

Shearlet-based regularized ROI reconstruction in fan beam computed tomography

T.A. Bubba^a, D. Labate^b, G. Zanghirati^a, S. Bonettini^a and B. Goossens^c

^aDept. of Mathematics and Computer Science, University of Ferrara, and INdAM-GNCS
via G. Saragat 1, 44122 Ferrara, Italy;

^bDepartment of Mathematics, University of Houston
651 Phillip G. Hoffman, Houston, TX USA 77204-3008;

^cIPI, Dept. of Telecommunications and Information Processing, Ghent University–iMinds
Sint-Pietersnieuwstraat 41, B-9000 Ghent, Belgium.

ABSTRACT

Region-of-interest (ROI) reconstruction in computed tomography (CT) is a problem receiving increasing attention in the medical imaging community, due to its potential to lower exposure to X-ray radiation and to reduce the scanning time. Since the ROI reconstruction problem requires to deal with truncated projection images, classical CT reconstruction algorithms tend to become very unstable and the solution of this problem requires either *ad hoc* analytic formulas or more sophisticated numerical schemes. In this paper, we introduce a novel approach for ROI CT reconstruction, formulated as a convex optimization problem with a regularized functional based on shearlets or wavelets. Our numerical implementation consists of an iterative algorithm based on the scaled gradient projection method. As illustrated by numerical tests in the context of fan beam CT, our algorithm is insensitive to the location of the ROI and remains very stable also when the ROI size is rather small.

Keywords: Computed tomography, region-of-interest reconstruction, shearlets, wavelets, gradient projection methods

1. INTRODUCTION

Computed Tomography (CT) is a noninvasive imaging technique designed to recover the internal structure of a body from projections obtained by illuminating the body with X-ray radiation from many directions. The impact of this technique in medical diagnostics has been enormous, even though exposure to X-ray radiation comes with health hazards for patients. To reduce radiation exposure, several strategies have been explored such as sparsifying the numbers of X-ray projections or – if the goal is to reconstruct with high accuracy only a relatively small region-of-interest (ROI) inside the body – truncating X-ray projections only to those rays intersecting the ROI.

However, reconstructing a density function from its projections is an ill-posed problem, and the ill-posedness may become rather severe when projections are truncated,¹ as in the case of ROI CT. In fact, the direct application of classical algorithms such as Filtered Back-Projection or the FDK algorithms² (with the missing projection data set to zero) typically produces unacceptable visual artifacts and is very unstable to noise.

To address the problem of ROI reconstruction from truncated projections, a variety of methods, both analytic and algebraic, were proposed in recent years (see review in Ref. 3). In particular, it was shown that it is possible to derive analytic ROI reconstruction formulas from truncated data, but such formulas usually require restrictive assumptions on the location of the ROI and depend on the acquisition setting.^{4–6} Algebraic methods, on the

Further author information: (Send correspondence to T.A.B.)

T.A.B.: E-mail: bbbtnl@unife.it, Telephone: +39 0532 974789

D.L.: E-mail: dlabate@math.uh.edu, Telephone: (713) 743-3492

G.Z.: E-mail: gaetano.zanghirati@unife.it, Telephone: +39 0532 97478

S.B.: E-mail: silvia.bonettini@unife.it, Telephone: +39 0532 974785

B.G.: E-mail: bart.goossens@telin.ugent.be, Telephone: +32 9 264 7966

other hand, are generally more flexible but computationally more demanding. They typically involve some form of regularization to ensure a stable ROI reconstruction,⁷⁻⁹ but the performance of existing methods is usually rather sensitive to the ROI size and the presence of noise. We refer to Ref. 10 for additional remarks about the sensitivity of existing regularized reconstruction methods to the ROI size.

In this paper, we use a state-of-the-art iterative minimization algorithm to solve the ROI CT reconstruction problem, formulated as a convex optimization problem with a regularized functional based on multiscale methods including shearlets. Thanks to their ability to sparsely represent images with edges, shearlets are useful to ensure stable and robust reconstruction, even in the presence of noise.¹¹ Our numerical implementation is based on the scaled gradient projection method,¹² which is an accelerated first-order descent method for convex and non-convex objective functions. It is particularly effective when the feasible region is given by “simple” constraints, that is, projecting onto such a feasible region is not a heavy task. As illustrated by our numerical tests, our algorithm produces accurate ROI reconstruction for any ROI location and also for ROI sizes that are small with respect to the field of view.

2. THE REGION-OF-INTEREST TOMOGRAPHY PROBLEM

We start by setting some mathematical formalism to present the problem of the reconstruction of a function from its X-ray projections.

2.1 The Radon transform and the ROI CT problem

The tomographic problem consists in recovering a density function f from its linear projections, computed by integration over straight lines. Mathematically, this problem is formulated through the notion of the Radon transform.

Given an integrable function of two variables $f \in L^1(\mathbb{R}^2)$, the *Radon transform* of f at (θ, τ) is the line integral of f over the lines (or rays) $\ell(\theta, \tau)$ perpendicular to $\mathbf{e}_\theta = (\cos \theta, \sin \theta) \in \mathbb{S}^1$ with (signed) distance $\tau \in \mathbb{R}$ from the origin. That is,

$$\mathcal{R}f(\theta, \tau) = \int_{\ell(\theta, \tau)} f(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^2} \delta(\tau - \mathbf{x} \cdot \mathbf{e}_\theta) f(\mathbf{x}) d\mathbf{x},$$

where $\ell(\theta, \tau) = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x} \cdot \mathbf{e}_\theta = \tau\}$. Therefore, the Radon transform maps f on \mathbb{R}^2 into the set of its linear projections defined on the tangent space

$$\mathcal{T} = \{(\theta, \tau) : \theta \in [0, 2\pi), \tau \in \mathbb{R}\}.$$

In the following, we will refer to the Radon projections as *full sinogram* and we shall denote it by:

$$y(\theta, \tau) = \mathcal{R}f(\theta, \tau), \quad \theta \in [0, 2\pi), \tau \in \mathbb{R}. \quad (1)$$

In the ROI tomographic problem, projections are only collected for those rays meeting a region of interest inside the field of view and the goal is to recover the density function in the ROI.

Denoting the ROI as $S \subset \mathbb{R}^2$, the set of ROI-truncated projections is identified by the set

$$\mathcal{P}(S) = \{(\theta, \tau) \in \mathcal{T} : \ell(\theta, \tau) \cap S \neq \emptyset\} \subset \mathcal{T}.$$

The ROI reconstruction problem can now be formulated as the problem of reconstructing the density function f restricted to the ROI S from the truncated Radon projections:

$$y_0(\theta, \tau) = M(\theta, \tau) \mathcal{R}f(\theta, \tau), \quad (2)$$

where the mask function M is given by

$$M(\theta, \tau) = 1_{\mathcal{P}(S)}(\theta, \tau)$$

and 1_A is the indicator function of the set A . We will refer to y_0 as the *truncated sinogram*.

It is easy to see that

$$y = y_0 + (1 - M)y. \quad (3)$$

Equation (3) suggests to interpret the ROI reconstruction problem as an extrapolation problem. That is, given y_0 on $\mathcal{P}(S)$, the goal is to extrapolate this function to the region outside $\mathcal{P}(S)$. Since we cannot choose *any* extrapolation scheme to define y outside $\mathcal{P}(S)$, y should be subject to the global constraint

$$Rf = y \quad (4)$$

to ensure that y is the image of the Radon transform of a function $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. By applying M and $1 - M$ to the left-hand and right-hand sides of (4), respectively, we obtain the following equations:

$$M\mathcal{R}f = My = y_0 \quad (\text{data fidelity}) \quad (5)$$

$$(1 - M)\mathcal{R}f = (1 - M)y \quad (\text{data consistency}) \quad (6)$$

The data fidelity equation defines a constraint inside the ROI and the data consistency equation enforces accurate reconstruction inside the ROI.

In the following, we will assume that the ROI is a disk $S \subset \mathbb{R}^2$ with center $\mathbf{p}_{\text{ROI}} \in \mathbb{R}^2$ and radius $R_{\text{ROI}} \in \mathbb{R}$. Thus, $\mathcal{P}(S) = \{|\tau - \mathbf{p}_{\text{ROI}} \cdot \mathbf{e}_\theta| < R_{\text{ROI}}\}$. Clearly, more general convex ROIs can be handled by finding the minimal enclosing disk for this ROI and reconstructing the image for this disk.

In the next section we show how to combine these equations with a regularization to obtain a unique solution.

2.2 Solution of the ROI problem

It is known that the solution of the ROI problem is not guaranteed to be unique.¹ However, even when the uniqueness is ensured, the inversion of the Radon transform is an ill-posed problem and the ill-posedness may be more severe when projections are truncated, as in the case of the ROI CT problem.

A natural approach for obtaining a stable reconstruction of f from equation (2) is by computing the least squares solution \hat{f}

$$\hat{f} = \arg \min_f \|\mathcal{M}\mathcal{R}f - y_0\|_2^2. \quad (7)$$

However, the minimizer of this problem is not unique in general, since the set of solutions is the affine subspace

$$V = \{f \in L^2(\mathbb{R}^2) : y = \mathcal{R}f \quad \text{and} \quad My = y_0\}.$$

A classical approach to achieve uniqueness is by using Tikhonov regularization, which imposes an additional norm condition by searching for the minimum-norm solution. In principle, this norm condition can be applied in the image domain or in the Radon-transform domain.

To define a norm condition in the projection domain, we recall the definition of the Riesz potential operator $I^{-\alpha}$ for a function $g \in L^2(\mathcal{T})$:

$$\mathcal{F}(I^\alpha g)(\theta, \xi) = |\xi|^{-\alpha} \mathcal{F}g(\theta, \xi), \quad \alpha < 2. \quad (8)$$

Using (8), the L^2 -norm of the function f can equivalently be calculated in the Radon-transform domain. Using the formula²

$$f = \frac{1}{4\pi} \mathcal{R}^* I^{-1} \mathcal{R}f, \quad (9)$$

where \mathcal{R}^* is the adjoint operator of \mathcal{R} , also known as backprojection operator, we have

$$\begin{aligned} \|f\|_2^2 &= \langle f, f \rangle = \frac{1}{4\pi} \langle f, \mathcal{R}^* I^{-1} \mathcal{R}f \rangle \\ &= \frac{1}{4\pi} \langle \mathcal{R}f, I^{-1} \mathcal{R}f \rangle \\ &= \frac{1}{4\pi} \langle I^{-\frac{1}{2}} \mathcal{R}f, I^{-\frac{1}{2}} \mathcal{R}f \rangle \\ &= \frac{1}{4\pi} \left\| I^{-\frac{1}{2}} \mathcal{R}f \right\|_2^2 \end{aligned} \quad (10)$$

Thus, we are lead to the following optimization problem:

$$\hat{f} = \arg \min_f \Upsilon_2(f) \quad \text{s. t.} \quad y = \mathcal{R}f \quad \text{and} \quad My = y_0 \quad (11)$$

with $\Upsilon_2(f) = \left\| I^{-\frac{1}{2}} \mathcal{R}f \right\|_2^2$.

3. DISCRETE FRAMEWORK AND OBJECTIVE FUNCTION

We now show how to apply the theoretical observations of Section 2 to make the proposed approach a computationally suitable algorithm for the ROI CT problem.

In the following, we will refer to the Radon-transform domain as the *projection domain* since our approach applies to different discrete projection geometries, including parallel and fan beam in 2D.

To state the discrete framework, we discretize equations (5) and (6). We shall denote by K the number of projection angles and by P the number of detector cells (*i.e.*, samples along the detector array); N is both the width and the height in pixels of the object to reconstruct. Given a projection geometry, the matrix \mathbf{W} of the forward problem, that represents the map from the image domain to the projection domain, has dimensions $KP \times N^2$. The mask corresponding to the ROI is a diagonal matrix \mathbf{M} of dimensions $KP \times KP$ whose entries are either 0 or 1. The unknown discrete density function \mathbf{f} to be reconstructed is represented as a vector of length N^2 in which the entries are stacked column by column. Similarly, the full sinogram \mathbf{y} and the truncated sinogram \mathbf{y}_0 are represented as vectors of length KP , obtained by stacking the entries column by column. We recall that \mathbf{y} and \mathbf{y}_0 are related to each other by the data fidelity equation

$$\mathbf{y}_0 = \mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{W}\mathbf{f},$$

and the data consistency equation, that sets the extrapolation scheme outside the ROI, reads as

$$(\mathbf{I}_{KP} - \mathbf{M})\mathbf{W}\mathbf{f} = (\mathbf{I}_{KP} - \mathbf{M})\mathbf{y},$$

where \mathbf{I}_{KP} is the $KP \times KP$ identity matrix.

3.1 Explicit and implicit formulation of the objective function

As in the continuous case, data fidelity and data consistency equations need to be coupled with regularization to yield a unique solution. Thus, similarly to (11), we obtain the following discrete optimization problem:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \|\Phi\mathbf{W}\mathbf{f}\|_2^2 \quad \text{s. t.} \quad \mathbf{y} = \mathbf{W}\mathbf{f} \quad \text{and} \quad \mathbf{M}\mathbf{y} = \mathbf{y}_0 \quad (12)$$

where Φ is a discrete filter corresponding to the Riesz potencial operator $I^{-\frac{1}{2}}$.

However, rather than requiring the exact equalities stated by data fidelity and data consistency equations, we shall minimize the L^2 -norm error associated to them, according to a maximum likelihood approach. By exploiting this idea, we can state two different optimization problems and objective functions. On the one hand, we can consider the L^2 -norm error of the data fidelity equation only and incorporate the data consistency information in the regularization term. In this case, the only variable we are minimizing on is the image \mathbf{f} to be reconstructed. This approach yields:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f} \in \Omega_{\mathbf{f}}} \Psi(\mathbf{f})$$

where

$$\Psi(\mathbf{f}) = \frac{1}{2} \|\mathbf{M}\mathbf{W}\mathbf{f} - \mathbf{y}_0\|_2^2 + \lambda \|\Phi((\mathbf{I}_{KP} - \mathbf{M})\mathbf{W}\mathbf{f} + \mathbf{y}_0)\|_2^2. \quad (13)$$

We will refer to (13) as the *implicit formulation*, since the full sinogram \mathbf{y} does not appear explicitly.

On the other hand, we can consider both the data fidelity and the data consistency L^2 -norm errors into the objective function to assess if, by explicitly incorporating the extrapolation scheme in the minimization problem,

we achieve a more accurate reconstruction. In this case, the image \mathbf{f} to be reconstructed and the full sinogram \mathbf{y} are both unknowns. This second approach yields:

$$(\hat{\mathbf{f}}, \hat{\mathbf{y}}) = \arg \min_{\substack{\mathbf{f} \in \Omega_{\mathbf{f}} \\ \mathbf{y} \geq \mathbf{0}}} \Psi(\mathbf{f}, \mathbf{y})$$

where

$$\Psi(\mathbf{f}, \mathbf{y}) = \frac{1}{2} \|\mathbf{M}\mathbf{W}\mathbf{f} - \mathbf{y}_0\|_2^2 + \frac{1}{2} \|(\mathbf{I}_{KP} - \mathbf{M})(\mathbf{W}\mathbf{f} - \mathbf{y})\|_2^2 + \lambda \|\Phi((\mathbf{I}_{KP} - \mathbf{M})\mathbf{y} + \mathbf{y}_0)\|_2^2. \quad (14)$$

In comparison to the previous formulation, we will refer to (14) as the *explicit formulation*. In both cases, λ denotes the regularization parameter and the feasible region $\Omega_{\mathbf{f}}$ is either defined as $\mathbf{f} \geq 0$ or $0 \leq \mathbf{f} \leq L$, where L is the maximum pixel value of the image and the inequalities are meant component-wise. We also notice that each term of the objective function is convex with respect to the unknowns.

Finally, we add a slightly modified version of both the objective functions, which incorporate a smoothed total variation (TV) term, because they are used as a comparison in the experimental phase. These two additional objective functions read as

$$\Psi(\mathbf{f}) = \frac{1}{2} \|\mathbf{M}\mathbf{W}\mathbf{f} - \mathbf{y}_0\|_2^2 + \lambda \|\Phi((\mathbf{I}_{KP} - \mathbf{M})\mathbf{W}\mathbf{f} + \mathbf{y}_0)\|_2^2 + \rho \text{TV}_{\delta}(\mathbf{f}) \quad (15)$$

for the implicit formulation and as

$$\Psi(\mathbf{f}, \mathbf{y}) = \frac{1}{2} \|\mathbf{M}\mathbf{W}\mathbf{f} - \mathbf{y}_0\|_2^2 + \frac{1}{2} \|(\mathbf{I}_{KP} - \mathbf{M})(\mathbf{W}\mathbf{f} - \mathbf{y})\|_2^2 + \lambda \|\Phi((\mathbf{I}_{KP} - \mathbf{M})\mathbf{y} + \mathbf{y}_0)\|_2^2 + \rho \text{TV}_{\delta}(\mathbf{f}) \quad (16)$$

for the explicit one. Here, ρ is a regularization parameter and δ is the TV smoothing parameter.

We remark that, in the absence of noise, the solution \mathbf{f} is the same for both the implicit and the explicit cases (up to a vector in the nullspace of \mathbf{W}). However, in the presence of noise, the explicit formulation is beneficial due to the fact that the norm $\|\mathbf{W}\mathbf{f} - \mathbf{y}\|_2^2$ is generally non-zero and can be controlled using an appropriate stopping criterion, typically depending on the noise variance.

4. EXPERIMENTAL RESULTS

In this section, we show how to solve the optimization problem defined in Section 3 by using the proposed SGP-based algorithm. The numerical experiments presented in Section 4.2 will show the effectiveness of this approach.

We start by specifying our choices for the construction of the matrix \mathbf{W} of the forward problem and for the discrete filter Φ .

We recall that \mathbf{W} is the $KP \times N^2$ matrix mapping the image domain into the projection domain. Every row of \mathbf{W} contains weights that relate the pixel intensities in the image domain to the corresponding sample in the projection domain. Obviously, the value of each weight depends on the interpolation scheme chosen. Well known approaches are, for instance, the pixel-driven method, typically used in FBP schemes, and the ray-driven method, used in ray-tracers. In this paper, we will consider another state-of-the-art approach: the distance-driven method.¹³ This method combines a highly sequential memory access pattern with relatively low arithmetic complexity, without introducing interpolations artifacts in the image or projection domains. Essentially, this approach is based on converting the projection problem into a 1D re-sampling problem. There are two main ingredients in the distance-driven method. The first one is the *kernel operation*:

$$b_n = \sum_j w_j c_j \quad \text{with} \quad w_j = \frac{[\min(\xi_{m+1}, v_{n+1}) - \max(\xi_m, v_n)]_+}{v_{n+1} - v_n}, \quad [x]_+ = \max(x, 0), \quad (17)$$

which allows one to compute the destination signal values $\{b_j\}_j$ from the sample values $\{c_i\}_i$ of a source signal, the sample source locations $\{\xi_i\}_i$ and the sample destination locations $\{v_j\}_j$. The second element of the method is that there is a (possibly zero) length of overlap between each image pixel and each detector cell due to the

bijection between the position on the detector and the position within an image row (or column). Thus, every point within an image row is uniquely mapped onto a point on the detector, and vice versa. In practice, to compute the overlap length, all pixel boundaries in an image row and all detector cell boundaries are mapped onto a common line, *e.g.*, a line parallel to a coordinate axis. That is, each length of overlap is the interval length between two adjacent intersections, obtained by connecting the boundaries midpoints of all detector cells and pixels in a image row to the X-ray source and by computing the intercepts of these lines with the common axis. The final weights are achieved by normalizing the overlap length by the detector cell width. This corresponds exactly to apply the kernel operation from equation (17). In our case, b_n is the theoretical (unblurred and noiseless) value measured at the n -th detector cell and c_j is the estimate of the j -th pixel attenuation function. For example, if the n -th detector cell is “shadowed” by only two pixels in a row, equation (17) reads as:

$$b_n = \frac{\xi_{m+1} - v_n}{v_{n+1} - v_n} c_m + \frac{v_{n+1} - \xi_{m+1}}{v_{n+1} - v_n} c_{m+1}.$$

For the matrix Φ , we will not consider a discrete matrix approximation of the Riesz potential operator. Instead, we will simply approximate it by using a one-level shearlet decomposition.¹¹ Recall that the shearlet decomposition is a multiscale method which refines the conventional wavelet framework by combining multiscale analysis and directional sensitivity.^{14,15} Thanks to this property, shearlets provide optimally sparse approximations with images containing C^2 -edges, outperforming conventional wavelets.¹⁴ This is particularly relevant in CT-like applications, because point-like structures in the image domain map onto sine-shaped curvilinear structures in the projection domain. For comparison, we also considered a one-level undecimated Daubechies 4 wavelet transform. Notice that, since both classical shearlets and Daubechies wavelets are tight frame, the equivalence $\Phi^T \Phi = \mathbf{I}_{KP}$ holds true and this improves the convergence of the algorithm due to simpler updating rules.

4.1 Scaled gradient projection method

In this section, we describe the algorithm we propose for the solution of the optimization problems (15)-(16), whose step-by-step description is given in Algorithm 1.

The proposed method, called *scaled gradient projection* (SGP) method,¹² is an iterative approach from the family of first-order descent methods that apply to convex (and non-convex), differentiable and constrained problems with “simple” feasible regions. This is indeed the case of our formulation, given that the feasible region $\Omega_{\mathbf{f}}$ is usually either a box or a non-negativity constraint and, as already indicated, our functionals are convex and differentiable.

Before presenting and discussing the numerical experiments, it is useful to make some remarks about the main steps of SGP.

When the objective function reads as in (15), the $(k + 1)$ -th iteration, with $k = 0, 1, 2, \dots$, is

$$\mathbf{f}^{(k+1)} = (1 - \lambda_k) \mathbf{f}^{(k)} + \lambda_k \mathcal{P}_{\Omega_{\mathbf{f}}} \left(\mathbf{f}^{(k)} - \alpha_k D_k \nabla \Psi(\mathbf{f}^{(k)}) \right) \quad (18)$$

where λ_k , α_k are suitable steplengths, D_k is the scaling matrix and $\mathcal{P}_{\Omega_{\mathbf{f}}}$ is the projector onto the feasible region. The main feature of this method consists in the combination of non-expensive diagonally scaled gradient directions with steplength selection rules specially designed for these directions. In details, any choice of the steplength α_k in a closed interval $[\alpha_{\min}, \alpha_{\max}] \subset \mathbb{R}^+$ and of the scaling matrix D_k in the compact set \mathcal{D}_L is allowed, where \mathcal{D}_L is the set of the symmetric positive definite matrices D such that $\|D\| \leq L$ and $\|D^{-1}\| \leq L$, for a given threshold $L > 1$. This is very important from a practical point of view because it makes their updating rules problem-related and performance-aware. In particular, SGP is equipped with an adaptive steplength selection based on the Barzilai-Borwein (BB) updating rules.^{16,17} In practice, by means of a variable threshold, one of the two different selection strategies

$$\alpha_k^{\text{BB1}} = \arg \min_{\alpha_k \in \mathbb{R}} \|B(\alpha_k) \mathbf{s}^{(k-1)} - \zeta^{(k-1)}\| \quad \text{and} \quad \alpha_k^{\text{BB2}} = \arg \min_{\alpha_k \in \mathbb{R}} \|\mathbf{s}^{(k-1)} - B(\alpha_k)^{-1} \zeta^{(k-1)}\|$$

Algorithm 1 Scaled Gradient Projection Method

Choose the starting point $\mathbf{f}^{(0)} \in \Omega_{\mathbf{f}}$, set the parameters $\beta, \gamma \in (0, 1)$ and $0 < \alpha_{\min} < \alpha_{\max}$. Fix a positive integer μ .

for $k = 0, 1, 2, \dots$ **do**

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

Step 2. Projection: $\mathbf{z}^{(k)} = \mathcal{P}_{\Omega_f}(\mathbf{f}^{(k)} - \alpha_k D_k \nabla \Psi(\mathbf{f}^{(k)}))$.

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

Step 3. Descent direction: $\mathbf{d}^{(k)} = \mathbf{z}^{(k)} - \mathbf{f}^{(k)}$;

Step 4. Set $\lambda_k = 1$ and $\Psi_{\max} = \max_{0 \leq j \leq \min(k, \mu-1)} \Psi(\mathbf{f}^{(k-j)})$;

Step 5. Backtracking loop:

if $\Psi(\mathbf{f}^{(k)} + \lambda_k \mathbf{d}^{(k)}) \leq \Psi_{\max} + \beta \lambda_k \nabla \Psi(\mathbf{f}^{(k)})^T \mathbf{d}^{(k)}$ **then**

go to Step 6;

else

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

end if

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

end for

is selected,¹⁸ where the matrix $B(\alpha_k) = (\alpha_k D_k)^{-1}$ is an approximation of the Hessian matrix $\nabla^2 \Psi(\mathbf{f}^{(k)})$, $\mathbf{s}^{(k-1)} = \mathbf{f}^{(k)} - \mathbf{f}^{(k-1)}$ and $\zeta^{(k-1)} = \nabla \Psi(\mathbf{f}^{(k)}) - \nabla \Psi(\mathbf{f}^{(k-1)})$. As far as the scaling matrix concerns, the updating rule for each entry $d_i^{(k)}$ is:

$$d_i^{(k)} = \min \left\{ \sigma, \max \left\{ \frac{1}{\sigma}, \mathbf{f}_i^{(k)} \right\} \right\} \quad i = 1, \dots, KP \quad (19)$$

where σ is an appropriate threshold.

We can extend the SGP method to the case of two unknowns, as in equation (16), by applying the SGP iteration to each variable separately. In particular, we apply the algorithm to the image \mathbf{f} to be reconstructed as described above, while in the iteration on the full sinogram \mathbf{y} we take $D_k = \mathbf{I}_{KP}$ for each $k = 0, 1, 2, \dots$. That is, the $(k+1)$ -th iteration, with $k = 0, 1, 2, \dots$, is

$$\mathbf{f}^{(k+1)} = (1 - \lambda_k) \mathbf{f}^{(k)} + \lambda_k \mathcal{P}_{\Omega_f} \left(\mathbf{f}^{(k)} - \alpha_k D_k \nabla_{\mathbf{f}} \Psi(\mathbf{f}^{(k)}, \mathbf{y}^{(k)}) \right) \quad (20)$$

$$\mathbf{y}^{(k+1)} = (1 - \lambda_k) \mathbf{y}^{(k)} + \lambda_k \mathcal{P}_{\mathbf{y} \geq 0} \left(\mathbf{y}^{(k)} - \alpha_k \nabla_{\mathbf{y}} \Psi(\mathbf{f}^{(k)}, \mathbf{y}^{(k)}) \right) \quad (21)$$

We notice also that global convergence properties are ensured by exploiting a nonmonotone line-search strategy along the feasible direction.^{19,20} Indeed, the nonmonotone line-search strategy, as implemented in step 5, ensures that $\Psi(\mathbf{f}^{(k+1)})$ (resp., $\Psi(\mathbf{f}^{(k+1)}, \mathbf{y}^{(k+1)})$) is lower than the maximum of the objective function on the last μ iterations; of course, if $\mu = 1$ then the strategy reduces to the standard monotone Armijo rule.

Finally, the SGP algorithm can be used not only as a minimization method for a regularized objective functional, but also as an iterative regularization method applied to the un-regularized functional by means of an early stopping technique.

4.2 Numerical experiments

In this section, we present preliminary numerical results for the ROI CT reconstruction problem in the framework of 2D fan beam geometry.

To demonstrate and validate our approach, we used the synthetic data set known as “modified Shepp-Logan phantom” available, for instance, in the Matlab Image Processing toolbox. The algorithms were implemented in Matlab 8.1.0 and the experiments performed on a dual CPU server, equipped with two 6-cores Intel Xeon X5690 at 3.46GHz, 188 GB DDR3 central RAM memory and up to 12 TB of disk storage.

The object to be imaged is the modified Shepp-Logan phantom sized $N \times N$ pixels with $N = 128$. We assume it to be placed in the first quadrant of the Cartesian coordinate system with the image lower left corner

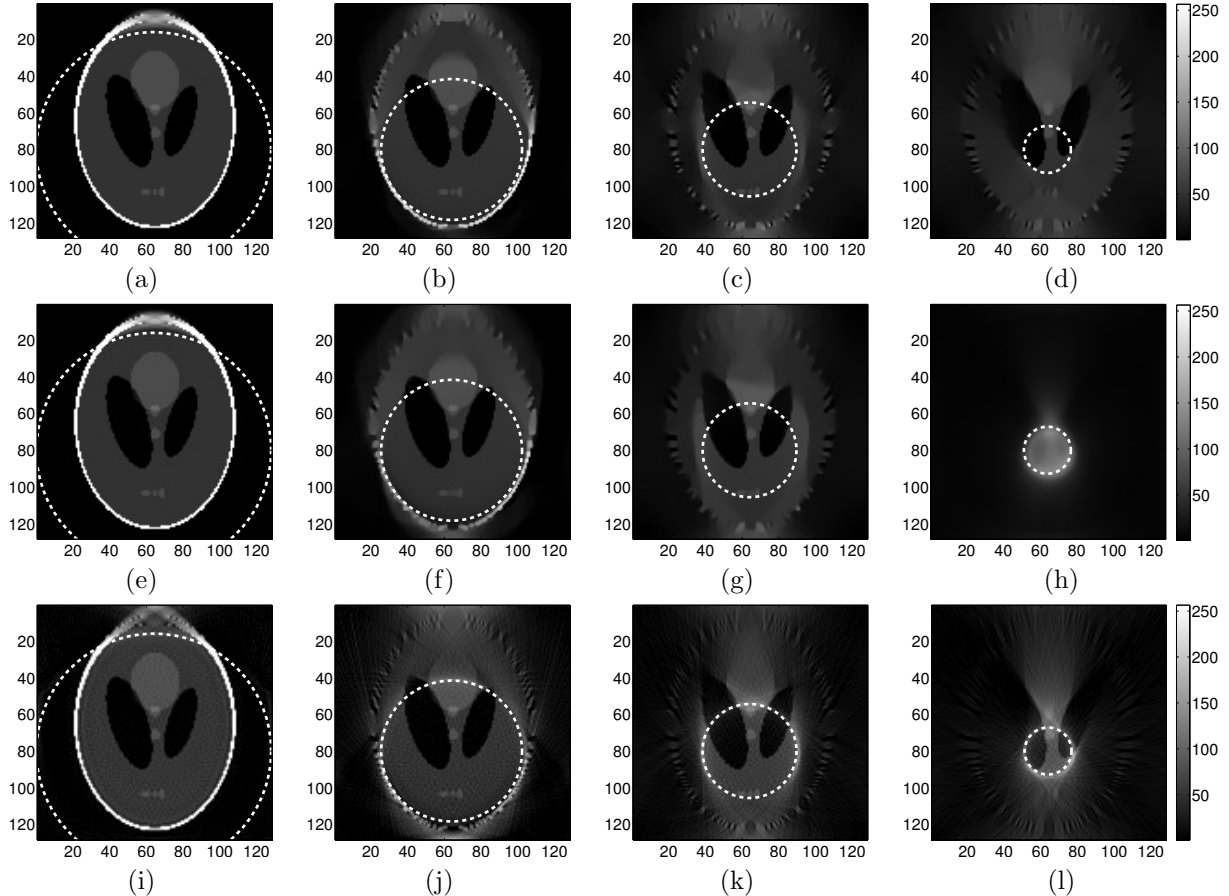


Figure 1. Optimal reconstructions of the Shepp-Logan phantom for decreasing *radii*: $\gamma = 0.5N$ for (a) and (e), $\gamma = 0.3N$ for (b) and (f), $\gamma = 0.2N$ for (c) and (g), $\gamma = 0.1N$ for (d) and (h). First row: implicit formulation. Second row: explicit formulation. Third row: LS-CG reconstruction.

at the origin. A full angle scan (182 views over 2π) with complete data was simulated and truncated projection data were obtained by discarding the samples outside the ROI projection $\mathcal{P}(S)$. We recall that this corresponds to a ROI disk in the image domain (see Section 2.1). In particular, we considered concentric ROI disks with decreasing *radius* γ , placed off-center with respect to the field of view. The results in this paper covers ROIs that are fully inside the object being imaged (in the hypothesis of the interior tomography problem). However, a larger *radius* that exceeds the object along one coordinate axis has been considered for comparison.

Radius	iter.	PSNR	Rel. Err.	Obj. Fun and Parameters
0.5	1277	51.35	0.013	$d(f) + r_1(y) + r_2(f)$ $\lambda = 5 \cdot 10^{-4}$ $\rho = 0.01$
0.3	2526	48.17	0.042	$d(f) + r_2(f)$ $\rho = 0.1$
0.2	1054	48.33	0.071	$d(f) + r_2(f)$ $\rho = 0.1$
0.1	2770	57.40	0.061	$d(f) + r_2(f)$ $\rho = 0.1$

Table 1. Optimal results for the *implicit* formulation of the objective function. Here, $d(f) = 0.5\|MWf - y_0\|_2^2$ is the discrepancy term and $r_1(y) = \|\Phi((\mathbf{I}_{KP} - M)y + y_0)\|_2^2$ and $r_2(f) = \text{TV}_\delta(f)$ are the regularization and the TV terms, respectively.

Radius	iter.	PSNR	Rel. Err.	Obj. Fun and Parameters
0.5	1642	49.59	0.016	$d_1(f) + d_2(f, y) + r_2(f)$ $\rho = 0.01$
0.3	2933	41.51	0.090	$d_1(f) + d_2(f, y) + r_2(f)$ $\rho = 1$
0.2	1910	44.90	0.106	$d_1(f) + d_2(f, y) + r_2(f)$ $\rho = 1$
0.1	2	35.97	0.718	$d_1(f) + d_2(f, y)$

Table 2. Optimal results for the *explicit* formulation of the objective function. Here, $d_1(f) = 0.5\|MWf - y_0\|_2^2$ and $d_2(f, y) = 0.5\|(\mathbf{I}_{KP} - M)(Wf - y)\|_2^2$ are the discrepancy terms and $r_2(f) = \text{TV}_\delta(f)$ is the TV term.

To validate the proposed approach with respect to both the objective function and the regularization parameters we performed extensive numerical tests.

We investigated the performance of the algorithm for the regularization parameter $\lambda = 5 \cdot 10^\ell$, $\ell = -4, -3, \dots, 1$; for the TV parameter ρ we sampled the values 10^{-2} , 10^{-1} and 1. We considered both the explicit and the implicit formulations of the objective function by designing different versions of the SGP method and exploiting the early stopping technique.

We report in Tables 1 and 2 the best results for the implicit and the explicit formulations, respectively. As a figure of merit, we use the peak-signal-to-noise ratio (PSNR) and the relative error. We recall that the PSNR, measured in dB, is defined as follows:

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MPV}^2}{\mathbf{e}_{\text{MSE}}} \right)$$

where MPV is the maximum pixel value and \mathbf{e}_{MSE} is the mean squared error. We stress that both PSNR and relative error are evaluated inside the ROI S only. This is consistent with the motivation of ROI CT, that only aims to recover the image only inside the ROI.

In Figure 1 we report the reconstruction obtained for the implicit (upper row) and the explicit (lower row) formulations, for decreasing *radii*. In those pictures, the ROI is identified by a dashed white circle. Due to space constraint, we only report results obtained using the shealets-based approach and omit the corresponding wavelet-based results. In all experiments we found the former to outperform the latter on a visual basis, even if the figures of merit are comparable. Referring to the first row of Figure 1, we observe that, when the *radius* is as large as half the number of pixels of the image edge (subfigure (a)), the best approach for the objective function is to combine the Tikhonov-like regularization with the TV term, as in equation (15). Notice that the values for the parameters λ and ρ are quite small since they are the smallest ones of the investigated ranges. For smaller *radii* (subfigures (b)-(d)), the best reconstruction is obtained by dropping the Tikhonov-like regularization term and by setting to a medium value the TV regularization parameter ρ . Notice that, when the *radius* decreases, the relative error has the same magnitude and the optimal reconstruction is achieved for the same formulation of the objective function and for the same value of the regularization parameter.

Referring to the second row of Figure 1, we observe that in this case the best approach for all *radii*, except for the smallest one, is obtained by dropping the Tikhonov-like regularization term. However, the largest radius (subfigure (e)) needs a small value for the regularization parameter ($\rho = 0.01$) while the middle ones (subfigure (f)-(g)) ask for a more severe TV regularization with $\rho = 1$. Notice that for *radius* equal to 20% of the number of pixels of the image edge (subfigure (g)), the error magnitude increases and for the smallest *radius* (subfigure (h)) the best approach is achieved by exploiting the early stopping technique. However, in this case, the reconstruction is not as good as in the previous cases, even if the fundamental structures are recognized after only two iterations.

For baseline comparison, we report in the third row of Figure 1 the reconstruction obtained with the Least Square Conjugate Gradient method (LS-CG).²¹ The LS-CG method minimizes $\|\mathbf{y}_0 - \mathbf{M}\mathbf{W}\mathbf{f}\|^2$ with respect to \mathbf{f} and hence does not include any regularization. The Conjugate Gradient method is then used to solve the sparse system of equations resulting from minimizing the least squares norm. Notice that the PSNR values for

the LS-CG method are considerably lower than the ones obtained with the SGP approach. For Radius = 0.5, 0.3, 0.2 and 0.1, we find PSNR = 27.84 dB, 25.46 dB, 27.16 dB and 22.98 dB, respectively. We believe that the lower performance is mainly due to the lack of regularization, especially the lack of a TV term, which promotes solutions that are piecewise constant.

5. CONCLUSION

This paper presents the numerical assessment on the solution of the ROI CT problem via a first-order iterative minimization method. Two different objective functions are considered, exploiting different levels of regularization and aiming to make the reconstruction from truncated data stable. The experiments show that, using noiseless projection data, the purely TV-based approach performs somehow better than the shearlets+TV approach for the cases with a small *radius*. Nevertheless, this is probably strongly dependent on the phantom features and may not hold for more general data. In future work, we will consider more realistic sinograms, also in the presence of noise, where the contribution of the shearlet term is expected to be relevant for the regularization. Overall, the reported results show that accurate ROI reconstructions can be obtained regardless of the location and size of the ROI and for rather small ROI sizes using both formulations of the objective function.

ACKNOWLEDGMENTS

T.A.B., S.B. and G.Z. are supported by the Italian national research project FIRB2012, grant n. RBFR12M3AC and by the local research project FAR2014 of the University of Ferrara. T.A.B. is supported by the Young Researchers Fellowship 2014 of the University of Ferrara and by the INdAM-GNCS Scholarship 2015. D.L. acknowledges partial support of NSF-DMS 1320910. B.G. acknowledges support by a postdoctoral fellowship of the Research Foundation - Flanders (FWO, Belgium).

REFERENCES

1. Natterer, F., [*The Mathematics of Computerized Tomography*], SIAM: Society for Industrial and Applied Mathematics (2001).
2. Natterer, F. and Wubbeling, F., [*Mathematical Methods in Image Reconstruction*], SIAM: Society for Industrial and Applied Mathematics (2001).
3. Clackdoyle, R. and Defrise, M., "Tomographic reconstruction in the 21st century. region-of-interest reconstruction from incomplete data," *IEEE Signal Processing* **60**, 60–80 (2010).
4. Noo, F., Defrise, M., Clackdoyle, R., and Kudo, H., "Image reconstruction from fan-beam projections on less than a short scan," *Physics in Medicine and Biology* **47**(14), 2525–2546 (2002).
5. R.Clackdoyle and Noo, F., "A large class of inversion formulae for the 2-d radon transform of functions of compact support," *Inverse Problems* **20**, 1281–1291 (2004).
6. Zou, Y., Pan, X., and Sidky, E., "Image reconstruction in regions-of-interest from truncated projections in a reduced fan-beam scan," *Phys. Med. Biol.* **50**, 13–28 (2005).
7. Hamelin, B., Goussard, Y., Dussault, J., Cloutier, G., Beaudoin, G., and Soulez, G., "Design of iterative roi transmission tomography reconstruction procedures and image quality analysis," *Medical Physics* **37**(9), 4577–4589 (2010).
8. Yang, J., Yu, H., Jiang, M., and Wang, G., "High-order total variation minimization for interior tomography," *Inverse Problems* **26**(3), 035013 (2010).
9. Ziegler, A., Nielsen, T., and Grass, M., "Iterative reconstruction of a region of interest for transmission tomography," *Medical Physics* **35**(4), 1317–1327 (2008).
10. Goossens, B., Labate, D., and Bodmann, B., "Region-of-interest computed tomography by regularity-inducing convex optimization," *submitted* (2014).
11. Easley, G. R., Labate, D., and Lim, W., "Sparse directional image representations using the discrete shearlet transform," *Appl. Comput. Harmon. Anal.* **25**, 25–46 (2008).
12. Bonettini, S., Zanella, R., and Zanni, L., "A scaled gradient projection method for constrained image deblurring," *Inverse Problems* **25**, 015002 (2009).

13. De Man, B. and Basu, S., “Distance-driven projection and backprojection in three dimensions,” *Physics in Medicine and Biology* **7**, 2463–2475 (2004).
14. Guo, K. and Labate, D., “Optimally sparse multidimensional representation using shearlets,” *SIAM J. Math. Analysis* **39**(1), 298–318 (2007).
15. Kutyniok, G. and Labate, D., [*Shearlets: Multiscale Analysis for Multivariate Data*], Springer (2012).
16. Barzilai, J. and Borwein, J. M., “Two point step size gradient methods,” *IMA J. Numer. Anal.* **8**, 141–148 (1988).
17. Fletcher, R., “On the barzilai-borwein method,” Tech. Rep. NA/207, Technical Report, Department of Mathematics, University of Dundee, Dundee, UK (2001).
18. Frassoldati, G., Zanghirati, G., and Zanni, L., “New adaptive stepsize selections in gradient methods,” *J. Industrial and Management Optim.* **4**, 299–312 (2008).
19. Grippo, L., Lampariello, F., and Lucidi, S., “A nonmonotone line-search technique for newtons method,” *SIAM J. Numer. Anal.* **23**, 707–716 (1986).
20. Dai, Y. H. and Fletcher, R., “On the asymptotic behaviour of some new gradient methods,” *Math. Programming* **103**, 541–559 (2005).
21. Hestenes, M. and Stiefel, E., “Methods of conjugate gradients for solving linear systems,” *Journal of Research of the National Bureau of Standards* **6**, 409–436 (1952).