Nonnegative image reconstruction from sparse Fourier data: a new deconvolution algorithm

2010 Inverse Problems 26 095001

(http://iopscience.iop.org/0266-5611/26/9/095001)

Inverse Problems **26** (2010) 095001 (18pp) doi:10.1088/0266-5611/26/9/095001

Nonnegative image reconstruction from sparse Fourier data: a new deconvolution algorithm

S Bonettini[1] **and M Prato**[2,3]

[1] Dipartimento di Matematica, Università di Ferrara, Via Saragat 1, 44122 Ferrara, Italy
[2] Dipartimento di Matematica Pura e Applicata, Università di Modena e Reggio Emilia, Via Campi 213/b, 41125 Modena, Italy
[3] CNR—SPIN, Via Dodecaneso 33, 16146 Genova, Italy

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

Received 8 March 2010, in final form 26 May 2010
Published 9 July 2010
Online at stacks.iop.org/IP/26/095001

Abstract
This paper deals with image restoration problems where the data are nonuniform samples of the Fourier transform of the unknown object. We study the inverse problem in both semidiscrete and fully discrete formulations, and our analysis leads to an optimization problem involving the minimization of the data discrepancy under nonnegativity constraints. In particular, we show that such a problem is equivalent to a deconvolution problem in the image space. We propose a practical algorithm, based on the gradient projection method, to compute a regularized solution in the discrete case. The key point in our deconvolution-based approach is that the fast Fourier transform can be employed in the algorithm implementation without the need of preprocessing the data. A numerical experimentation on simulated and real data from the NASA RHESSI mission is also performed.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The reconstruction of an image from the knowledge of a nonuniform sampling of its Fourier transform is a common problem in several scientific areas, such as radioastronomy, x-ray astronomy, computed tomography and magnetic resonance imaging [7, 10, 16]. In the general case, given N points irregularly distributed in the frequency space, we want to recover the corresponding image in the physical space. This problem presents both theoretical and computational difficulties: from the theoretical point of view, we are dealing with an inverse problem which is ill-posed in the sense of Hadamard [17]. In fact, due to the sparsity of the samples in the frequency space, infinite solutions are admissible; moreover, some of these solutions could not be physically meaningful because of the error which potentially affects

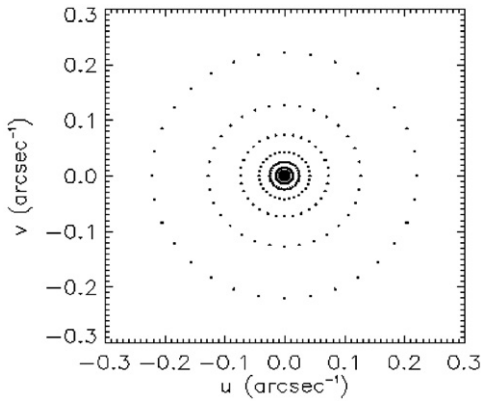


Figure 1. Example of a typical sampling of RHESSI data in the frequency plane.

the data. On the other hand, from the computational point of view, the irregularity of the data sampling prevents the straightforward application of the fast Fourier transform (FFT) algorithm, with the result of a heaviness of the inversion procedure.

Several methods have been proposed in the recent literature to face these difficulties, most of which are based on a two-step procedure (see [32] and references therein). At the beginning, the data are interpolated with a suitable function fixed *a priori*. This interpolation step (or *gridding*) has a twofold purpose: to fill the areas in the frequency space which lack in measured data and to allow in this way a uniform re-sampling of the data themselves. Thanks to this procedure, it is possible to invert the Fourier transform and reconstruct the desired image through some FFT-based direct or iterative algorithm.

Such a strategy results to be particularly efficient when

- the measured data are available in a large number and well distributed (even if not uniformly) in the frequency space. In this way, the choice of the interpolation function should not engrave heavily on the re-sampled data and different interpolation functions should lead to very similar reconstructed images;
- the signal-to-noise ratio is sufficiently high. In contrast, if the measured data were affected by a high level of noise, the interpolation phase may lead to an amplification of this noise level on the re-sampled data with the result of artefacts formation and undesirable effects in the corresponding reconstructed image.

Unfortunately, these two conditions fail in a large number of datasets provided by the NASA satellite *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 and γ-ray images of solar eruptions [21, 23]. The RHESSI satellite is still working and keeps on providing every day a large amount of data with very high quality.

The crucial point in RHESSI technology is the fact that the spatial information is encoded in the measured data as rapid temporal variations of the captured radiation [19]. In practice, the raw data given by RHESSI for a fixed x-ray energy interval is a set of complex numbers, called *visibilities*, which represents a sparse and nonuniform sampling of the radiation spatial Fourier transform. Due to RHESSI's hardware, the measured data correspond to spatial frequencies distributed in the frequency plane on nine concentric circles whose radii form a geometric sequence with a common ratio $\sqrt{3}$ (figure 1).

For this reason, especially in the high frequencies where the sampling is more sparse, the impact of the interpolation method can be nonnegligible. Moreover, particularly in the recent period of depressed solar activity, most of the RHESSI datasets correspond to solar eruptions with poor intensity and, therefore, are characterized by a low signal-to-noise ratio (due to the Poissonian nature of the noise affecting the data).

In the present paper we introduce a new method (that we called *Space-D*) to reconstruct an image from a nonuniform sampling of its Fourier transform which acts straightly on the data without interpolation and re-sampling operations, exploiting in this way the real nature of the data themselves. To this aim, we study the problem from the theoretical viewpoint in both continuous and discrete settings and we reformulate it as a constrained minimum problem. In particular, we show that the minimization of the data discrepancy is equivalent to a deconvolution problem with a suitable kernel.

In our approach, the solution is effectively computed by means of a gradient projection method with an adaptive steplength parameter. In particular, our choice is based on a suitable alternation of the two Barzilai–Borwein rules [1] which recently obtained very good results in signal and image denoising and deblurring problems [2, 8, 37]. Since the objective function involves a convolution operator, the algorithm can be effectively implemented exploiting the fast Fourier transform.

The main strength of our approach resides in a convenient problem formulation coupled with an efficient optimization algorithm. Moreover, it has been experimentally shown that the gradient projection method exhibits a semi-convergence behaviour [2, 8], thus providing regularized solutions. Finally, super-resolution is achieved through projections of the images obtained at each iteration of the algorithm on the admissible region [25].

The plan of the paper is the following: in section 2 we formulate the mathematical problem in a semidiscrete framework, providing the formal expression of its generalized solution and deducing the equivalent minimum problem through the introduction of a particular convolution operator. In particular, in section 2.3, we fully discretize the problem and we follow the same road covered in the continuous case to get the analogous minimum problem. In section 3, the optimization algorithm is described and the regularization issue is discussed. Applications of such a method on synthetic and real data provided by RHESSI are given in section 4, together with a comparison with the two state-of-the-art reconstruction algorithms from visibilities already available in the RHESSI software. Some conclusions and ideas for future developments of the method are offered in section 5.

2. Mathematical formulation

The problem of reconstructing a function given a finite sampling of its inverse Fourier transform can be considered as a semidiscrete problem [3, 4]: an unknown object belonging to an infinite-dimensional functional space has to be recovered from a finite number of data. More precisely, we define a mapping from an appropriate functional space into a finite-dimensional vector space that relates a function with some samples of its Fourier transform. To this end, let us define the square $\mathcal{D} = [X_1, X_2] \times [Y_1, Y_2]$ in the image plane and the points

$$(u_k, v_k), \qquad k = 1, \ldots, N$$

in the frequency plane. For a given two-dimensional function

$$\mathcal{D} \ni (x, y) \mapsto f(x, y) \in \mathbb{R},$$

the discrete data $g \in \mathbb{C}^N$ (also called *visibilities*) are defined as

$$g_k = \int_{\mathcal{D}} f(x, y)\, \mathrm{e}^{2\pi \mathrm{i}(u_k x + v_k y)}\, \mathrm{d}x\, \mathrm{d}y, \qquad k = 1, \ldots, N \tag{1}$$

and can be considered a nonuniform sampling of the Fourier transform[4]

$$V(u, v) = \int_{\mathcal{D}} f(x, y)\, \mathrm{e}^{2\pi \mathrm{i}(ux+vy)}\, \mathrm{d}x\, \mathrm{d}y, \qquad (u, v) \in \mathbb{R}^2. \tag{2}$$

We consider the Hilbert space $\mathcal{X} = L^2(\mathcal{D}, \mathbb{C})$ and the semidiscrete operator $\mathcal{A} : \mathcal{X} \to \mathbb{C}^N$ defined as

$$(\mathcal{A}f)_k = \int_{\mathcal{D}} f(x, y)\, \mathrm{e}^{2\pi \mathrm{i}(u_k x+v_k y)}\, \mathrm{d}x\, \mathrm{d}y, \qquad k = 1, \ldots, N, \quad f \in \mathcal{X}. \tag{3}$$

The adjoint of the operator $\mathcal{A}$, $\mathcal{A}^* : \mathbb{C}^N \to \mathcal{X}$, also called back-projection, is defined as follows: given $c \in \mathbb{C}^N$,

$$(\mathcal{A}^* c)(x, y) = \sum_{k=1}^{N} c_k\, \mathrm{e}^{-2\pi \mathrm{i}(u_k x+v_k y)}, \qquad (x, y) \in \mathcal{D}. \tag{4}$$

Indeed, we have

$$\begin{aligned} \langle \mathcal{A}f, c\rangle_{\mathbb{C}^N} &= \sum_{k=1}^{N} c_k \int_{\mathcal{D}} \overline{f(x, y)}\, \mathrm{e}^{-2\pi \mathrm{i}(u_k x+v_k y)}\, \mathrm{d}x\, \mathrm{d}y \\ &= \int_{\mathcal{D}} \overline{f(x, y)} \sum_{k=1}^{N} c_k\, \mathrm{e}^{-2\pi \mathrm{i}(u_k x+v_k y)}\, \mathrm{d}x\, \mathrm{d}y = \langle f, \mathcal{A}^* c\rangle_{\mathcal{X}} \end{aligned}$$

where $\langle \cdot, \cdot \rangle_{\mathbb{C}^N}$ and $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ denote the inner products of $\mathbb{C}^N$ and $\mathcal{X}$, respectively.

Thus, the image reconstruction problem can be formulated as a linear inverse problem with discrete data [4]

$$\mathcal{A}f = g \tag{5}$$

which is ill-posed since it has infinite solutions.

2.1. Generalized solution

In this section we derive the formal expression of the minimum norm solution, or generalized solution, of problem (5), which can be reformulated in the following equivalent form: given the set of functions $\{\varphi_1, \ldots, \varphi_N\}$ in the Hilbert space $\mathcal{X}$ and the vector $g \in \mathbb{C}^N$, find $f \in \mathcal{X}$ such that

$$g_k = \langle f, \varphi_k\rangle_{\mathcal{X}}, \qquad k = 1, \ldots, N. \tag{6}$$

In this way, the kth component ($k = 1, \ldots, N$) of the element $\mathcal{A}f$ in the Euclidean space $\mathbb{C}^N$ corresponds to the value of the bounded linear functional described by

$$(\mathcal{A}f)_k = \langle f, \varphi_k\rangle_{\mathcal{X}}. \tag{7}$$

From (3) and (7) we can derive the following form for the function φ_k ($k = 1, \ldots, N$):

$$\varphi_k(x, y) = \mathrm{e}^{-2\pi \mathrm{i}(u_k x+v_k y)}, \qquad (x, y) \in \mathcal{D}. \tag{8}$$

The functions $\varphi_1, \ldots, \varphi_N$ defined in (8) lead to the generalized solution $f^\dagger$ of the inverse problem (i.e. the unique solution of the minimal norm of (6)) through the introduction of the Gram matrix $G = (G_{mn})$ whose entries are given by

$$G_{mn} = \langle \varphi_m, \varphi_n\rangle_{\mathcal{X}}, \qquad m, n = 1, \ldots, N. \tag{9}$$

[4] Coherently with the definition of visibility given in [24, 27], we define the Fourier transform with a positive sign in the exponent.

In fact, it is well known that the relation

$$\mathcal{A}\mathcal{A}^* = G$$

holds[5]; therefore, the eigendecomposition of the matrix G provides directly the singular vectors $a_1, \ldots, a_N \in \mathbb{C}^N$ and the singular values $\sigma_1 \geqslant \cdots \geqslant \sigma_N \in (0, \infty)$ of the operator $\mathcal{A}$. The singular system of $\mathcal{A}$ can be completed by straightforward calculation of the singular functions $\psi_1, \ldots, \psi_N \in \mathcal{X}$ through the formulae

$$\psi_k(x, y) = \frac{1}{\sigma_k} \sum_{j=1}^{N} (a_k)_j \varphi_j(x, y), \qquad k = 1, \ldots, N, \quad (x, y) \in \mathcal{D}. \tag{10}$$

From the singular system of $\mathcal{A}$ we can derive the generalized solution $f^\dagger$ of the inverse problem as

$$f^\dagger(x, y) = \sum_{k=1}^{N} \frac{\langle g, a_k \rangle_{\mathbb{C}^N}}{\sigma_k} \psi_k(x, y), \qquad (x, y) \in \mathcal{D} \tag{11}$$

(see also [3, 4]). In our case in which the functions $\varphi_1, \ldots, \varphi_N$ are defined by (8), the entry G_{mn} of the Gram matrix G $(m, n = 1, \ldots, N)$ is given by

$$G_{mn} = \frac{\left(\mathrm{e}^{2\pi \mathrm{i}(u_n - u_m)X_1} - \mathrm{e}^{2\pi \mathrm{i}(u_n - u_m)X_2}\right)\left(\mathrm{e}^{2\pi \mathrm{i}(v_n - v_m)Y_1} - \mathrm{e}^{2\pi \mathrm{i}(v_n - v_m)Y_2}\right)}{4\pi^2 (u_n - u_m)(v_n - v_m)}.$$

Once computed the eigendecomposition of G, formula (11) can be directly applied to obtain the function $f^\dagger(x, y)$. However, the generalized solution may not have a physical meaning (for example, in the image reconstruction problems we typically look for nonnegative solutions). Besides, the condition number of the operator $\mathcal{A}$, given by $\mathcal{K}(\mathcal{A}) = \sigma_1/\sigma_N$, can be very large, making the generalized solution sensitive to any perturbation of the data. Indeed, for the (u, v) distribution in figure 1 (which is realistic for the numerical experiments shown in section 4) and for symmetric squares $\mathcal{D}$ of different sizes $L = 64, 128, 256$ (i.e. $X_1 = Y_1 = -L/2$, $X_2 = Y_2 = L/2$), we have $\mathcal{K}(\mathcal{A}) = 1.2 \times 10^{18}, 6.8 \times 10^{13}, 2.4 \times 10^4$, respectively. For these reasons, we consider a further formulation of our problem, including suitable constraints on the solutions set and which can be practically solved, in its discrete version, with efficient and stable numerical methods.

2.2. *Convolution model*

In this section we show that the generalized solution $f^\dagger$ of the inverse problem (5) can be viewed as a solution of a deconvolution problem in the space domain.

Indeed, $f^\dagger$ is a minimum point of the least-squares functional

$$J(f) \equiv \tfrac{1}{2} \|\mathcal{A}f - g\|^2_{\mathbb{C}^N}, \qquad f \in \mathcal{X} \tag{12}$$

and, therefore, is a solution of the equation $\nabla J(f) = 0$, where

$$\nabla J(f) = \mathcal{A}^*\mathcal{A}f - \mathcal{A}^*g. \tag{13}$$

From the definition of $\mathcal{A}$ and $\mathcal{A}^*$ we can easily rewrite the operator $\mathcal{A}^*\mathcal{A}$ as a convolution mapping with an appropriate kernel $D \in L^2(\mathbb{R}^2, \mathbb{C})$ (also called *dirty beam* [22]). Indeed, for each $f \in \mathcal{X}$, from (3) and (4) we have

$$\mathcal{A}^*\mathcal{A}f = D * f, \tag{14}$$

[5] For the sake of correctness, this relation is mathematically true if we consider the left-hand side the matrix is naturally associated with the discrete linear operator $\mathcal{A}\mathcal{A}^*$.

where

$$D(x, y) = \sum_{k=1}^{N} \mathrm{e}^{-2\pi \mathrm{i}(u_k x + v_k y)}, \qquad (x, y) \in \mathbb{R}^2. \tag{15}$$

Moreover, the term $A^* g$ in (13) is the so-called *dirty image* f_d defined by

$$f_d(x, y) = \sum_{k=1}^{N} g_k \, \mathrm{e}^{-2\pi \mathrm{i}(u_k x + v_k y)}, \qquad (x, y) \in \mathcal{D}. \tag{16}$$

Thus, the problem of minimizing (12) can also be formulated as the deconvolution problem of finding a function $f \in \mathcal{X}$ such that

$$D * f = f_d. \tag{17}$$

Since it is natural in image restoration to require the radiation flux to be real and nonnegative, we are interested in solving the constrained minimization problem

$$\min_{\substack{f \in L^2(\mathcal{D}, \mathbb{R}) \\ f \geqslant 0}} \quad J(f). \tag{18}$$

Moreover, in some applications, it could be useful to insert also an equality constraint

$$\int_{\mathcal{D}} f(x, y) \, \mathrm{d}x \, \mathrm{d}y = F,$$

where F represents the total flux emitted by the source and inferred in some way from the data. We observe that problem (18) can be still ill-posed, since the set of nonnegative functions is closed and convex in L^2 but not compact.

2.3. Discrete formulation

In this section we want to introduce a discrete version of the problem previously described in a continuous setting. Therefore, we consider a uniform grid over the square $\mathcal{D}$ given by the points

$$x_j = X_1 + (j - 1)\Delta x, \qquad y_h = Y_1 + (h - 1)\Delta y, \qquad j, h = 1, \ldots, n, \tag{19}$$

and we discretize the integral of equation (1) by the rectangular rule obtaining a vector $g \in \mathbb{C}^N$ with

$$g_k \approx \sum_{j,h=1}^{n} f(x_j, y_h) \, \mathrm{e}^{2\pi \mathrm{i}(u_k x_j + v_k y_h)} \Delta x \Delta y, \qquad k = 1, \ldots, N.$$

Since the choice of $\mathcal{D}$ and n automatically determines the band of the image, the value of n must be chosen in such a way that the available (u_k, v_k) points ($k = 1, \ldots, N$) result to be included in the band itself.

In the discrete scenario, we look for an approximation of the values

$$f_{jh} = f(x_j, y_h)\Delta x \Delta y, \qquad j, h = 1, \ldots, n.$$

Therefore, given a vector $f_{jh} \in \mathbb{C}^{n^2}$, we define the linear operator A as

$$(Af)_k = \sum_{j,h=1}^{n} f_{jh} \, \mathrm{e}^{2\pi \mathrm{i}(u_k x_j + v_k y_h)}, \qquad k = 1, \ldots, N \tag{20}$$

and its adjoint is

$$(A^*c)_{jh} = \sum_{k=1}^{N} c_k \, \mathrm{e}^{-2\pi \mathrm{i}(u_k x_j + v_k y_h)}, \qquad j, h = 1, \ldots, n, \tag{21}$$

where $c \in \mathbb{C}^N$. Operators (20) and (21) can be considered a numerical approximation of (3) and (4); in these settings, the discrete inverse problems are given by

$$Af = g.$$

As a result of this, our optimal real positive radiation flux image will be the solution of the minimization problem in the space domain

$$\min_{\substack{f \in \mathbb{R}^{n^2} \\ f \geqslant 0}} \quad J(f) \equiv \tfrac{1}{2} \|Af - g\|^2_{\mathbb{C}^N}. \tag{22}$$

Following the process described in section 2.2, the gradient of the objective function J can be expressed again by means of a discrete convolution. With the same settings introduced before, we define the points

$$x_j = X_1 + (j-1)\Delta x, \qquad y_h = Y_1 + (h-1)\Delta y, \qquad j, h = -n, -(n-1), \ldots, 0$$

and denote by D_{jh} the values $D(x_j, y_h)\Delta x \Delta y$ $(j, h = -n, \ldots, n)$.

Thus, the discrete counterparts of (14) and (16) become

$$(D * f)(x_j, y_h) = \sum_{p,q=1}^{n} f_{pq} D_{j-p,h-q} = (A^*Af)_{jh} \tag{23}$$

and

$$f_d(x_j, y_h) = \sum_{k=1}^{N} g_k \, \mathrm{e}^{-2\pi \mathrm{i}(u_k x_j + v_k y_h)} = (A^*g)_{jh}, \tag{24}$$

where (x_j, y_h) is a grid point $(j, h = 1, \ldots, n)$.

As in the semi-discrete case, the flux conservation condition can be easily imposed by adding the equality

$$\sum_{j,h=1}^{n} f_{jh} = F$$

to the constraints of problem (22), where again F represents the total flux emitted by the source.

We conclude this section by pointing out two practical considerations:

- the computation of both the objective function and its gradient are performed entirely in the space domain through equations (23) and (24). In particular, once the grid (19) is fixed, we do not need to explicitly calculate the associated sampling of the frequency plane given by the discrete Fourier transform;
- we can assume that the (u, v) points corresponding to the measured visibilities are symmetric with respect to the origin in the frequency plane. This requirement can be easily accomplished by adding the further visibilities

  $$g_{N+k} = V(-u_k, -v_k) = \overline{g_k}, \qquad k = 1, \ldots, N.$$

  As a consequence of this, the dirty image and the dirty beam are automatically real-valued matrices.

3. Description of the algorithm

For the numerical solution of the minimization problem (22) we propose to employ a gradient projection algorithm [5, 6, 8, 33], which is well suited in the case of simple constraints. In particular, we adopt the variable steplength approach [6, 8, 33], which applies to any minimization problem

$$\min_{f\in\mathcal{C},} \quad J(f) \tag{25}$$

where J is a differentiable real-valued function and $\mathcal{C}$ is a convex closed subset of $\mathbb{R}^m$, $m \in \mathbb{Z}_+$. Clearly, problem (22) is a special case of (25), since, even if the data are complex numbers, the objective function is real valued and its argument is constrained in a subset of $\mathbb{R}^{n^2}$.

The method starts from a feasible point $f^{(0)} \in \mathcal{C}$ and generates a sequence

$$f^{(k+1)} = f^{(k)} + \lambda_k d^{(k)}, \qquad k = 0, 1, 2, \ldots,$$

where each $f^{(k)} \in \mathcal{C}$ and the search direction is computed as

$$d^{(k)} = \mathcal{P}_{\mathcal{C}}(f^{(k)} - \alpha_k \nabla J(f^{(k)})) - f^{(k)}, \qquad k = 0, 1, 2, \ldots.$$

Here, λ_k, α_k are positive scalar parameters and $\mathcal{P}_{\mathcal{C}}$ denotes the orthogonal projection on the set $\mathcal{C}$. If α_k is chosen in a bounded interval $[\alpha_{\min}, \alpha_{\max}]$, with $0 < \alpha_{\min} < \alpha_{\max}$, and λ_k satisfies an Armijo-type condition, then the limit points of the sequence $\{f^{(k)}\}$ are stationary for problem (25) [6, 8]. An overview of the gradient projection method is given in algorithm 1.

Algorithm 1. Gradient projection (GP) method

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

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

STEP 1. Choose the parameter $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$.
STEP 2. Projection: $y^{(k)} = P_{\mathcal{C}}(f^{(k)} - \alpha_k \nabla J(f^{(k)}))$.
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 $J_{\text{new}} \leqslant J(f^{(k)}) + \beta\lambda_k \nabla J(f^{(k)})^T d^{(k)}$ THEN
    go to Step 6;
  ELSE
    set $\lambda_k = \theta\lambda_k$ and go to Step 5.
  ENDIF
STEP 6. Set $f^{(k+1)} = f^{(k)} + \lambda_k d^{(k)}$.

END

In spite of its simplicity, the gradient projection method offers good practical performances on large-scale problems compared to other optimization algorithms, especially when the Hessian matrix of the objective function is not explicitly available.

However, the effectiveness of the algorithm strongly depends on its implementation. In particular, in section 3.1, we focus on two aspects, the choice of the steplength parameter α and the computation of the objective function in (22) and its gradient.

Besides, we have to take into account the ill-conditioning of problem (22) and the noise that may affect the data: to this end, some regularization techniques are needed. Even if a rigorous theoretical study is not yet available, the numerical experience on image reconstruction from noisy data demonstrates that the GP method exhibits the semi-convergence property [3, 14] (see section 4.2): the results of the iteration first provide better and better approximations of the true object but after a certain point they turn to the worse, due to increased noise propagation. Therefore, regularization is obtained in practice by an appropriate early stopping of the iterations: in this way, the gradient projection method is employed as an iterative regularization method. The choice of the stopping criterion is discussed in section 3.2. A general description of our approach to image reconstruction from sampling of the Fourier transform, that we called *Space-D*, is given in algorithm 2.

3.1. Implementation details

A crucial feature for the effectiveness of the method is the choice of the steplength parameter α_k. Recent studies show that significant improvements in the convergence speed of gradient methods can be obtained with a suitable alternation of the two Barzilai and Borwein (BB) rules [1], which are defined as follows:

$$\alpha_k^{(1)} = \frac{s^{(k-1)^T} s^{(k-1)}}{s^{(k-1)^T} z^{(k-1)}} \qquad \text{and} \qquad \alpha_k^{(2)} = \frac{s^{(k-1)^T} z^{(k-1)}}{z^{(k-1)^T} z^{(k-1)}}, \tag{26}$$

where $s^{(k-1)} = (f^{(k)} - f^{(k-1)})$ and $z^{(k-1)} = (\nabla J(f^{(k)}) - \nabla J(f^{(k-1)}))$. The BB rules are motivated by the quasi-Newton approach, where the inverse of the Hessian is replaced by a multiple of the identity matrix $B(\alpha) = \alpha I$. Then, omitting the iteration number, the two BB formulae are given by $\alpha^{(1)} = \arg\min \|B(\alpha)s - y\|$ and $\alpha^{(2)} = \arg\min \|s - B(\alpha)^{-1} y\|$.

The recent literature on the steplength selection in gradient methods suggests to design steplength-updating strategies by alternating the two BB rules. Here we will use the adaptive alternation strategy proposed in [8, 15] that gave remarkable convergence rate improvements in many different applications. Given an initial value α_0, the steplengths α_k ($k = 1, 2, \ldots$) are defined by the following criterion:

if $\alpha_k^{(2)} / \alpha_k^{(1)} \leqslant \tau_k$ THEN

$\alpha_k = \min\left\{\alpha_j^{(2)},\ j = \max\{1, k - M_\alpha\}, \ldots, k\right\}$; $\qquad \tau_{k+1} = \tau_k * 0.9$;

ELSE

$\alpha_k = \alpha_k^{(1)}$; $\qquad \tau_{k+1} = \tau_k * 1.1$;

ENDIF,

where M_α is a prefixed nonnegative integer and $\tau_1 \in (0, 1)$.

As observed in several experimental studies [2, 8], the GP method equipped with this alternation of the two BB formulae is more efficient with respect to the same algorithm with other steplength selection rules and also to other iterative approaches as the projected Landweber method.

About the complexity of the algorithm, the main tasks are the computation of the objective function $J(f)$ at every backtracking loop and of the gradient $\nabla J(f)$ at each main iteration, which are both depending on the quantities $A^*Af^{(k)}$ and A^*g. We remark that A^*g that is the dirty image f_d needs to be computed only once at the beginning by formula (24). Moreover, thanks to the equivalence (23), the matrix A^*A has a block Toeplitz structure and, therefore, the matrix–vector product $A^*Af^{(k)}$ can be efficiently computed with the FFT algorithm (see [36]). Thus, taking into account of the zero boundary conditions on the reconstructed image, the cost per iteration is $\mathcal{O}(4n^2 \log(4n^2))$.

3.2. Stopping criterion and regularization

The stopping criterion for the iterative regularization method in the Space-D algorithm is based on the Morozov's discrepancy principle [14, 35]. In particular, we assume that the measured data g^δ can be represented in the form $g^\delta = g + \delta g$, where δg denotes the noise affecting and $g = Af$, where f is the object to be reconstructed. Furthermore, we assume that the quantity $\eta = \|\delta g\|$ (or an estimate of it, as in the case of the application described in the following section) is known. Then, a regularized solution $f^{(k)}$ is computed if we terminate the optimization procedure at the first iteration $\bar{k}$ in which the inequality

$$\|Af^{(k)} - g^\delta\| \leqslant \eta, \qquad k = 0, 1, 2, \ldots \tag{27}$$

is satisfied.

As far as we know, no result exists proving that this criterion, applied to the GP iteration, leads to a regularization method for the constrained least-squares problem (22). However, in [2] it has been shown the practical effectiveness of the GP method equipped with criterion (27) on image deconvolution problems.

Moreover, we include a further stopping condition based on the relative difference of the objective function between two successive iterates

$$|J(f^{(k)}) - J(f^{(k-1)})| < \epsilon |J(f^{(k)})|, \tag{28}$$

where ϵ is a prefixed tolerance. The criterion (28) is a quite standard condition to check the convergence of an iterative optimization algorithm. Its main purpose is to devise the point where no significant decrease in the objective function is obtained (see for example [37]). We include this condition in algorithm 2 since, from the numerical experience, the discrepancy criterion alone may terminate the optimization procedure too soon [2], especially when the error norm η is overestimated. However, due to the ill-posedness of the problem, too small values of ϵ must be avoided to guarantee a sufficient amount of regularization.

4. Numerical experiments: the RHESSI mission

In this section we test our algorithm in a real-world application arising in x-ray astronomy. The specific problem we are interested in is to reconstruct the spatial distribution of the x-ray

Algorithm 2. Space-D algorithm

STEP 1. *Back-projection*

Compute the dirty beam and the dirty image at the given grid points

$$D(x_j, y_h) \qquad j, h = -n, \ldots, n$$
$$f_d(x_j, y_h) \qquad j, h = 1, \ldots, n$$

using formulae (15), (24).

STEP 2. *Deconvolution*

Solve the problem

$$\min_{\substack{f \in \mathbb{R}^{n^2} \\ f \geqslant 0,}} \quad J(f)$$

where $J(f)$ is defined in (22) with the GP algorithm, stopping the procedure at the first iteration in which both the discrepancy principle (27) and the inequality (28) are satisfied.

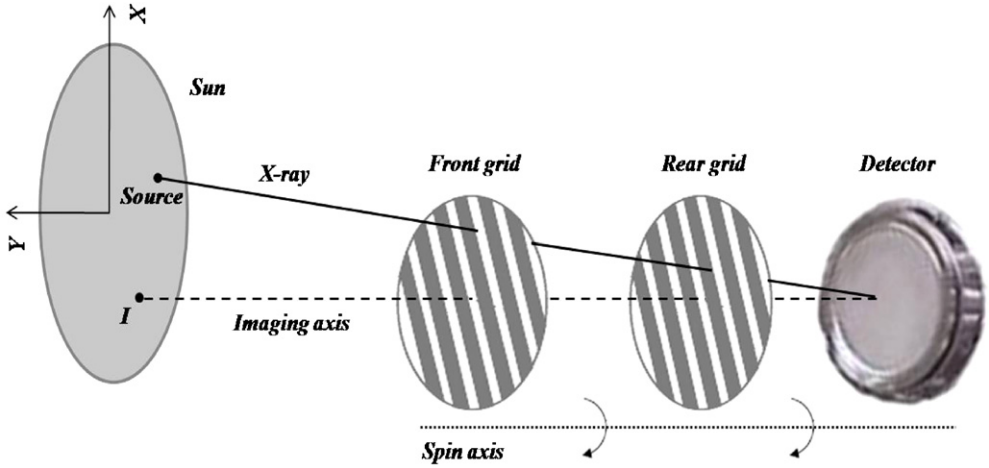


Figure 2. Schematic representation of a rotating modulation collimator.

radiation emitted by accelerated particles in the solar chromosphere during the solar flares. The datasets are provided by the Reuven Ramaty high energy solar spectroscopic imager (RHESSI) satellite under the form of visibilities, which are calibrated measurements of the source distribution's two-dimensional spatial Fourier components. We compared our method with the two state-of-the-art reconstruction algorithms from visibilities already available in the RHESSI software by using both simulated data and real events.

4.1. The RHESSI spacecraft

The NASA RHESSI mission has been launched with the precise intent to investigate the particle acceleration mechanisms during the solar flares through the spatial and spectral analysis of the emitted x-ray radiation [21, 23]; RHESSI is currently operating and its datasets are at disposal of the scientific community.

RHESSI is the last descendant of a generation of collimator-based satellites (HINOTORI, YOHKOH/HXT, HEIDI) originated in 1980 from the hard x-ray imaging spectrometer (HXIS), carried in the Solar Maximum Mission [19, 20, 30, 34]. It observes hard x-ray emission from the entire solar disc through a set of nine co-aligned pairs of rotating modulation collimators (RMCs—see figure 2); the transmitted radiation is recorded on a set of cooled HPGe detectors, which provide spectral resolution of $\sim$1 keV over a wide spectral range. Because of the RMC design of the instrument, imaging information is recorded in its most 'native' form, as a set of visibilities, measured at spatial frequencies (u, v) corresponding to angular resolutions of each RMC pair. Therefore, the visibilities measured by RHESSI are arranged around concentric circles in the frequency plane (see figure 1). A detailed description of the relation between collimator and Fourier transform is presented in [19].

The reconstruction algorithms from visibilities available within the RHESSI mission are the following ones.

- *uv-smooth* [24]. The key point of this method is the coverage of the frequency plane through an interpolation step applied on the measured visibilities. The final image is obtained by applying the Gerchberg–Papoulis method to the interpolated visibilities, in which a projection onto the convex set of all nonnegative functions is performed at each

iteration. The very fast interpolation routine and the low number of iterations required make uv-smooth the fastest image reconstruction algorithm available within the RHESSI mission.

- *MEM-NJIT* [9]. This method has been implemented by the solar group at New Jersey Institute of Technology and is the first visibility-based method created for RHESSI data analysis. The idea at the basis of MEM-NJIT is the maximization of an entropy functional with constraints on the chi-square and the total flux of the image. A gridding initial step is performed on the measured visibilities in order to make possible the use of the FFT algorithm in the iterative optimization procedure adopted to solve the maximization problem.
- *forward-fit*. The best-fit values of the parameters corresponding to assumed simple functional forms for the source are determined through a fitting procedure. The actual routine only provides for a limited choice of shapes (one or two circular Gaussians, one elliptical Gaussian, one curved elliptical Gaussian).

The corresponding routines written in interactive data language (IDL) are available within the Solar SoftWare (SSW) tree (*uv_smooth.pro*, *mem_njit.pro* and *hsi_vis_fwdfit.pro*, respectively).

Since we are not interested in considering pre-selected morphologies but want to introduce a general method able to reconstruct any kind of source geometry, in the following we will compare the performances of Space-D only with the first two algorithms.

4.2. Simulated data

Following the idea introduced in [24], instead of using sources described by mathematical forms, we created a set of simulated images starting from real solar flare maps. To this aim, we considered one of the most famous events since RHESSI is in orbit, which occurred on 23 July 2002 [13, 18, 26]. We selected a time range of about 1 min during the phase of maximum emission (00:29:10–00:30:19 UT) and, starting straightly from the counts collected by RHESSI, we used the Clean algorithm [19] to build the 64×64 (with 1 arcsec $\times$ 1 arcsec pixel size) x-ray images corresponding to three different energy ranges: 20–22 keV, 41–46 keV and 156–177 keV. We remark that with these settings the band of the image is given by the square $[-0.5, 0.5] \times [-0.5, 0.5]$, which includes all the spatial frequencies (u_k, v_k) $(k = 1, \ldots, N)$ provided by RHESSI (see figure 1).

The selected images represent very different source morphologies (see figure 3—first column):

- in the low energy case, a single wide asymmetric x-ray source is present;
- for intermediate energies, a more complicated structure combining the superposition of several sources is involved;
- at higher energies, the emission is limited to two compact sources plus a weaker one in-between.

A threshold-based filter was applied to the Clean images in order to eliminate artefacts potentially introduced by the Clean algorithm. The resulting maps have been considered as the target distributions for the reconstruction algorithms; the corresponding visibilities have been calculated through numerical integration of the Fourier transform and corrupted by realistic noise with SSWs routines *hsi_vis_map2vis.pro* and *hsi_vis_randomize.pro*.

We remark that the Clean algorithm is a deconvolution procedure which builds an image from the count profiles provided by RHESSI. Since the starting point of this algorithm is not a sampling of the Fourier transform of the image, we did not consider it for the comparison with Space-D but we simply used it as a sort of 'fair referee' to build our realistic simulated

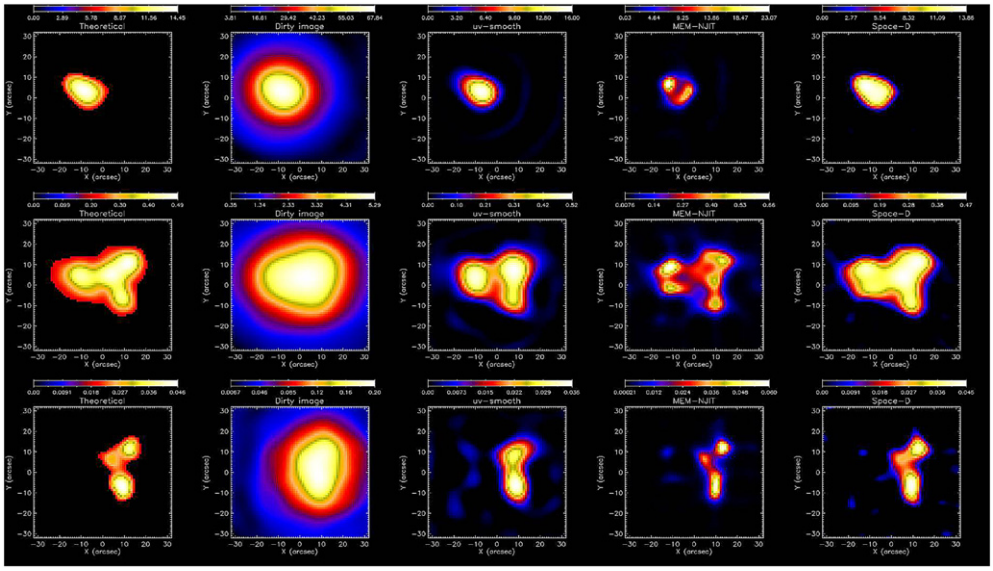


Figure 3. Results of the simulated tests for the three energy ranges considered: 20–22 keV (top row), 41–46 keV (middle row) and 156–177 keV (bottom row). From *left* to *right*, the theoretical image, the dirty image and the reconstructions with uv-smooth, MEM-NJIT and Space-D are presented, respectively.

maps. We refer the reader to [11, 31] for a comparison of visibility and non-visibility-based image reconstruction algorithms from RHESSI data.

In figure 3 we collect (from left to right, respectively) the original sources, the corresponding dirty images and the reconstructed maps for the three simulated cases obtained with uv-smooth, MEM-NJIT and Space-D. The considered algorithms ran on a PC equipped with a 1.66 GHz Intel Core Duo T5500 in IDL environment with their default settings: in particular, for Space-D we chose $\epsilon = 10^{-4}$ in the stopping rule (28) and a constant image with total flux equal to $\max_k |g_k|$ as the starting point $f^{(0)}$.

The aim of our experimentation is to compare the accuracy of the approximate solutions: we just point out that the overall computation was performed in few seconds in all cases (1–2 s for uv-smooth and Space-D and 4–5 s for MEM-NJIT), which is negligible with respect to the time needed for loading the data.

For all the reconstructions we also report in table 1 both the Euclidean relative errors, i.e. the quantity $\|\bar{f} - f\|_2/\|f\|_2$ where f is the true object and $\bar{f}$ is the approximation provided by the algorithms, and the ratios between reconstructed and original total flux calculated by simply summing up the pixel contents of the corresponding images.

From figure 3 and table 1 we can observe that, in general, the more accurate reconstructions are obtained by the Space-D algorithm. The uv-smooth method performs very well on smooth and geometrically simple sources (simulation 1). Some difficulty might be encountered in recovering more structured diffuse objects (simulation 2) or multiple close compact sources (simulation 3), possibly due to the smoothness forced in the interpolated data. On the other side, MEM-NJIT seems to overfit the data, as also remarked in [11, 12, 24]: diffuse objects are reconstructed as a combination of sharper and narrower sources. However, it is capable to devise the weaker source in the third simulation. From our simulations, Space-D reveals

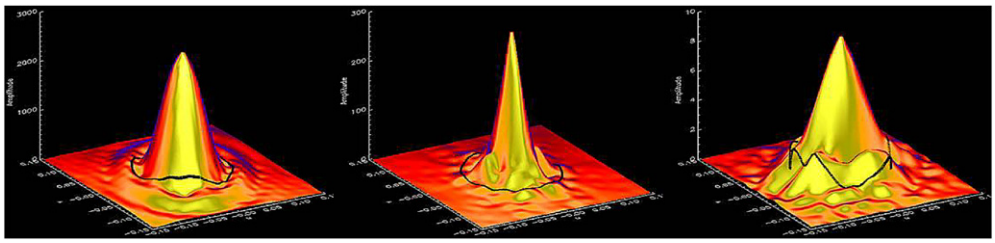


Figure 4. Fourier transform's amplitude of the Space-D reconstructed image in the three energy ranges considered: 20–22 keV (left), 41–46 keV (middle) and 156–177 keV (right). The black strip correspond to the boundary of the band containing the measured data.

Table 1. Relative reconstruction errors and ratio between reconstructed and theoretical total fluxes for the three simulated datasets.

		uv-smooth	MEM-NJIT	Space-D
20–22 keV	Error	0.237 681	0.422 465	0.145 208
	Flux ratio	1.061 657	0.981 008	1.002 756
41–46 keV	Error	0.260 160	0.360 863	0.163 929
	Flux ratio	0.978 989	0.993 390	1.005 529
156–177 keV	Error	0.382 557	0.425 764	0.294 643
	Flux ratio	1.105 895	1.063 128	1.039 846

a good capability both to recover simple geometries and to detect some details present in the original image, thanks to the out-of-band extrapolation achieved through the projection at each iteration of the image on the set of constraints $\mathcal{C}$ [25] (see figure 4). We have to admit that extremely complicated morphologies as the one in simulation 2 represent a hard task also for our methodology, even if on the whole the general structure is well reconstructed.

Finally, we conclude this section with some comments about the stopping criterium adopted and the related parameter ϵ. To this aim, for each simulated case we report in figure 5 the Euclidean relative error as a function of the iteration number. The optimal iteration number suggested by the discrepancy principle is highlighted with a square, while a triangle (resp. a diamond, an asterisk and a plus sign) denotes the optimal number of iterations provided by the stopping criterium (28) with $\epsilon = 10^{-2}$ (resp. 10^{-3}, 10^{-4}, 10^{-5}). In our tests, the optimal number of iterations provided by the algorithm is always the one highlighted with the asterisk, i.e. 77 iterations for simulation 1, 57 for simulation 2 and 76 for simulation 3.

From figure 5 we can clearly see the following.

- The semi-convergence property of the iterative procedure which shows up in the presence of a 'best' reconstruction with minimum relative error.
- The good stability of the reconstructions with respect to the parameter ϵ. Also in simulation 3 (figure 5(c)), where the effect of noise is more evident, any value of ϵ in the range $[10^{-4}, 10^{-3}]$ provides essentially the same reconstruction error. As declared before, for our images we arbitrarily chose $\epsilon = 10^{-4}$.
- The effect of the stopping criteria (27) and (28). Especially in simulations 1 and 2 (figures 5(a) and (b)), the number of iterations suggested by the discrepancy is particularly low, possibly due to an overestimation of the errors on the input visibilities. Thus, the

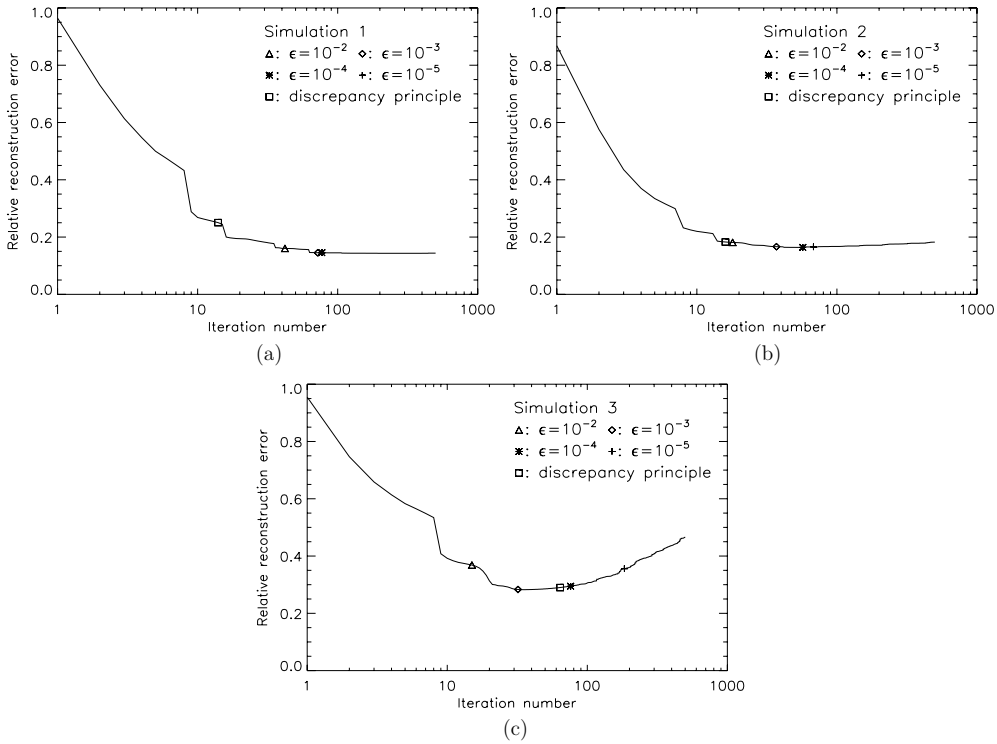


Figure 5. Euclidean relative errors as functions of the iteration number for the Space-D algorithm in the three energy ranges considered: (a) 20–22 keV; (b) 41–46 keV; (c) 156–177 keV. The squares correspond to the iteration numbers provided by the discrepancy principle, while the triangles (resp. diamonds, asterisks, plus signs) denote the numbers of iterations provided by the stopping criterium (28) with $\epsilon = 10^{-2}$ (resp. 10^{-3}, 10^{-4}, 10^{-5}).

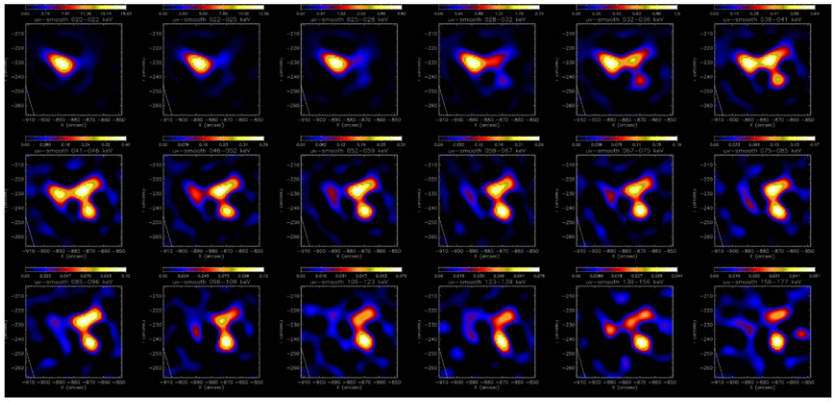
addition of the stopping criterium (28) with an intermediate value for ϵ may help the algorithm in going further with the iterations.

4.3. Real data

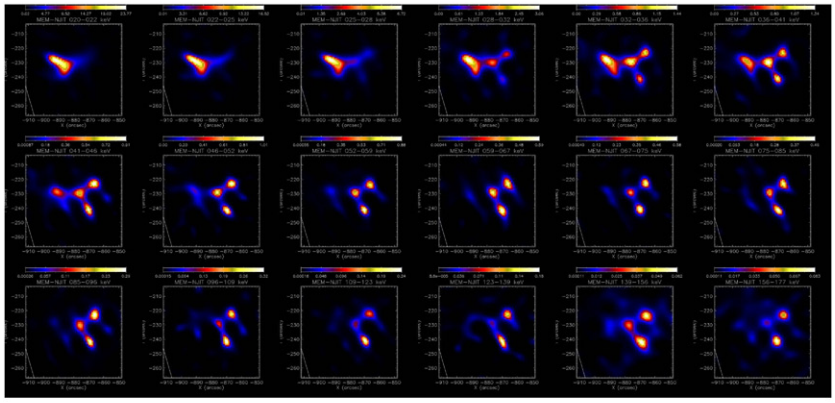
After the analysis of synthetic datasets, we complete the numerical experiments by testing the performance of Space-D on real flare data provided by RHESSI. The three algorithms ran with the same settings as in the previous simulations. Despite (of course) the true distributions are unknown, we show again the images obtained with the other two methods considered before, in order to compare the reconstructions and make some comments.

Given the importance of this event, we consider again the solar eruption of 23 July 2002 used to build the simulated images in the previous section. Following [13], we selected 18 energy intervals with increasing width in order to avoid that poor statistics would reflect into extremely noisy visibilities. The reconstructions provided by uv-smooth, MEM-NJIT and Space-D are given in figures 6(a)–(c), respectively, in lexicographic order for increasing energies.

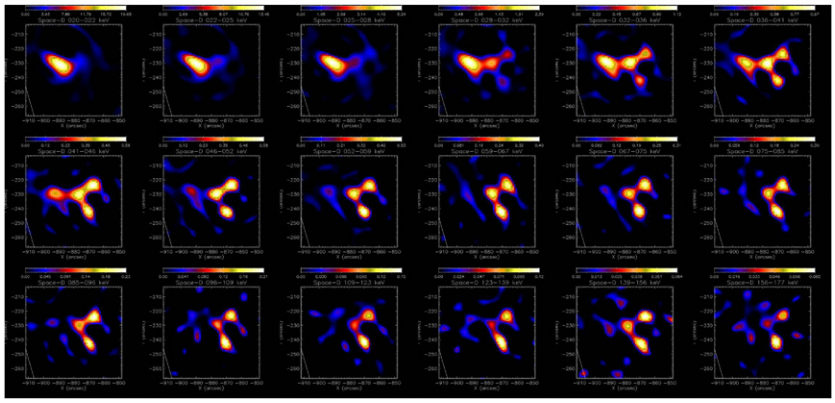
Each algorithm presents a behaviour similar to the one observed in the simulations with respect to the different source morphologies. At low energies, a single extended emitting region is revealed by all the three methods, with a narrower width in the MEM-NJIT case.



(a)



(b)



(c)

Figure 6. Real data: (a) uv-smooth; (b) MEM-NJIT; (c) Space-D.

At intermediate and high energies, the source structure increases in complexity splitting into several compact sources. The main discrepancy among the images provided by the different algorithms resides in the northern area: while both Space-D and MEM-NJIT clearly devise two separated emitting regions, for uv-smooth they collapse in a single source.

5. Conclusions and future work

In this paper we presented a new deconvolution approach to recover an image from a sparse and nonuniform sampling of its Fourier transform. The strength of our method is that, unlike many classical and recent approaches, the reconstruction algorithm is applied straightly on the measured data, avoiding that arbitrary assumptions may affect the results.

Applications of this method are possible in image and signal restoration problems arising from several scientific and medical areas. We showed the potential of our algorithm when applied to analyse real and synthetic data from the NASA RHESSI mission. From the numerical experience, the proposed methodology seems to provide very accurate reconstructions in very few seconds, thanks to the well-tried optimization algorithm adopted for the deconvolution.

Future studies will be addressed in both theoretical and practical issues. We could consider further formulations of the problem including more *a priori* information as, for example, a suitable weighting of the measured visibilities. Moreover, it will be interesting to analyse the behaviour of the algorithm when the regularization is performed by including an explicit penalty term in the objective function (e.g. Tikhonov, entropy, total variation).

From the practical point of view, further comparisons on several real datasets with all the algorithms already available for RHESSI imaging (visibility and non-visibility based) are mandatory. Moreover, if combined with an existing software which infers the Fourier transform of the emitting source (instead of the emission flux) [28, 29], images of the emitting particles could be reproduced.

Acknowledgments

We are very grateful to the anonymous referees for their very useful comments and suggestions. This work has been partially supported by a grant of the Italian GNCS-INdAM, by Italian MIUR and by a grant of the Swiss International Space Science Institute (ISSI).

References

[1] Barzilai J and Borwein J M 1988 Two point step size gradient methods *IMA J. Numer. Anal.* **8** 141–8
[2] Benvenuto F, Zanella R, Zanni L and Bertero M 2010 Nonnegative least-squares image deblurring: improved gradient projection approaches *Inverse Problems* **26** 025004
[3] Bertero M and Boccacci P 1998 *Introduction to Inverse Problems in Imaging* (Bristol: Institute of Physics Publishing)
[4] Bertero M, De Mol C and Pike E R 1985 Linear inverse problems with discrete data: I. General formulation and singular system analysis *Inverse Problems* **1** 301–30
[5] Bertsekas D 1999 *Nonlinear Programming* (Belmont: Athena Scientific)
[6] Birgin E G, Martinez J M and Raydan M 2003 Inexact spectral projected gradient methods on convex sets *IMA J. Numer. Anal.* **23** 539–59
[7] Blahut R E 2001 *Theory of Remote Image Formation* (Cambridge: Cambridge University Press)
[8] Bonettini S, Zanella R and Zanni L 2009 A scaled gradient projection method for constrained image deblurring *Inverse Problems* **25** 015002
[9] 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
[10] Bracewell R N 2000 *The Fourier Transform and Its Applications* (New York: McGraw-Hill)
[11] 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
[12] Donoho D L, Johnstone I M, Hoch J C and Stern A S 1992 Maximum entropy and the nearly black object *J. R. Stat. Soc.* B **54** 41–81

[13] Emslie A G, Kontar E P, Krucker S and Lin R P 2003 RHESSI hard x-ray imaging spectroscopy of the large gamma-ray flare of 2002 July 23 *Astrophys. J.* **595** L107–10
[14] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)
[15] Frassoldati G, Zanghirati G and Zanni L 2008 New adaptive stepsize selections in gradient methods *J. Ind. Manag. Optim.* **4** 299–312
[16] Gull S F and Daniell G J 1978 Image reconstruction from incomplete and noisy data *Nature* **272** 686–90
[17] Hadamard J 1923 *Lectures on Cauchy's Problem in Linear Partial Differential Equations* (New Haven, CT: Yale University Press)
[18] Holman G D, Sui L, Schwartz R A and Emslie A G 2003 Electron bremsstrahlung hard X-ray spectra, electron distributions, and energetics in the 2002 July 23 solar flare *Astrophys. J.* **595** L97–101
[19] Hurford G J *et al* 2002 The RHESSI imaging concept *Solar Phys.* **210** 61–86
[20] Kosugi T *et al* 1991 The hard x-ray telescope (HXT) for the solar-A mission *Solar Phys.* **136** 17–36
[21] Krucker S *et al* 2008 Hard x-ray emission from the solar corona *Astron. Astrophys. Rev.* **16** 155–208
[22] Lannes A, Anterrieu E and Maréchal P 1997 CLEAN and WIPE *Astron. Astrophys. Suppl. Ser.* **123** 183–98
[23] Lin R P *et al* 2002 The Reuven Ramaty High-Energy Solar Spectroscopic Imager (RHESSI) *Solar Phys.* **210** 3–32
[24] 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
[25] Piana M and Bertero M 1997 Projected Landweber method and preconditioning *Inverse Problems* **13** 441–64
[26] Piana M, Massone A M, Kontar E P, Emslie A G, Brown J C and Schwartz R A 2003 Regularized electron flux spectra in the July 23, 2002 solar flare *Astrophys. J.* **595** L127–30
[27] Piana M, Massone A M, Hurford G J, Prato M, Emslie A G, Kontar E P and Schwartz R A 2007 Electron flux spectral imaging of solar flares through regularized analysis of hard x-ray source visibilities *Astrophys. J.* **665** 846–55
[28] 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
[29] Prato M, Piana M, Emslie A G, Hurford G H, Kontar E P and Massone A M 2009 A regularized visibility-based approach to astronomical imaging spectroscopy *SIAM J. Imag. Sci.* **2** 910–30
[30] Prince T A, Hurford G J, Hudson H S and Crannell C J 1988 Gamma-ray and hard x-ray imaging of solar flares *Solar Phys.* **118** 269–90
[31] Schmahl E J, Pernak R L, Hurford G J, Lee J and Bong S 2007 Analysis of RHESSI flares using a radio astronomical technique *Solar Phys.* **240** 241–52
[32] Schomberg H and Timmer J 1995 The gridding method for image reconstruction by Fourier transformation *IEEE Trans. Med. Imaging* **14** 596–607
[33] Serafini T, Zanghirati G and Zanni L 2005 Gradient projection methods for quadratic programs and applications in training support vector machines *Optim. Methods Softw.* **20** 353–78
[34] Strong K T, Saba J L R, Haisch B M and Schmelz J T 1999 *The Many Faces of the Sun: A Summary of the Results from NASA's Solar Maximum Mission* (New York: Springer)
[35] Tikhonov A N, Goncharsky A V, Stepanov V V and Yagola A G 1995 *Numerical Methods for the Solution of Ill-Posed Problems* (Dordrecht: Kluwer)
[36] Vogel C R 2002 *Computational Methods for Inverse Problems* (Philadelphia, PA: SIAM)
[37] 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