worth stressing that this estimate concerns the objective function value, instead of the sequence of the iterates.

4. Handling with regularization

The presence of noise on the measured data makes any minimizer of (4) meaningless. Two main strategies are typically adopted to overcome this drawback: (a) arrest the iterative procedure employed to solve (4) before convergence; or (b) add a (suitably weighted) regularization term
to the KL which preserves certain features of the unknown target distribution and solve the resulting minimum problem exactly. Both approaches then need a criterion for the choice of the so-called regularization parameter, which in the former case is the number of iterations to be performed, while in the latter case is the constant multiplying the regularization term. In our approach, we investigated the effectiveness of a recently proposed discrepancy principle for data affected by Poisson noise [17, 36, 37], with the main result being the following lemma.

**Lemma 4.1.** Let $Y_\lambda$ be a Poisson random variable with expected value $\lambda$ and consider the following function of $Y_\lambda$:

$$F(Y_\lambda) = 2 \left\{ Y_\lambda \ln \left( \frac{Y_\lambda}{\lambda} \right) + \lambda - Y_\lambda \right\}.$$  

Then, for large $\lambda$, the following asymptotic estimate of the expected value of $F(Y_\lambda)$ holds true:

$$E[F(Y_\lambda)] = 1 + O \left( \frac{1}{\lambda} \right).$$

Therefore, according to this discrepancy principle, an optimal regularized solution $f_\eta$ should be the one satisfying:

$$\sum_{j=1}^{N} D_{KL}(c, P_{f_\eta}) = 1 + \varepsilon,$$

where $\varepsilon$ depends on the noise level of the measured data. For medium/high levels of statistics, the default value $\varepsilon = 1/m$, where $m = \text{mean}(c)$, provided within the SGP code results in a reliable choice. This is not true for count profiles characterized by a low level of statistics. In this case, the term $O \left( \frac{1}{\lambda} \right)$ in lemma 4.1 has to be accurately estimated.

We tried to estimate $\varepsilon$ by following the proof of lemma 4.1, which is based on the Taylor formula of the function:

$$\ln \left( \frac{1}{Y_\lambda} \right) = \ln \left( 1 + \frac{Y_\lambda - \lambda}{\lambda} \right) =: \ln (1 + \xi).$$

Since:

$$\ln (1 + \xi) = \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1}}{\ell} \xi^\ell,$$

after some computations, we get the following expression for $F(Y_\lambda)$:

$$F(Y_\lambda) = \sum_{\ell=1}^{\infty} \frac{2(-1)^{\ell+1}}{\ell(\ell + 1)} \cdot \frac{(Y_\lambda - \lambda)^{\ell+1}}{\lambda^\ell}.$$  

By introducing the central moments:

$$\mu_\ell = \mathbb{E}[(Y_\lambda - \lambda)^\ell]$$

and by recalling that, for a Poisson random variable, $\mu_2 = \mu_3 = \lambda$ and $\mu_{\ell+1}, \ell > 2$, can be derived from the recursive rule [38]:

$$\mu_{\ell+1} = \ell \mu_{\ell} + \lambda \mu_\ell,$$

the expected value of $F(Y_\lambda)$ becomes:

$$E[F(Y_\lambda)] = \sum_{\ell=1}^{\infty} \frac{2(-1)^{\ell+1}}{\ell(\ell + 1)} \cdot \frac{\mu_{\ell+1}}{\lambda^{\ell+1}} = 1 + \sum_{\ell=2}^{\infty} \frac{2(-1)^{\ell+1}}{\ell(\ell + 1)} \cdot \frac{\mu_{\ell+1}}{\lambda^{\ell}}.$$  

(10)
Therefore, the quantity $\varepsilon$ in (9) can be estimated from the last term in (10). We implemented a routine in Matlab (The MathWorks, Natick, MA) which approximates $\varepsilon$ by truncating the series at a given number of terms $\ell$. In our tests we decided to arrest the series at $\ell = 10$ since we noticed some numerical instability for a higher numbers of terms.

With regards to the way of introducing regularization, we first investigate the use of the proposed gradient methods as iterative regularization methods for solving (4) (in this case the role of the regularization parameter is played by the iteration number $k$); then we use the SGP method to solve the minimization problem:

$$
\min_{f \geq 0} J_R(f) := D_{KL}(c, Pf) + \eta \|Lf\|_2^2,
$$

where $L$ is the identity matrix (zero order Tikhonov regularization) or the discrete gradient operator (first order Tikhonov regularization), for different values of $\eta > 0$. In the former case, the value $1 + \varepsilon$ estimated from (10) is used as stopping criterion for all the gradient algorithms, which are arrested at the first iteration $k$ satisfying the condition:

$$
\frac{2}{\sum_{j=1}^9 N_j} D_{KL}(c, Pf^{(k)}) \leq 1 + \varepsilon.
$$

When the Tikhonov term is added in the objective function, we compute the solution $f_\eta$ of (11) with SGP for different values of $\eta$ and choose the regularized solution $f_\eta^*$ satisfying:

$$
\frac{2}{\sum_{j=1}^9 N_j} D_{KL}(c, Pf_\eta^*) \approx 1 + \varepsilon.
$$

For the computation of the Tikhonov solutions, the stopping criterion:

$$
|J_R(f^{(k+1)}) - J_R(f^{(k)})| \leq 10^{-9} |J_R(f^{(k+1)})|
$$

has been used.

We point out that the modifications needed by SGP to be applied to (11) are minimal, since we only have to change $J$ with $J_R$ in algorithm 1 and in the computation of the steplength $\alpha_k$ and to adopt the scaling matrix:

$$
D_k = \text{diag} \left( \min \left( L_2, \max \left( L_1, \frac{f^{(k)}}{P^T \ell + \epsilon \eta f^{(k)}} \right) \right) \right)
$$

suggested by the scaled gradient method proposed by Lantéri et al [39, 40]. The constant $\ell$ in (14) is equal to 1 for zero order regularization while $\ell = 4$ has to be set if a first order Tikhonov term is adopted.

5. Numerical experiments

In this section, we test the effectiveness of the considered optimization schemes in providing good reconstructions of realistic images with a notable reduction in the number of iterations required with respect to EM. As a further comparison, we also consider the images recovered by a visibility-based strategy. Among the possible choices, we chose the uv-smooth algorithm [8], which is the approach providing the most reliable results among the algorithms implemented in SSW which use the visibilities as input data. For the sake of completeness, we summarize the main features of uv-smooth in the following section.

5.1. The uv-smooth algorithm

The uv-smooth method consists of a two-step process: (a) interpolation to generate a smooth continuum of visibilities within the disk in the Fourier plane spanned by the available data;
Algorithm 3. The Gerchberg–Papoulis method.

Put the starting point $f^{(0)}$ equal to the null map, and set the parameter $\tau > 0$.

For $k = 0, 1, 2, \ldots$, do the following steps:

Step 1. Compute the Fourier transform $\mathcal{F} f^{(k)}$ of $f^{(k)}$;

Step 2. Compute
$$
\mathcal{F} f^{(k+1)} = \mathcal{F} f^{(k)} + \tau (V - \chi_B \mathcal{F} f^{(k)});
$$

Step 3. Compute the inverse Fourier transform $f^{(k+1)}$ of $\mathcal{F} f^{(k+1)}$;

Step 4. Project $f^{(k+1)}$ on the set of the real positive numbers.

End.

(b) out-of-band extrapolation through an FFT-based iterative method with the imposition of image positivity. In more detail, the measured visibilities are interpolated in the Fourier plane through a thin-plate spline algorithm, by means of the IDL routine grid_tps.pro [41]. The resulting surface is then resampled on a uniform grid. The advantage achievable with the interpolation step is that information is also inferred for spatial frequencies corresponding to ‘virtual’ subcollimators with angular resolution between the minimum and the maximum values available with the RHESSI hardware. Moreover, with the new (uniform) resampling on the visibilities in the Fourier plane, the FFT routine can be applied. With regards to the extrapolation step, the idea is the extraction of information also for higher frequencies, to get a super-resolution effect of the reconstructed image [21, 42]. This can be done with a projected Gerchberg–Papoulis method [43, 44], which consists of finding a positive distribution $f(x, y)$ such that:

$$
V(u, v) = \chi_B(u, v) (\mathcal{F} f)(u, v),
$$

where $(u, v)$ denotes a pair of spatial frequencies, $V(u, v)$ is the corresponding visibility, $B$ is the band in which RHESSI provides the visibilities, $\chi_B$ is the characteristic function of $B$ and $\mathcal{F}$ is the Fourier transform. The steps of the computational procedure adopted are summarized in algorithm 3. The steplength parameter $\tau$ has to be properly chosen in order to assure the convergence of the algorithm. Regularization is achieved through an early stopping of the iterations based on the descent of $\| \chi_B \mathcal{F} f - V \|$.

5.2. Simulated datasets

For the creation of the synthetic datasets, we employed the following procedure.

• We considered several real flare events and, starting from the radiation collected by RHESSI, we reconstructed the corresponding $64 \times 64$ images with the CLEAN algorithm [45], available in SSW. In particular, we considered three energy ranges of the 23 July 2002 flare (00:29:10–00:30:19 UT, 20–22, 41–46 and 156–177 keV) and one energy range of the 15 April 2002 flare (00:05:00–00:10:00 UT, 12–14 keV), the 31 August 2004 (05:33:00–05:38:00 UT, 10–12 keV) and the 1 June 2005 flares (02:37:00–02:41:00 UT, 10–15 keV).

• These images have been cleared of possible artifacts introduced by the reconstruction method, and suitably scaled in order to have different statistics (and, therefore, different noise levels). The resulting distributions are shown in the first row of figures 2 and 3.

• For the $j$th detector ($j = 1, \ldots, 9$), we computed the corresponding modulated count profiles $c^{(j)}$ by means of a realistic transmission probability $P^{(j)}(x, y; t)$ (even if we simplified the procedure by neglecting the hardware-based parameter $K^{(j)}$ in (1)).

As
Figure 2. The original images (first row) of the Sim1, Sim2 and Sim3 datasets with the reconstructions provided by EM (second row), SGP (third row), GPE (fourth row) and AS_CBB (fifth row) as iterative regularization methods applied to (4) and combined with the discrepancy principle (12). The images obtained with the visibility-based uv-smooth algorithm are also shown (last row).
concerns the discretization of the observational time, we used $N_j = 640$ time bins for each detector.

- We perturbed the count profiles with Poisson noise by using the `imnoise` function available in the Image Processing Toolbox of Matlab. We will denote the resulting six synthetic datasets as $\text{Sim1}, \ldots, \text{Sim6}$, following the order of the original events described in the first item.

Once these six datasets had been created, we reconstructed the target distributions with the EM, SGP, GPE and AS_CBB methods by stopping the iterations according to the discrepancy principle described in section 4. All the algorithms have been initialized with a constant image whose total flux has been deduced by the modulated count profiles (see the appendix of [6] for more details). The numerical experiments for these methods have been performed by means of routines implemented by ourselves in Matlab.

The choice of the subcollimators used in the inversion process requires a brief comment. Due to the thin slits of the grids, detectors 1 and 2 are typically discarded since the corresponding count profiles are characterized by a very low SNR. Therefore, we performed 3 sets of simulations, by using all the subcollimators (set 1), subcollimators 2–9 (set 2) and 3–9 (set 3). We found the best results in the second case for the datasets $\text{Sim1}, \text{Sim4}$ and $\text{Sim5}$, while lower reconstruction errors have been registered in the third case for $\text{Sim2}, \text{Sim3}$ and $\text{Sim6}$ (whose original images have lower flux).

As concerns the uv-smooth method, we found the best reconstructions by using detectors 3–9, in agreement with the fact that a fitting step on the raw count profiles amplifies the effect of the noise on the processed data (and, therefore, the information provided by subcollimator 2, which can be exploited in the count-based strategies, here have a corruptive effect on the reconstructions). We remark that in this case we used the original IDL routine available in SSW.

5.3. Results

The reconstructed images are shown in the last five rows of figures 2 and 3, while in table 1 we reported the relative reconstruction errors (in Euclidean norm) and the iterations required by the gradient methods. In order to evaluate the effectiveness of the stopping criterion for the count-based algorithms, we also added in the table the analogous values corresponding to the best reconstructions (i.e. the ones with the lowest reconstruction error). Finally, in figures 4 and 5 we plotted the reconstruction errors and the discrepancy functions $\sum_{j=1}^9 D_{\text{KL}}(c, Pf^{(k)})$ against the iteration number, where $d_{\text{min}}$ is 2 or 3 according to the detectors used in our simulations (see discussion above).

From the reconstructed images and the corresponding errors we can make the following observations:

- The accelerated methods greatly reduce the number of iterations required by EM to provide the reconstruction, with a gain ranging from a factor of about 2 to a factor of about 17. We remark that the cost per iteration of all the methods are essentially the same, since the main computations are the two matrix–vector products $Pf^{(k)}$ and $PT(cPf^{(k)})$ (few further scalar products are needed by SGP and AS_CBB to compute the steplength, while few further matrix–vector products could be required by GPE due to its backtracking procedure). Therefore, since the IDL implementation of EM in SSW requires computational times of a few tens of seconds [11], we expect that, once the accelerated algorithms are implemented in IDL and included in SSW by the RHESSI Software Team, the overall CPU effort required to produce the desired image will be very low. It will also be comparable to that
Figure 3. The original images (first row) of the Sim4, Sim5 and Sim6 datasets with the reconstructions provided by EM (second row), SGP (third row), GPE (fourth row) and AS_CBB (fifth row) as iterative regularization methods applied to (4) and combined with the discrepancy principle (12). The images obtained with the visibility-based uv-smooth algorithm are also shown (last row).
Figure 4. Plots of the reconstruction errors provided by EM, SGP, GPE and AS_CBB as functions of the iteration number for the six datasets described in the text. The squares indicate the values provided by the discrepancy principle (12), while the circles indicate those corresponding to the minimum errors.

Figure 5. Plots of the discrepancy functions provided by EM, SGP, GPE and AS_CBB as functions of the iteration number for the six datasets described in the text.

of the visibility-based methods (whose computational cost includes the time required to compute the visibilities by fitting the count profiles).

- Quite surprisingly, besides the computational gain, the accelerated methods are also able to provide some improvements in the quality of the final images, as attested numerically by the lower reconstruction errors in table 1 and visually by the pictures in figures 2 and 3. We remark that these improvements do not depend on the stopping criterion we adopted,
Table 1. Reconstruction errors provided by EM, SGP, GPE and AS_CBB used as iterative regularization methods for solving (4) in the six simulations described in the text. Both the values obtained with the discrepancy principle (12) and the optimal values (i.e. the ones minimizing the reconstruction error) are shown. The reconstruction errors obtained with the visibility-based uv-smooth algorithm are also reported. The lower values for each simulation are highlighted in bold.

|        | Sim1 | Sim2 | Sim3 | Sim4 | Sim5 | Sim6 |
|--------|------|------|------|------|------|------|
| EM discr | Err  | 0.212 | 0.274 | 0.409 | 0.214 | 0.341 | 0.273 |
|        | Iter | 4369 | 669 | 803 | 17632 | 533 | 1256 |
| SGP discr | Err | **0.169** | **0.243** | 0.338 | **0.112** | 0.295 | **0.176** |
|        | Iter | 253 | 78 | 369 | 1511 | 234 | 354 |
| GPE discr | Err | 0.251 | 0.245 | **0.320** | 0.278 | **0.248** | 0.196 |
|        | Iter | 634 | 118 | 190 | 1038 | 121 | 172 |
| AS_CBB discr | Err | 0.170 | 0.248 | 0.330 | **0.112** | 0.278 | 0.181 |
|        | Iter | 1098 | 120 | 82 | 2568 | 89 | 210 |
| EM best | Err  | 0.202 | 0.255 | 0.395 | 0.155 | 0.337 | 0.236 |
|        | Iter | 12536 | 1766 | 516 | 1952 | 450 | 3271 |
| SGP best | Err  | **0.141** | **0.182** | 0.329 | **0.088** | 0.273 | 0.164 |
|        | Iter | 2361 | 254 | 462 | 377 | 353 | 442 |
| GPE best | Err  | 0.192 | 0.244 | **0.312** | 0.161 | **0.238** | 0.192 |
|        | Iter | 109 | 123 | 212 | 198 | 136 | 161 |
| AS_CBB best | Err  | 0.148 | **0.182** | 0.330 | **0.088** | 0.274 | **0.163** |
|        | Iter | 5000 | 426 | 82 | 1267 | 76 | 318 |
| uv-smooth | Err  | 0.278 | 0.280 | 0.446 | 0.173 | 0.323 | 0.252 |

since they still hold true if one also looks at the best errors achievable by the different methods.

- The discrepancy principle we chose as stopping criterion seems to work quite well in general, since it is able to stop the iterates very close to the optimal ones. Of course some exceptions are present, and in some cases the algorithms are arrested when the reconstruction error is already growing back, while in other cases they are stopped too early (see figure 4). However, we think that this behavior is honestly unavoidable when a general automatic stopping procedure is adopted.

- We did not find any clear prevalence of one accelerated scheme with respect to the others. The SGP and AS_CBB algorithms provide similar reconstruction errors, while the performance of GPE seems to be more unstable (in Sim3 and Sim5 it provides the lowest errors, while in Sim1 and Sim4 the highest ones). Also in the number of iterations required we found different problem-dependent behaviors.

- As concerns the comparison with uv-smooth, we observe that, as could be expected, the accuracy of the reconstructions is in general improved by the proposed algorithms, since no systematic error is present on the raw data.

As a final test, we applied SGP to problem (11) in order to investigate the effect of the addition of a zero order or first order Tikhonov term in the objective function to ensure a certain smoothness of the reconstructed image. In this case, the regularized solutions have been obtained according to the procedure described in section 4. We reported the reconstruction errors in table 2, together with the optimal values for both the regularization parameters $\eta$ and the tolerance for the stopping criterion $\varepsilon$. Our simulations do not show clear indications of the effectiveness of the presence of the Tikhonov term, since only in some cases improvements in the reconstruction can be noticed. We can conclude that, since the computational procedure
to find a regularized solution of (11) is much heavier than the use of SGP as an iterative regularization method to solve (4), the latter approach is definitely preferable.

6. Conclusions

This paper deals with the real-world problem of reconstructing an x-ray image of a solar flare by means of the data collected by RHESSI. This problem can be addressed in two ways, both of which present advantages and drawbacks. The visibility-based methods are very fast but act on preprocessed data characterized by systematic errors. The count-based strategies can be applied straightforwardly on the raw data, but require the computation of the transmission probabilities, which overloads the cost per iteration of the reconstruction algorithms. Our work belongs to the latter category and, starting from the statistical model proposed in [11], aims to significantly decrease the number of iterations needed by means of very powerful optimization schemes, thus leading to a computational effort comparable with that of the methods acting on the visibilities.

In order to achieve a meaningful solution, we introduced regularization both by early stopping of the proposed methods applied to the minimization of the KL divergence, and by using SGP as an optimization method to exactly minimize a weighted sum of the KL and the Tikhonov functional. As automatic criterion for the choice of the number of iterations or the regularization parameter, we adopted a recently proposed discrepancy principle for Poisson data, suitably adapted to take into account different levels of noise.

Our numerical experiments on synthetic datasets show that the accelerated gradient algorithms, used as iterative regularization methods instead of EM, are able to substantially reduce not only the number of iterations but also the reconstruction error, thus leading to a faster recovery of higher quality images. An analysis of the same data showed that the approach of adding a Tikhonov term to the KL does not provide sufficient improvements to justify the much higher computational effort required. Finally, the comparison with a state-of-the-art visibility-based method showed that the reconstruction errors provided by the proposed methods are systematically lower. Starting from the analysis of the execution times carried out in [11] and taking into account the gain in the iterations shown in this paper with respect to EM, we are confident that these methods implemented within SSW will run in a computational time comparable with that of the visibility-based methods. Moreover, the integration of these algorithms in SSW will allow an intensive comparison with all the other reconstruction algorithms on the real datasets provided by RHESSI.

Future work will involve the application of an accelerated gradient method to the RHESSI spectroscopy problem, in which the measured count spectra are related to the electron spectra, which emitted those counts by a Volterra integral equation (see e.g. [46]).
Acknowledgments

This work has been partially supported by MIUR (Italian Ministry for University and Research), under the projects FIRB—Futuro in Ricerca 2012, contract RBFR12M3AC, and PRIN 2012, contract 2012MTE38N. The Italian GNCS—INdAM (Gruppo Nazionale per il Calcolo Scientifico—Istituto Nazionale di Alta Matematica) is also acknowledged.

References

[1] Hurford G J 2010 X-ray imaging with collimators, masks and grids Observing Photons in Space (ISSI Scientific Reports Series vol 9) ed M C E Huber, A Pauluhn, J L Culhane, J G Timothy, K Wilhelm and A Zehnder (ISSI) (New York: Springer) pp 223–34
[2] Ramsey B D, Austin R A and Decher R 1994 Instrumentation for x-ray astronomy Space Sci. Rev. 69 139–204
[3] Krucker S et al 2008 Hard x-ray emission from the solar corona Astron. Astrophys. Rev. 16 155–208
[4] Lin R P et al 2002 The Reuven Ramaty high-energy solar spectroscopic imager (RHESSI) Solar Phys. 210 3–32
[5] Hurford G J et al 2002 The RHESSI imaging concept Solar Phys. 210 61–86
[6] Bonettini S, Cornelio A and Prato M 2013 A new semiblind deconvolution approach for Fourier-based image restoration: an application in astronomy SIAM J. Imaging Sci. 6 1736–57
[7] Bong S C, Lee J, Gary D E and Yun H S 2006 Spatio-spectral maximum entropy method: I. Formulation and test Astrophys. J. 636 1159–65
[8] Massone A M, Emslie A G, Hurford G J, Prato M, Kontar E P and Piana M 2009 Hard x-ray imaging of solar flares using interpolated visibilities Astrophys. J. 703 2004–16
[9] Bonettini S and Prato M 2010 Nonnegative image reconstruction from sparse Fourier data: a new deconvolution algorithm Inverse Problems 26 095001
[10] Prato M 2013 A deconvolution algorithm for imaging problems from Fourier data Boll. Unione Mat. Ital. 6 389–404
[11] Benvenuto F, Schwartz R, Piana M and Massone A M 2013 Expectation maximization for hard x-ray count modulation profiles Astron. Astrophys. 555 A61
[12] Csiszár I 1991 Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems Ann. Stat. 19 2032–66
[13] Bonettini S, Zanella R and Zanni L 2009 A scaled gradient projection method for constrained image deblurring Inverse Problems 25 015002
[14] Bertsekas D 2009 Supplementary chapter 6 on convex optimization algorithms Convex Optimization Theory (Belmont: Athena Scientific)
[15] Hager W W, Mair B A and Zhang H 2009 An affine-scaling interior-point CBB method for box-constrained optimization Math. Program. 119 1–32
[16] Tikhonov N A and Arsenin V Y 1977 Solution of Ill Posed Problems (New York: Wiley)
[17] Bertero M, Boccacci P, Talenti G, Zanella R and Zanni L 2010 A discrepancy principle for Poisson data Inverse Problems 26 105004
[18] Hadamard J 1923 Lectures on Cauchy’s Problem in Linear Partial Differential Equations (New Haven, CT: Yale University Press)
[19] Reed M and Simon B 1972 Methods of Modern Mathematical Physics I: Functional Analysis (New York: Academic)
[20] Kirsch A 1996 An Introduction to the Mathematical Theory of Inverse Problems (New York: Springer)
[21] Bertero M and Boccacci P 1998 Introduction to Inverse Problems in Imaging (Bristol: Institute of Physics Publishing)
[22] Dennis B R and Pernak R L 2009 Hard x-ray flare source sizes measured with the Ramaty high energy solar spectroscopic imager Astrophys. J. 698 2131–43
[23] Landi G and Piccolomini E L 2012 An improved Newton projection method for nonnegative deblurring of Poisson-corrupted images with Tikhonov regularization Numer. Algorithms 60 169–88
[24] Bonettini S and Serafini T 2009 Non-negatively constrained image deblurring with an inexact interior point method J. Comput. Appl. Math. 231 236–48
[25] Setzer S, Steidl G and Teuber T 2010 Deblurring Poissonian images by split Bregman techniques J. Vis. Commun. Image Represent. 21 193–9
[26] Figueiredo M A T and Bioucas-Dias J M 2010 Restoration of Poissonian images using alternating direction optimization IEEE Trans. Image Process. 19 3133–45
[27] Beck A and Teboulle M 2009 A fast iterative shrinkage-thresholding algorithm for linear inverse problems SIAM J. Imaging Sci. 2 183–202
[28] Nesterov Y 2005 Smooth minimization of non-smooth functions Math. Program. 103 127–52
[29] Prato M, Cavicchioli R, Zanni L, Boccacci P and Bertero M 2012 Efficient deconvolution methods for astronomical imaging: algorithms and IDL-GPU codes Astron. Astrophys. 539 A133
[30] Bonettini S, Landi G, Piccolomini E Loli and Zanni L 2013 Scaling techniques for gradient projection-type methods in astronomical image deblurring Int. J. Comput. Math. 90 9–29
[31] Cornelio A, Porta F, Prato M and Zanni L 2013 On the filtering effect of iterative regularization algorithms for discrete inverse problems Inverse Problems 29 125013
[32] Barzilai J and Borwein J M 1988 Two-point step size gradient methods IMA J. Numer. Anal. 8 141–8
[33] Frassoldati G, Zhangirati G and Zanni L 2008 New adaptive stepsize selections in gradient methods J. Indist. Manag. Optim. 4 299–312
[34] Iusem A N 2003 On the convergence properties of the projected gradient method for convex optimization Comput. Appl. Math. 22 37–52
[35] Polyak B 1987 Introduction to Optimization (New York: Optimization Software Inc.)
[36] 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
[37] Zanella R, Boccacci P, Zanni L and Bertero M 2013 Corrigendum: efficient gradient projection methods for edge-preserving removal of Poisson noise Inverse Problems 29 119501
[38] Johnson N L, Kats S and Kemp A W 1993 Univariate Discrete Distributions (New York: Wiley)
[39] Lantéri H, Roche M, Cuevas O and Aime C 2001 A general method to devise maximum likelihood signal restoration multiplicative algorithms with non-negativity constraints Signal Process. 81 945–74
[40] Lantéri H, Roche M and Aime C 2002 Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms Inverse Problems 18 1397–419
[41] Barrodale I, Skea D, Berkley M, Kuwahara R and Poeckert R 1993 Warping digital images using thin plate splines Pattern Recognit. 26 375–6
[42] Youla D C and Webb H 1982 Image restoration by the method of convex projections: part 1—theory IEEE Trans. Med. Imaging 1 81–94
[43] Gerchberg R W 1974 Super-resolution through error energy reduction Opt. Acta 21 709–20
[44] Papoulis A 1975 A new algorithm in spectral analysis and band-limited extrapolation IEEE Trans. Circuits Syst. 22 735–42
[45] Högberg J A 1974 Aperture synthesis with a non-regular distribution of interferometer baselines Astron. Astrophys. Suppl. Ser. 15 417–26
[46] Prato M 2009 Regularization methods for the solution of inverse problems in solar X-ray and imaging spectroscopy Arch. Comput. Methods Eng. 16 109–60