ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Reconstruction of 3D X-Rays CT images from reduced samplings by a scaled gradient projection algorithm

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Loli Piccolomini E, C.V.L. (2018). Reconstruction of 3D X-Rays CT images from reduced samplings by a scaled gradient projection algorithm. COMPUTATIONAL OPTIMIZATION AND APPLICATIONS, 71(1), 171-191 [10.1007/s10589-017-9961-2].

Availability:
This version is available at: https://hdl.handle.net/11585/633541 since: 2018-04-04

Published:

DOI: http://doi.org/10.1007/s10589-017-9961-2

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

16 September 2026

This is the final peer-reviewed accepted manuscript of:

Piccolomini, E.L., Coli, V.L., Morotti, E. *et al.* Reconstruction of 3D X-ray CT images from reduced sampling by a scaled gradient projection algorithm. *Comput Optim Appl* 71, 171–191 (2018).

The final published version is available online at: **https://doi.org/10.1007/s10589-017-9961-2**

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/)

When citing, please refer to the published version.

Reconstruction of 3D X-ray CT images from reduced sampling by a scaled gradient projection algorithm

E. Loli Piccolomini*, V.L. Coli†, E. Morotti‡ and L. Zanni†

October 10, 2017

Abstract

We propose a scaled gradient projection algorithm for the reconstruction of 3D X-ray tomographic images from limited data. The problem arises from the discretization of an ill-posed integral problem and, due to the incompleteness of the data, has infinite possible solutions. Hence, by following a regularization approach, we formulate the reconstruction problem as the nonnegatively constrained minimization of an objective function given by the sum of a fit-to-data term and a smoothed differentiable Total Variation function. The problem is challenging for its very large size and because a good reconstruction is required in a very short time. For these reasons, we propose to use a gradient projection method, accelerated by exploiting a scaling strategy for defining gradient-based descent directions and generalized Barzilai-Borwein rules for the choice of the step-lengths. The numerical results on a 3D phantom are very promising since they show the ability of the scaling strategy to accelerate the convergence in the first iterations.

Keywords. 3D Computed Tomography, image reconstruction, Total Variation regularization, nonnegatively constrained minimization, scaled gradient projection methods.

1 Introduction

In this paper we consider a very challenging optimization problem arising in 3D X-ray Computed Tomography (CT) image reconstruction from low sampling acquisitions, i.e., when the CT system acquires only a reduced

*Department of Mathematics, University of Bologna, e-mail: elena.loli@unibo.it

†Department of Physics, Informatics and Mathematics, University of Modena and Reggio Emilia, e-mail: vannalisa.coli@unimore.it, luca.zanni@unimore.it

‡Department of Mathematics, University of Padova, e-mail: morotti@math.unipd.it

set of data. This application has recently received growing attention in the medical community, since sub-sampling acquisitions have several advantages over the traditional complete sampling acquisitions in speeding up the imaging process, thus reducing the exposure to ionizing radiations and increasing the patient safety [31, 11, 20].

In these cases the traditional analytical reconstruction methods such as the Filtered Back-Projection (called Feldkamp method in 3D [17]) produce images of low quality, with extreme artifacts and high noise. Iterative Image Reconstruction (IIR) methods are generally preferred because they can introduce a priori information about the unknown object and they can exploit the Compressed Sensing (CS) theory for reconstructing a signal or an image from a reduced number of acquisitions with respect to the Nyquist theory [20]. The drawback of IIR algorithms is their higher computational cost with respect to the analytical methods, but thanks to the dramatic improvement of CPUs speed and the possibility to perform parallel computation at low cost on GPUs, the time for the IIR algorithms execution is now acceptable even in the clinical setting [3].

The IIR methods solve a minimization problem of the form:

$$\min_{x \geq 0} f(x) = J(x) + \lambda R(x) \tag{1}$$

where $J(x)$ is the fit-to-data function whose expression is related to the kind of noise on the data, $R(x)$ is a regularization function and $\lambda > 0$ is the regularization parameter.

Since tomographic data are affected by mixed Gaussian and Poisson noise, it is desirable to be able to efficiently solve problem (1) with $J(x)$ equal to the linear Least Squares (LS) functional, when Gaussian noise is dominant, or equal to the nonlinear Kullback-Leibler (KL) divergence, when Poisson noise is dominant. We propose a unifying optimization framework which allows to choose the proper form of $J(x)$, depending on the dominant kind of noise of a specific physical CT system. The function $R(x)$ should reduce the noise, regularize the ill-conditioned problem arising from the discretization of an ill-posed Fredholm integral equation and impose some sparsity on the computed solution following the CS theory. Since many medical images are almost uniform inside the organs, the most widely used regularization function for the CT problems is the Total Variation (TV) function [36, 38, 35, 34, 37, 40, 24, 32]

$$TV(x) = \int_{\Omega} |\nabla x| dx \tag{2}$$

that forces the sparsity in the gradient domain of the solution.

In real applications, the problem has a very large size, of the order of billions, and the problem solution is very challenging, because in the clinical applications a good image must be reconstructed in at most 1-2 minutes.

For these reasons, the IIR methods are not executed until convergence, but they are stopped after the desired time. With these premises, it is essential to use a reconstruction algorithm with non-expensive iterations and fast convergence in the very first iterations. To this end, some suggestions are available in the recent literature: first-order optimization methods based on accelerated gradient schemes have been proposed in [37, 21, 34], optimization transfer methods have been applied to CT image reconstruction for example in [25, 15, 23], a momentum approach can be found in [24] (based on Nesterov method in [29]) and a Fixed Point method using approximate second-order information has been used in [27].

The aim of this paper is to propose, for the solution of (1), a first-order solver belonging to the class of the Scaled Gradient Projection (SGP) methods [6, 10], exploiting special diagonal scaling matrices for defining the descent directions and selection strategies derived by the Barzilai-Borwein (BB) rules for the choice of the step-length parameter [1]. SGP algorithms have been successfully exploited in image restoration applications [7, 10], where the problem formulation is very similar to (1), while the problem size is much smaller than in 3D CT imaging and the matrix-vector products are performed by Fast Fourier Transforms thanks to the particular structure of the matrix involved in the objective function. Hence, the efficiency of an SGP approach in the considered CT application is not predictable and its ability in satisfying the practical request to provide accurate reconstructions in a short time needs to be investigated. In our study, we evaluate the behavior of the proposed SGP algorithm in comparison with a standard non-scaled BB method and the state-of-the-art accelerated gradient method [21], which has been successfully applied in tomographic reconstruction problems.

The paper is organized as follows. In Section 2 we describe the 3D CT discrete model and we formulate the constrained optimization problem; in Section 3 we present the proposed SGP algorithm; in Section 4 we show the numerical results obtained on a 3D phantom; finally, in Section 5 we draw some conclusions.

2 Problem formulation

In a 3D cone beam tomography system a cone of X-rays is emitted by a source rotating along an arc or a circular trajectory around the object of interest (see Figure 1) from a fixed number of positions (or angles). The rays, partially absorbed by the object, are projected on a flat panel detector (that can possibly move with the source) and then recorded.

Following the Lambert-Beer's model that relates the recorded value $b_i^{(\theta)}$ at each pixel i of the detector, for a fixed angle θ, with the attenuation coefficient μ at each position $\mathbf{w}$ of the object we obtain the image formation

model for X-rays tomographic images; in details,

$$b_i^{(\theta)} = \exp\left(-\int_{L_\theta} \mu(\mathbf{w}) d\ell\right) + \eta_i^{(\theta)}, \quad i = 1, \ldots, N_p, \quad \theta = 1, \ldots N_\theta, \tag{3}$$

where N_p is the number of pixels in the detector, N_θ is the number of angles, L_θ is the line followed by the X-ray beam through the object, $\mu(\mathbf{w})$ is the linear attenuation coefficient at the position $\mathbf{w}$, depending on the material in the object and characterizing the structures inside the object, and, lastly, $\eta_i^{(\theta)}$ is the noise measured at the detector (pixel i, angle θ) and it includes scattering and electronic noise.

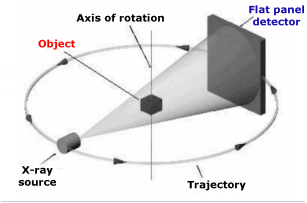


Figure 1: scheme of a 3D X-rays CT tomography scanning geometry.

The IIR methods consider the discrete linearization of (3):

$$Ax = b \tag{4}$$

where $b \in \mathbb{R}^{N_p \times N_\theta}$ $(b > 0)$ is the vector of recorded projections affected by noise, $x \in \mathbb{R}^{N_v}$ represents the discretization of $\mu(\mathbf{w})$ in the N_v voxels of the object (lexicographically ordered in a vector) and $A \in \mathbb{R}^{(N_p \times N_\theta) \times N_v}$ is the matrix describing the system geometry. In real applications, N_p is of the order of millions and N_v is of order of few billions and N_θ is of order 10^1 for sparse tomography. Different algorithms can be found in literature for the computation of A; we use here the Siddon algorithm [33] based on geometrical ray-tracing. Each element $a_{i,j}^{\theta}$ of A represents the length of the intersection of the ray, emitted at angle θ, recorded by pixel i of the detector, with the voxel j (in this notation, the pixels of the detector and the voxels of the object are lexicographically ordered in vectors). Moreover, in a reasonable physical setting, the matrix A has elements greater than zero in each column, because each voxel is projected at least once onto the detector. In the case of reduced sampling acquisition as is our case, $N_p \times N_\theta < N_v$ hence the linear system (4) has infinite possible solutions. Furthermore, since the linear system comes from the discretization of an integral equation of the first kind (3), the problem is ill-conditioned and some of the solutions of (4) are dominated by noise; thus, regularization strategies are necessary. Then, the problem can be reformulated as a penalized optimization problem of the form [35]:

$$\min_{x \geq 0} f(x) = J(x) + \lambda TV_\beta(x) \tag{5}$$

where $TV_\beta(x)$ is a smoothed differentiable version of the TV function defined as [39]:

$$TV_\beta(x) = \sum_{j=1}^{N_v} (\|\nabla x_j\|_2^2 + \beta^2)^{1/2} \tag{6}$$

with β a positive small parameter.

For what concerns the fit-to-data function $J(x)$, its expression is related to the noise on the data. Following a Maximum Likelihood approach [4], if the noise has a Gaussian distribution, the Least Squares function:

$$J(x) = \frac{1}{2}\|Ax - b\|_2^2 \tag{7}$$

gives the appropriate fit-to-data function, while, if the noise has a Poisson distribution, the Kullback-Leibler divergence

$$J(x) = \sum_{i=1}^{N_p \times N_\theta} \left(\sum_{j=1}^{N_v} A_{ij}x_j + bg - b_i - b_i log \frac{\sum_{j=1}^{N_v} A_{ij}x_j + bg}{b_i} \right) \tag{8}$$

($bg > 0$ is the background value) is the suitable term. The noise on the CT data is mixed Poisson (due to the X-rays particles behaviour) and Gaussian (due to the recording digital system) and the dominant one depends on the particular system considered. Hence, we consider in this paper the two different cases in which the fit-to-data term $J(x)$ is defined as in (7) or as is (8). In both cases, the objective function of the problem (5) is coercive and strictly convex on the nonnegative orthant, therefore the problem has a unique solution.

2.1 Problem discretization

For the discussion in the next section, it is useful to introduce the discretization of the problem in the 3D setting, by using the notation j_x, j_y, j_z to indicate the indices of a voxel of the discrete object on the three cartesian axes.

The $TV_\beta(x)$ function is discretized by forward differences with boundary periodic conditions. The discrete $TV_\beta(x)$ function can be written as:

$$TV_\beta(x) := \frac{1}{2} \sum_{j_x=1}^{N_x} \sum_{j_y=1}^{N_y} \sum_{j_z=1}^{N_z} \phi(\delta^2 x_{j_x,j_y,j_z}) \tag{9}$$

where $N_x \times N_y \times N_z = N_v$,

$$\delta^2 x_{j_x,j_y,j_z} = (x_{j_x+1,j_y,j_z} - x_{j_x,j_y,j_z})^2 + (x_{j_x,j_y+1,j_z} - x_{j_x,j_y,j_z})^2 + (x_{j_x,j_y,j_z+1} - x_{j_x,j_y,j_z})^2$$

and

$$\phi(t) = 2\sqrt{t + \beta^2}.$$

Table 1: SGP algorithm.

Initialize: $x^{(0)} \geq 0,\ \gamma, \sigma \in (0,1),\ 0 < \alpha_{min} \leq \alpha_{max},\ \alpha_0 \in [\alpha_{min}, \alpha_{max}],\ D_0 \in \mathcal{D}_{\rho_0}$;
for $k = 0, 1, \ldots$
  $d^{(k)} = P_+\left(x^{(k)} - \alpha_k D_k \nabla f(x^{(k)})\right) - x^{(k)}$; (scaled gradient projection step)
  $\eta_k = 1$;
  while $f(x^{(k)} + \eta_k d^{(k)}) > f(x^{(k)}) + \sigma \eta_k \nabla f(x^{(k)})^T d^{(k)}$
    $\eta_k = \gamma \eta_k$; (backtracking step)
  end
  $x^{(k+1)} = x^{(k)} + \eta_k d^{(k)}$;
  define the diagonal scaling matrix $D_{k+1} \in \mathcal{D}_{\rho_{k+1}}$; (scaling updating rule)
  define the step–length $\alpha_{k+1} \in [\alpha_{min}, \alpha_{max}]$; (step–length updating rule)
end

In order to better explain some details of the SGP algorithm presented in the next section, it is convenient to recall also the form of the (j_x, j_y, j_z) entry of the gradient of $TV_\beta(x)$:

$$\begin{aligned}
\frac{\partial TV_\beta}{\partial x_{j_x,j_y,j_z}}(x) &= \frac{1}{2}\frac{\partial}{\partial x_{j_x,j_y,j_z}}\left(\phi(\delta^2 x_{j_x,j_y,j_z}) + \phi(\delta^2 x_{j_x-1,j_y,j_z}) + \phi(\delta^2 x_{j_x,j_y-1,j_z}) + \phi(\delta^2 x_{j_x,j_y,j_z-1})\right) \\
&= \left(\phi'(\delta^2 x_{j_x,j_y,j_z})(3x_{j_x,j_y,j_z} - x_{j_x+1,j_y,j_z} - x_{j_x,j_y+1,j_z} - x_{j_x,j_y,j_z+1})\right) + \\
&+ \left(\phi'(\delta^2 x_{j_x-1,j_y,j_z})(x_{j_x,j_y,j_z} - x_{j_x-1,j_y,j_z})\right) + \\
&+ \left(\phi'(\delta^2 x_{j_x,j_y-1,j_z})(x_{j_x,j_y,j_z} - x_{j_x,j_y-1,j_z})\right) + \\
&+ \left(\phi'(\delta^2 x_{j_x,j_y,j_z-1})(x_{j_x,j_y,j_z} - x_{j_x,j_y,j_z-1})\right).
\end{aligned}$$

3 A scaled gradient approach for CT image reconstruction

In this section we recall a general framework that allows to design first-order descent methods for problem (5) able to exploit both scaled gradient directions and effective step-length selections. This framework is provided by the class of the Scaled Gradient Projection (SGP) methods [6, 10] that can be described as in Table 1.

The following notation is used in the SGP description: $\mathcal{D}_\rho$ denotes the set of diagonal matrices D with entries $\bar{d}_{j,j},\ j = 1, \ldots, n$, such that $\frac{1}{\rho} \leq \bar{d}_{j,j} \leq \rho$, with $\rho > 1$, and $P_+(z)$ is the euclidean projection of the vector $z \in \mathbb{R}^n$ onto the nonnegative orthant. At the k-th iteration of the SGP algorithm, a matrix $D_k \in \mathcal{D}_{\rho_k}$ is used to define the scaled gradient direction $-D_k \nabla f(x^{(k)})$, along which a step depending on the step-length $\alpha_k > 0$ is performed starting from $x^{(k)}$; then, by projecting the resulting vector onto the nonnegative orthant, the descent direction $d^{(k)}$ is obtained. This direction is exploited in a line-search procedure for generating the new approximation $x^{(k+1)}$ in which the objective function achieves a sufficient reduction with respect to $f(x^{(k)})$.

Finally, the scaling matrix and the step-length parameter are prepared for a new iteration by means of suitable adaptive updating rules.

For an effective application of SGP to problem (5), important theoretical and practical aspects need to be discussed. The main convergence properties of SGP algorithm are stated in the following proposition, whose proof can be found in [9].

Proposition 3.1. *Let $\{x^{(k)}\}$ be the sequence generated by applying the SGP method to problem* (5) *with the fit-to-data term $J(x)$ defined as in* (7) *or* (8)*. The following properties hold true:*

i) if $\rho_k^2 = 1 + \zeta_k,\ \zeta_k \geq 0,\ \sum_{k=0}^{\infty} \zeta_k < \infty$, then the sequence $\{x^{(k)}\}$ converges to the solution of the problem.

ii) if f^* denotes the optimal value of the objective function of problem (5)*, then*

$$f(x^{(k)}) - f^* = \mathcal{O}(1/k)$$

Proposition 3.1 ensures that SGP algorithm converges without restrictive assumptions on the step-length parameter α_k and the diagonal scaling matrix D_k, whose choices can be directed to accelerate the convergence rate of the scheme. Even if the theoretical convergence rate $\mathcal{O}(1/k)$ on the objective function values is lower than the rate $\mathcal{O}(1/k^2)$ of some optimal first-order methods exploiting extrapolation/inertial steps [28, 5, 2, 21], the practical performance of SGP method, achievable by suitable selections of D_k and α_k, is very well comparable with the convergence rate of the optimal algorithms [9, 30, 8, 13]. In the following we provide the updating rules for D_k and α_k that allow SGP to efficiently solve problem (5).

In order to update the diagonal scaling matrix D_k, we adapt the strategy proposed in [26] to the SGP framework. This strategy has shown the ability to force convergence acceleration, especially in the first steps of the iterative process [10, 7]. We define the diagonal scaling matrix by means of special splittings of the gradient of the objective function:

$$\nabla f(x) = V^f(x) - U^f(x), \qquad V^f(x) > 0, \quad U^f(x) \geq 0, \tag{10}$$

where $V^f(x)$ and $U^f(x)$ are obtained as:

$$V^f(x) = V^J(x) + \lambda V^{TV}(x), \quad U^f(x) = U^J(x) + \lambda U^{TV}(x),$$

with

$$\begin{array}{rcl} \nabla J(x) & = & V^J(x) - U^J(x), \qquad V^J(x) > 0, \quad U^J(x) \geq 0, \\ \nabla TV_\beta(x) & = & V^{TV}(x) - U^{TV}(x), \quad V^{TV}(x) > 0, \quad U^{TV}(x) \geq 0. \end{array}$$

Given the splitting (10), in [26] the choice $d_{j,j}^{(k+1)} = \frac{x_j^{(k+1)}}{V_j^f(x^{(k+1)})}, \quad j = 1, \ldots, n,$ is suggested for the diagonal entries of the matrix D_{k+1}. In order to exploit

a similar idea within the SGP scheme, we propose to update the diagonal scaling matrix in the following way:

$$\bar{d}_{j,j}^{(k+1)} = \min\left(\rho_{k+1}, \max\left(\frac{1}{\rho_{k+1}}, \frac{x_j^{(k+1)}}{V_j^f(x^{(k+1)})}\right)\right), \qquad j = 1, \ldots, n.$$

The vectors $V^J(x)$ and $V^{TV}(x)$ defining $V^f(x)$ are set by taking into account the special form of $\nabla J(x)$ and $\nabla TV_\beta(x)$, respectively.
When the $J(x)$ is the LS function (7), the gradient of the fit-to-data term has the form:

$$\nabla J(x) = A^T Ax - A^T b;$$

in this first case, we choose:

$$V^J(x) = A^T Ax, \quad U^J(x) = A^T b.$$

When the fit-to-data term $J(x)$ is the KL function (8), we have that:

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

where $\mathbf{1} \in \mathbb{R}^n$ is a vector whose components are all equal to one and $Y = \text{diag}\,(Ax + bg)$ is a diagonal matrix with the entries of $(Ax + bg)$ on the main diagonal; in this second case we choose:

$$V^J(x) = A^T \mathbf{1}, \quad U^J(x) = A^T Y^{-1} b.$$

Following the notation introduced in Section 2 for the entries of $\nabla TV_\beta(x)$, we set the components of $V^{TV}(x)$ as

$$\begin{aligned} V^{TV}_{j_x,j_y,j_z}(x) \quad = \quad & \big(3\phi'(\delta^2 x_{j_x,j_y,j_z}) + \phi'(\delta^2 x_{j_x-1,j_y,j_z}) + \\ & \phi'(\delta^2 x_{j_x,j_y-1,j_z}) + \phi'(\delta^2 x_{j_x,j_y,j_z-1})\big)\, x_{j_x,j_y,j_z}. \end{aligned}$$

Following the suggestions in [8, 13], the parameter ρ_{k+1} is chosen as $\rho_{k+1} = \sqrt{1 + 10^{15}/(k+1)^{2.1}}$.

Once the scaling matrix D_{k+1} has been defined, a new value for the step-length α_{k+1} can be computed with the aim to achieve further acceleration of the iterative process. The wide literature of the last decades identifies the Barzilai-Borwein (BB) rules as effective selection rules for the step-length updating in gradient methods [1]. In particular, recent studies have shown that selection rules based on special adaptive alternations of the two classical BB rules generally provide the best performance [41, 18]. For these reasons, first we derive the two BB rules for the step used by SGP and then we exploit these step-lengths rules within the alternating strategy proposed in [18], that has provided interesting convergence rate acceleration in many imaging algorithms [10, 30]. Due to the use of scaled gradient directions,

by applying the quasi-Newton properties on which the classical BB rules are based, the following step-lengths are obtained:

$$\alpha_{k+1}^{\mathrm{BB1}} = \arg\min_{\alpha_k \in \mathbb{R}} \|(\alpha_k D_{k+1})^{-1} s^{(k)} - z^{(k)}\|_2 = \frac{{s^{(k)}}^T D_{k+1}^{-1} D_{k+1}^{-1} s^{(k)}}{{s^{(k)}}^T D_{k+1}^{-1} z^{(k)}} \qquad (11)$$

and

$$\alpha_{k+1}^{\mathrm{BB2}} = \arg\min_{\alpha_k \in \mathbb{R}} \|s^{(k)} - (\alpha_k D_{k+1}) z^{(k-1)}\|_2 = \frac{{s^{(k)}}^T D_{k+1} z^{(k)}}{{z^{(k)}}^T D_{k+1} D_{k+1} z^{(k)}} \qquad (12)$$

where $s^{(k)} = \left(x^{(k+1)} - x^{(k)}\right)$ and $z^{(k)} = \left(\nabla f(x^{(k+1)}) - \nabla f(x^{(k)})\right)$.
The alternating strategy introduced in [18] leads to the following selection rule:

if $\alpha_{k+1}^{\mathrm{BB2}} / \alpha_{k+1}^{\mathrm{BB1}} < \tau_k$

$\alpha_{k+1} = \min\left\{\alpha_j^{\mathrm{BB2}} \,:\, j = \max\{1, k+1-m_\alpha\}, \ldots, k+1\right\}$, $\tau_{k+1} = 0.9\,\tau_k$,

else

$\alpha_{k+1} = \alpha_{k+1}^{\mathrm{BB1}}$ $\tau_{k+1} = 1.1\,\tau_k$,

end;

where m_α is a nonnegative integer and τ_k is a positive real number (we refer to the next section for the setting of these parameters). Other step-length updating rules are currently investigated in literature [14, 19, 16]; in particular, in [12] it has been shown that, for the CT image recontruction application, the step-length selection proposed in [30] seems to well compare with the above alternated BB rule and deeper studies are in progress on this topic.

4 Numerical results

In this section we present the numerical results performed on a MacBook Pro, 3GHz Intel Core i7, 8 Gb of RAM, equipped with Matlab (Release 2015a). For performing the tests, we used some functions of the `TVReg` Matlab Toolbox, `http://www.imm.dtu.dk/~pcha/TVReg/` [22].

Test problem. We consider as the true object x^* the digital Shepp Logan phantom discretized in $N_v = N_x \times N_y \times N_z = 61 \times 61 \times 61 = 226981$ voxels lexicographically ordered in a vector. In Figure 2 the slices number 24, 31 and 35 in the z direction are shown. The projections have been created as:

$$b^* = A \cdot x^*$$

where A is the projection matrix, obtained with the functions in the `TVReg` Toolbox for a 3D geometry with random angles over the half of a sphere. The

detector is supposed with $N_p = 61 \times 61$ pixels and the number of angles N_θ varies in the set $\{19, 37, 55\}$. In all the cases the problem is underdetermined. The projections are corrupted by noise, with both Gaussian and Poisson distribution, as specified in the following subsections.

Stopping criterion and parameters. We describe here the stopping criterion for the SGP algorithm and the setting of its main parameters. If k is the index of the current iteration and $S_k^f := \frac{|f(x^{(k+1)}) - f(x^{(k)})|}{|f(x^{(k)})|}$ is the relative distance between successive values of the objective function, we consider the conditions

$$S_k^f \leq \epsilon_1, \qquad \frac{1}{p} \sum_{j=0}^{p-1} S_{k-j}^f \leq \epsilon_2 \ \text{ if } \ k \geq p-1,$$

where $\epsilon_1 = 10^{-6}$, $p = 20$ and $\epsilon_2 = 10^{-5}$; the SGP stopping criterion consists in satisfying both the conditions or performing a maximum number of $k = 1000$ iterations.

For what concerns the SGP parameters, the setting reported below is used:

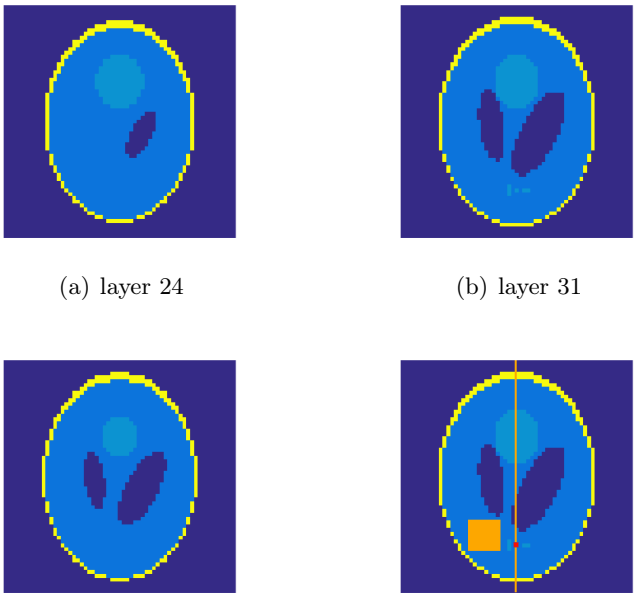
- $\gamma = 0.4$ and $\sigma = 10^{-4}$ as backtracking parameters;
- $\alpha_{min} = 10^{-10}$, $\alpha_{max} = 10^5$, $\alpha_0 = 1$, $m_\alpha = 2$ and $\tau_0 = 0.5$ for the step-length selection.

Results evaluation. In order to evaluate the reconstruction results, we consider the following parameters: the Relative Error (*Relerr*) between the exact volume x^* and the reconstructed image $\tilde{x}$ ($Relerr = \|x^* - \tilde{x}\|_2 / \|x^*\|_2$) and the Standard Deviation (*StdDev*) of the image values inside a small crop (represented by the yellow 8×8 pixel square in Figure 2(d)) on a uniform region in the central layer. The reconstructed images are also evaluated by plotting the profile of the yellow vertical line in Figure 2(d), vertical profile (VP), and the profile over the 61 layers in the z direction of the red pixel in Figure 2(d), depth profile (DP).

We show the results obtained by the algorithms at three different temporal windows: at 5 seconds (10-15 iterations), for simulating a real-time execution; at 20 seconds (50-60 iterations), corresponding to an over-time of few minutes in real applications; at the convergence, i.e., when the convergence criterium is satisfied (this is a long execution that can be performed only off-line in a real application). We think that each of these three different outputs may have a practical interest and together they represent the evolution of the methods in time.

4.1 Case of Gaussian noise. Results with the LS data function.

In this paragraph we show the results obtained on the simulated data $b = b^* + e$, where e is the vector representing white Gaussian noise with



(a) layer 24 (b) layer 31

(c) layer 35 (d) layer 31

Figure 2: different layers in the z-direction of the original phantom. For the analysis of the results, in (d) some interesting features in layer 31 are highlighted: a yellow small crop where the Standard Deviation is compute, a yellow line along which we analyse the vertical profile and a red pixel to examine the depth profile.

level defined as $\nu = \frac{\|e\|_2}{\|b^*\|_2}$; we consider here $\nu = 0.01$, corresponding to a Signal-to-Noise Ratio SNR $:= 20 \cdot log_{10}(\frac{||b||_2}{||b-b^*||_2})$ of about 40. We consider the fit-to-data function $J(x)$ as the LS function and we set the TV smoothing parameter β equal to 0.001 in all the experiments. The regularization parameter λ has been heuristically set to 0.09; we have experimented that for this test the model is not very sensitive to the value of λ (similar results have been obtained with different values of λ in the interval $[0.005, 0.5]$).

We compare the results of the proposed SGP method with the non-scaled GP method (GP) (with the step-length selection used by SGP, but with $D_{k+1} = I$ in the definitions (11) and (12) of the BB rules) and with the UPN method proposed in [21], implemented in the `TVReg` toolbox. The UPN method has been equipped with the same stopping criterion used for SGP and its parameters have been set at their best values after a careful tuning.

In Table 2 we present the results obtained with different number of views ($N_\theta = 19, 37, 55$) for the GP, SGP and UPN methods. In the columns

from left to right we report the Relative Error, the Standard Deviation, the objective function value and the number of performed iterations in the three considered temporal windows: at 5 seconds, at 20 seconds and at convergence. From the table, we see that the SGP method outperforms the others in the first iterations (5 and 20 seconds) for all the considered angles; at convergence, all the methods give very similar results. The reconstructions of central layer (layer 31) obtained with the three considered methods in the case $N_\theta = 37$ are shown in Figure 3.

(a) GP method at 5 seconds, 20 seconds, convergence.

(b) SGP method at 5 seconds, 20 seconds, convergence.

(c) UPN method at 5 seconds, 20 seconds, convergence.

Figure 3: reconstructions obtained in case of Gaussian noise on the data. From the left to the right: reconstructions after 5 seconds, after 20 seconds, at convergence.

In Figure 4 the errors versus the iterations (on the left) and the objective function values versus the iterations (on the right) are shown in log-log scale. We compare here the GP method (blue line), the SGP method (red line) and the UPN method (green line) up to the convergence of the methods. The advantage of using the scaling matrix is evident, especially in the first iterations.

			Relerr	*StdDev*	fun	iters
$N_\theta = 19$	GP	5 secs	0.3816	0.0559	5574.36	16
		20 secs	0.1548	0.0312	1596.75	69
		conv	0.0559	0.0052	1498.49	263
	SGP	5 secs	0.2637	0.0463	3396.96	19
		20 secs	0.1178	0.0241	1560.48	71
		conv	0.0543	0.0052	1498.57	198
	UPN	5 secs	0.3785	0.0640	6075.55	11
		20 secs	0.1786	0.0414	1652.62	48
		conv	0.0580	0.0056	1484.46	606
$N_\theta = 37$	GP	5 secs	0.3475	0.0331	11537.90	16
		20 secs	0.0898	0.0241	1795.02	64
		conv	0.0245	0.0028	1645.77	154
	SGP	5 secs	0.1840	0.0261	4335.56	18
		20 secs	0.0477	0.0777	1689.30	66
		conv	0.0247	0.0030	1646.39	194
	UPN	5 secs	0.4001	0.0281	19918.90	8
		20 secs	0.1045	0.0343	1917.49	46
		conv	0.0241	0.0028	1632.07	224
$N_\theta = 55$	GP	5 secs	0.3091	0.0438	14306.80	15
		20 secs	0.0779	0.0276	1997.80	60
		conv	0.0199	0.0044	1783.11	142
	SGP	5 secs	0.2148	0.0306	9662.70	16
		20 secs	0.0277	0.0077	1814.38	60
		conv	0.0199	0.0044	1783.60	147
	UPN	5 secs	0.4315	0.0325	40865.00	6
		20 secs	0.0677	0.0293	2033.66	46
		conv	0.0199	0.0043	1769.47	200

Table 2: results obtained on the test problems with data affected by Gaussian noise.

Figure 5 displays the VP (on the left) and DP (on the right) after 5 seconds, 20 seconds and at convergence. We compare again the GP reconstruction (blue line), SGP reconstruction (red line) and UPN reconstruction (green line) with the phantom profile (grey line). The VP plots confirm that after few iterations (5 seconds) we can identify, in the signal reconstructed by the SGP method, all the objects with a good approximation of their intensity; in the DP plot after 20 seconds the SGP method has almost completely eliminated the noise, while the GP and UPN plots show a residual noise yet. In Figure 6 we show a zoom of Figure 5 for the pixels between 4 and 27 for the VP plots on the left and between 19 and 43 for the DP plots on the right, in order to clearer represent the behaviour of the method for reconstructing the small object represented by the red dot in Figure 2(d). We can see that the SGP profiles are less noisy than the others and in the DP the peak of the SGP line is the closest the exact one.

4.2 Case of Poisson noise. Results with the KL data function.

We consider now some tests where the projections are affected by Poisson noise, with SNR $\simeq 40$ and background $bg = 10^{-5}$. The problem is solved by using the KL fit-to-data function in (8). In this case the regularization parameter λ has been heuristically set to 0.03 and the TV smoothing parameter $\beta = 0.01$; we have experimented that, as in the case of Gaussian noise, similar results have been obtained with different values of λ in the interval $[0.001, 0.1]$.

For this model, we compare the results obtained with the GP and the SGP methods, since the UPN method is provided only for the case of LS fit-to-data function. Table 3 reports the results in the case $N_\theta = 19, 37, 55$, with the same information of Table 2. For the KL model, the performance improvement due to the scaling is more consistent than in the LS model, as it can be seen by the Relative Error and Standard Deviation values. If the number of performed iterations in the last column is equal to 1000 an asterisk reminds that the algorithm has stopped after reaching the maximum number of iterations. We want to stress that this happens only for the GP method, confirming its slower convergence rate. In Figure 7 we plot the Relative Error versus the iterations in the left panel, while the objective function values versus the iterations are displayed in the right panel. In Figure 8 the reconstructions of the layer 31 obtained with both GP and SGP methods after 5 seconds, 20 seconds and at convergence are represented. Both the plots and the images confirm the convergence acceleration of the scaled algorithm with respect to the non-scaled one.

The analysis of VP and DP profiles in the case of $N_\theta = 37$ in Figure 9 shows that the scaling allows recovering very good profiles in very short time: after 20 seconds the line of the reconstructed image almost overlap the

			Relerr	*StdDev*	fun	iters
$N_\theta = 19$	GP	5 secs	0.6305	0.0154	16781.30	12
		20 secs	0.5685	0.0204	9031.95	61
		conv	0.4353	0.0290	3085.58	877
	SGP	5 secs	0.2145	0.0332	767.85	19
		20 secs	0.0984	0.0097	524.80	72
		conv	0.0869	0.0063	522.01	172
$N_\theta = 37$	GP	5 secs	0.6914	0.0085	55813.10	11
		20 secs	0.6349	0.0107	31273.40	53
		conv	0.4045	0.0180	4913.71	1000*
	SGP	5 secs	0.1752	0.0201	1127.98	18
		20 secs	0.0798	0.0129	597.74	66
		conv	0.0335	0.0031	545.41	393
$N_\theta = 55$	GP	5 secs	0.6919	0.0105	85188.70	12
		20 secs	0.6495	0.0121	54467.70	52
		conv	0.4191	0.0222	8215.40	1000*
	SGP	5 secs	0.1745	0.0226	1552.37	16
		20 secs	0.0496	0.0107	596.71	60
		conv	0.0387	0.0068	580.17	198

Table 3: results obtained on the test problems with data affected by Poisson noise.

line of the exact phantom (the only exception is the small peak in the center of the DP).

4.3 Algorithms reliability with regard to noise

Finally, in this paragraph we test the reliability of the proposed SGP algorithm with regard to noise of increasing levels on the projections. We have performed some tests by adding noise of different intensity on the data and solving the reconstruction problem with a suitable regularization parameter in each case. We show in Figure 10 the results obtained in the different executions after 20 seconds. We plot the relative errors as a function of the SNR in the case of Gaussian noise (left panel) and Poisson noise (right panel). The relative error obviously increases with increasing noise but the SGP method always performs better for all noise levels. We highlight in particular the remarkable difference between the errors obtained with SGP and GP methods in the case of Poisson noise.

5 Conclusions

In this paper we have presented a scaled gradient projection method for the reconstruction of 3D X-rays tomographic images from limited data. The problem is modeled in a variational framework as the nonnegative minimization of a function constituted by a fit-to-data term and a smoothed version of the Total Variation regularization function weighted by a positive parameter. In this framework we have considered both the Least Squares and the Kullback-Leibler fit-to-data functions. The proposed strategy for scaling the gradient directions by means of diagonal matrices derived by suitably splitting the objective gradient, combined with a step-length selection based on the Barzilai-Borwein rules, greatly improve the performance of the algorithm especially in the first iterations, as required by the specific application, where a good image is required in a very short time for clinical requirements. The results obtained on a medium size 3D phantom are very encouraging for a future use of the SGP method on real data for clinical reconstructions. Work in progress concerns the analysis of recent step-length selections [30, 16], which appear valid alternatives to the Barzilai-Borwein rules, and the study of suitable techniques for combining these step-length selections with the scaling strategy presented in this paper.

Acknowledgement

This work has been partially supported by the Italian Institute GNCS - INdAM and by the FAR2015 project of the University of Modena and Reggio Emilia, Italy.

References

[1] J. Barzilai and J.M. Borwein. Two-point step size gradient methods. *IMA J. Numer. Anal.*, 8, 141–148, 1988.

[2] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1), 183–202, 2009.

[3] M. Beister, D. Kolditz and W. Kalender. Iterative reconstruction methods in X-ray CT, *Physica Medica*, 28, 94–108, 2012.

[4] M. Bertero, H. Lantéri, and L. Zanni. Iterative image reconstruction: a point of view. *Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT)*, Y. Censor et al. eds., Birkhauser-Verlag, 37-63, 2008.

[5] D. Bertsekas. *Convex optimization theory. Supplementary Chapter 6 on convex optimization algorithms.* Athena Scientific, Belmont, 2009.

[6] E.G. Birgin, J.M. Martinez and M. Raydan. Inexact spectral projected gradient methods on convex sets. *IMA J. Numer. Anal.*, 23, 539–559, 2003.

[7] S. Bonettini, G. Landi, E. Loli Piccolomini and L. Zanni. Scaling techniques for gradient projection-type methods in astronomical image deblurring. *Int. J. Comput. Math.*, 90(1), 9–29, 2013.

[8] S. Bonettini, F. Porta and V. Ruggiero. A variable metric inertial method for convex optimization. *SIAM J. Sci. Comput.*, 31(4), A2558–A2584, 2016.

[9] S. Bonettini and M. Prato. New convergence results for the scaled gradient projection method. *Inv. Probl.*, 31(9), 1196–1211, 2015.

[10] S. Bonettini, R. Zanella and L. Zanni. A scaled gradient projection method for constrained image deblurring. *Inv. Probl.*, 25(1), 015002, 2009.

[11] D.J. Brenner and E. Hall. Computed tomography: an increasing source of radiation exposure, *N. Engl. J. Med.*, 357, 2277–2284, 2007.

[12] V.L. Coli, E. Loli Piccolomini, E. Morotti and L. Zanni. A fast gradient projection method for 3D image reconstruction from limited tomographic data. Journal of Physics : Conference Series. IOP Publishing, to appear.

[13] V.L. Coli, V. Ruggiero, and L. Zanni. Scaled first-order methods for a class of large-scale constrained least squares problems. In *Numerical Computations: Theory and Algorithms (NUMTA-2016)*, pages 040002–1 – 040002–4. AIP Publishing, 2016.

[14] R. De Asmundis , D. di Serafino, W. Hager, G. Toraldo and H. Zhang. An efficient gradient method using the Yuan steplength. *Comput. Optim. Appl.*, 59(3), 541–563, 2014.

[15] M. Defrise, C. Vanhove and X. Liu. An algorithm for total variation regularization in high-dimensional linear problems. *Inv. Probl.*, 52, 329–356, 2011.

[16] D. di Serafino, V. Ruggiero, G. Toraldo and L. Zanni. On the steplength selection in gradient methods for unconstrained optimization. *Appl. Math. Comput.*, 2017, http://dx.doi.org/10.1016/j.amc.2017.07.037 .

[17] L. Feldkamp, L. Davis and J. Kress. Practical cone-beam algorithm. *J. Opt. Soc. Am.*, 1, 612–619, 1984.

[18] G. Frassoldati, L. Zanni and G. Zanghirati. New adaptive stepsize selections in gradient methods. *J. Ind. Manag. Optim.*, 4(2), 299–312, 2008.

[19] C. Gonzaga and R.M. Schneider. On the steepest descent algorithm for quadratic functions. *Comput. Optim. Appl.*, 63(2), 523–542, 2016.

[20] C. Graff and E. Sidky. Compressive sensing in medical imaging. *Appl. Opt.*, 54(8), C23–C44, 2015.

[21] T.L. Jensen, J.H. Jørgensen, P.C. Hansen and S.H. Jensen. Implementation of an optimal first-order method for strongly convex total variation regularization. *BIT Numer. Math.*, 52, 329–356, 2012.

[22] J.H. Jørgensen, T.L. Jensen, P.C. Hansen, S.H. Jensen, E.Y. Sidky and X. Pan. Accelerated gradient methods for total-variation-based CT image reconstruction. In *11th Fully 3D Image Reconstruction in Radiology and Nuclear Medicins*, 435–438, 2011.

[23] D. Kim, D. Pal, J. Thibault and J.A. Fessler., Accelerating Ordered Subsets Image Reconstruction for X-ray CT Using Spatially Nonuniform Optimization Transfer *IEEE Trans. Med. Imag.* 32(11), 1965–1978, 2013.

[24] D. Kim, S. Ramani and J.A. Fessler. Combining ordered subsets and momentum for accelerated X-rays CT imaging reconstruction *IEEE Trans. Med. Imag.* 34(1), 167–178, 2015.

[25] K. Lange, D. Hunter and I. Yang. Optimization transfer using surrogate objective functions. *J. Comput. Graph. Stat.*, 9(1), 1–20, 2000.

[26] H. Lantéri, M. Roche and C. Aime. Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms. *Inv. Probl.*, 18(5), 1397–1419, 2002.

[27] E. Loli Piccolomini and E. Morotti. A fast TV-based iterative algorithm for digital breast tomosynthesis image reconstruction. *J. Algor. and Computat. Techn.*, 10(4), 277–289, 2016.

[28] Y. Nesterov. *Introductory lectures on convex optimization: a basic course*. Applied Optimization, Kluwer Academic Publ., Dordrecht, 2004.

[29] Y. Nesterov. Gradient methods for minimizing composite functions. *Math. Program.*, 140, 125–161, 2013.

[30] F. Porta, M. Prato and L. Zanni. A new steplength selection for scaled gradient methods with application to image deblurring. *J. Sci. Comput.*, 65(d), 895–919, 2015.

[31] R. Rangayyan, A. Dhawan and R. Gordon. Algorithms for limited-view computed tomography: an annotated bibliography and a challenge, *Appl. Optics*, 24(23), 4000–4012, 1985.

[32] S. Rose, M. Andersen, E.Y. Sidky and X. Pan. Noise properties of CT images reconstructed by use of constrained total-variation, data-discrepancy minimization. *Med. Phys.*, 42(5), 2690–2698, 2015.

[33] R.L. Siddon. Fast calculation of the exact radiological path for a three-dimensional CT array. *Med. Phys.*, 12(2), 252–255, 1985.

[34] E.Y. Sidky, J.H. Jørgensen and X. Pan. Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm. *Phys. Med. Biol.*, 57(10), 3065–3091, 2012.

[35] E.Y. Sidky, J.H. Jørgensen and X. Pan. First-order convex feasibility for x-ray CT. *Med. Phys.*, 40(3), 3115–1–15, 2013.

[36] E.Y. Sidky, C.M. Kao and X. Pan. Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. *J. Xray Sci. Technol.*, 14(2), 119–139, 2006.

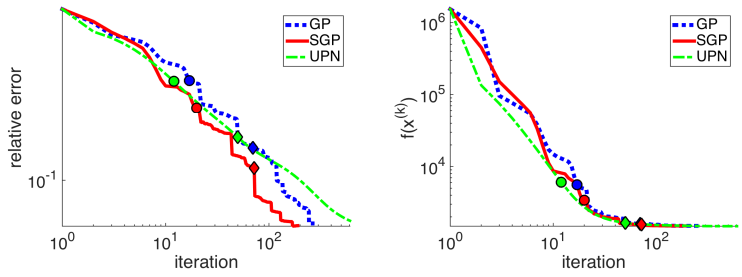
[37] E.Y. Sidky and X.Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Phys. Med. Biol.*, 53, 4777–4807, 2008.

[38] E.Y. Sidky, X. Pan, I.S. Reiser and R.M. Nishikawa. Enhanced imaging of microcalcifications in digital breast tomosynthesis through improved image-reconstruction algorithms. *Med. Phys.*, 36(11), 4920–4932, 2009.

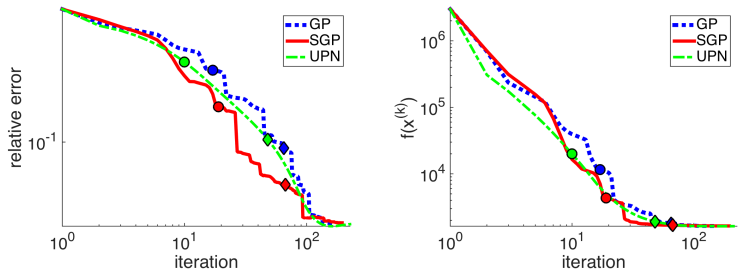
[39] C.R. Vogel. *Computational Methods for Inverse Problems.* SIAM, Philadelphia, PA, USA, 2002.

[40] H. Yu and G. Wang. A soft-threshold filtering approach for reconstruction from a limited number of projections. *Phys. Med. Biol.*, 55, 3905–3916, 2010.

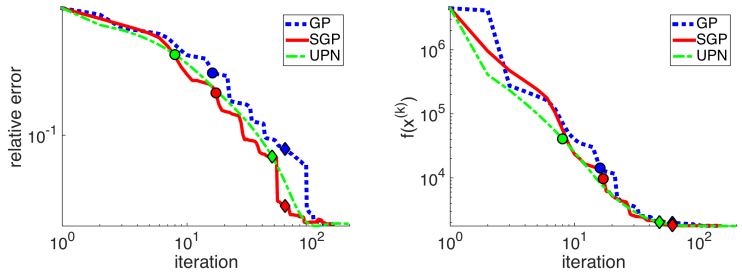
[41] B. Zhou, L. Gao and Y.H. Dai. Gradient methods with adaptive step-sizes. *Comput. Optim. Appl.*, 35(1), 69–86, 2006.



(a) $N_\theta = 19$

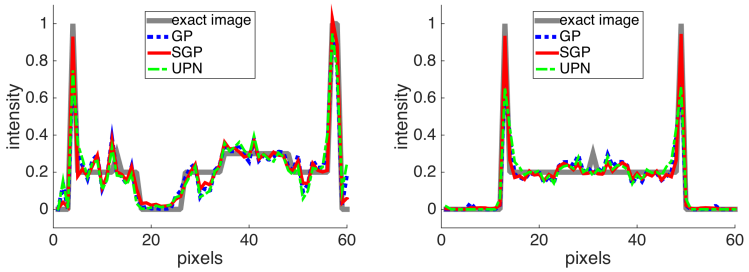


(b) $N_\theta = 37$

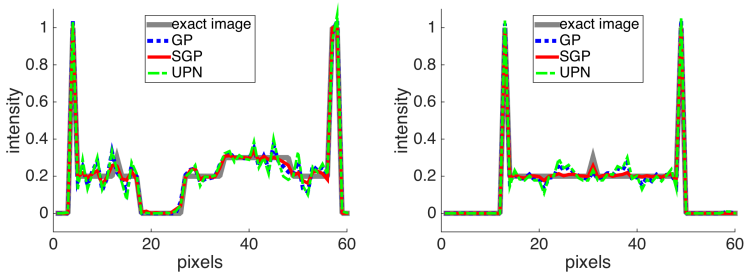


(c) $N_\theta = 55$

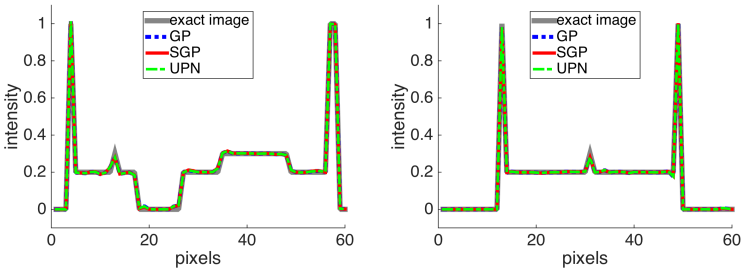
Figure 4: case of Gaussian noise. On the left: errors vs iterations; on the right: function values vs iterations. The circles and the diamonds represent the values at 5 and 20 seconds, respectively.



(a) Profiles after 5 seconds

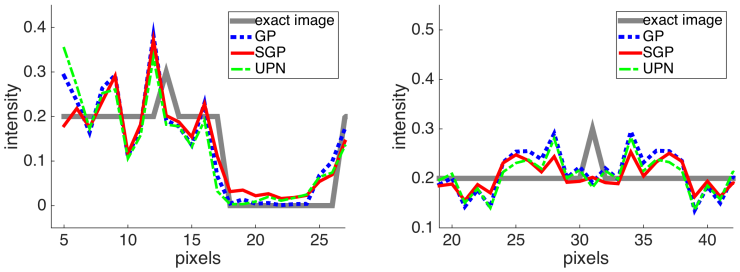


(b) Profiles after 20 seconds

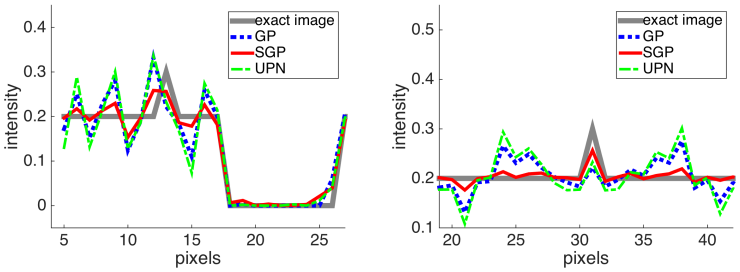


(c) Profiles at convergence

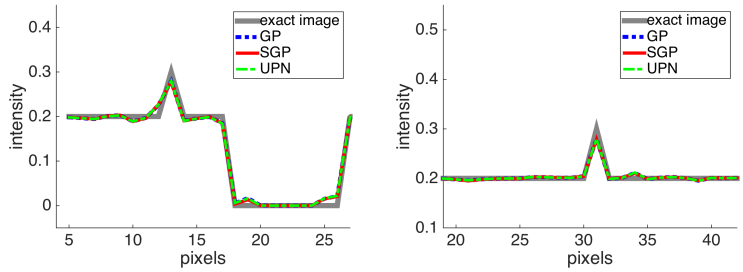
Figure 5: case of Gaussian noise. Profiles for 37 angles: on the left VP plots and on the rigth DP plots at different temporal windows.



(a) Profiles after 5 seconds

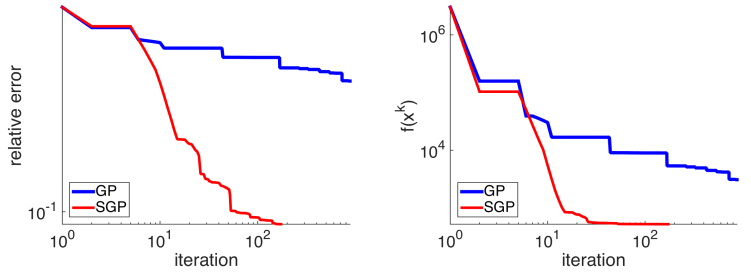


(b) Profiles after 20 seconds

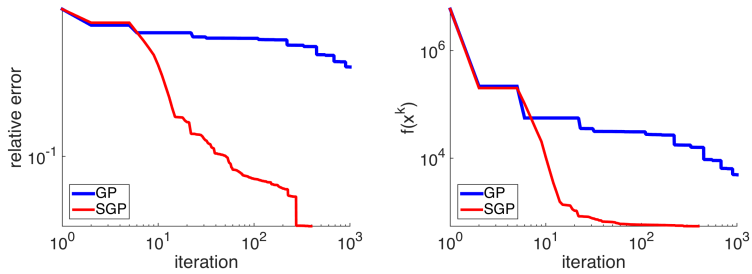


(c) Profiles at convergence

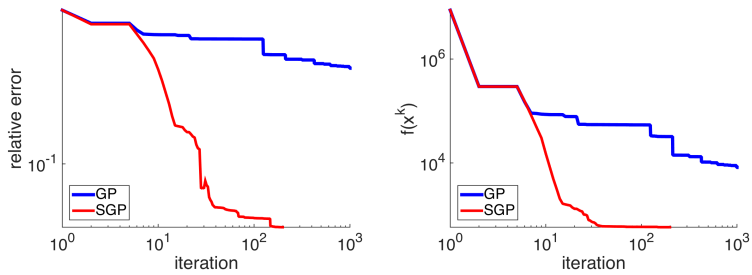
Figure 6: zoom of the plots in Figure 5.



(a) $N_\theta = 19$

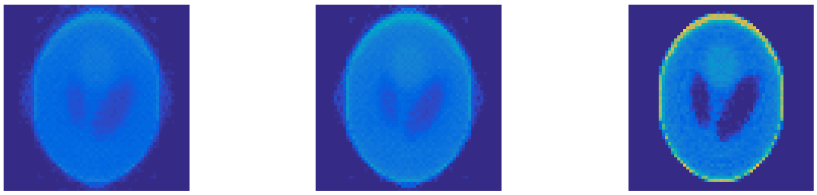


(b) $N_\theta = 37$



(c) $N_\theta = 55$

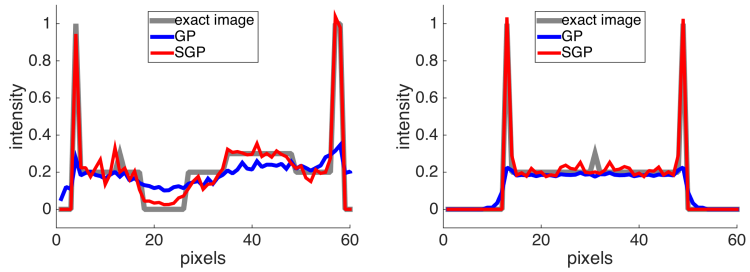
Figure 7: case of Poisson noise. On the left: errors vs iterations; on the right: function values vs iterations.



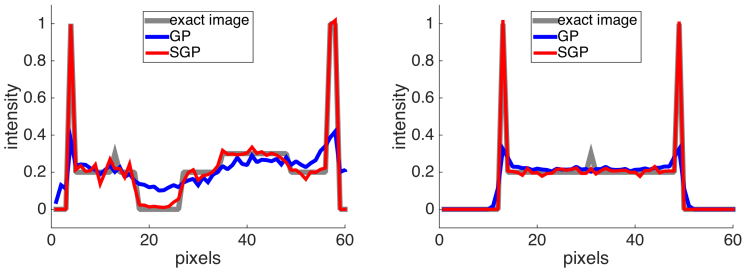
(a) GP method at 5 seconds, 20 seconds, convergence.

(b) SGP method at 5 seconds, 20 seconds, convergence.

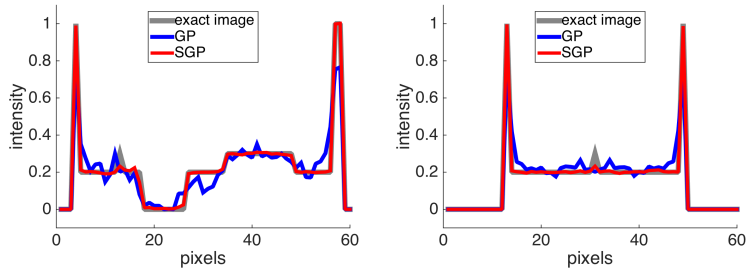
Figure 8: reconstructions obtained in case of Poisson noise on the data. From the left to the right: reconstructions after 5 seconds, after 20 seconds, at convergence.



(a) Profiles after 5 seconds



(b) Profiles after 20 seconds



(c) Profiles at convergence

Figure 9: case of Poisson noise. Profiles for 37 angles: on the left VP plots and on the rigth DP plots at different temporal windows.

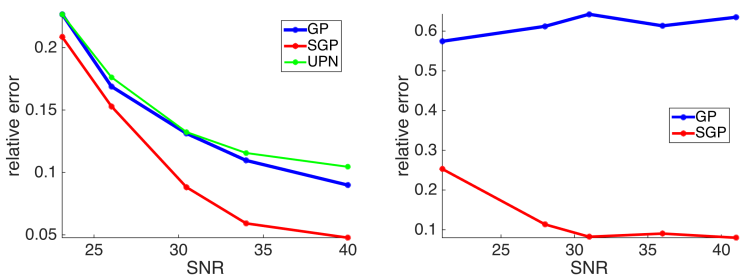


Figure 10: relative errors vs SNR after 20 seconds for all the methods. On the left panel, the results with Gaussian noise on the data; on the right panel, the results with Poisson noise on the data.