A practical local tomography reconstruction algorithm based on a known sub-region

Pierre Paleo,[a,b]* Michel Desvignes[c] and Alessandro Mirone[a]*

[a]ESRF, 71 avenue des Martyrs, 38000 Grenoble, France, [b]Université de Grenoble, GIPSA-Lab, 11 Rue des Mathématiques, 38400 Saint-Martin-d'Hères, France, and [c]GIPSA-Lab, Grenoble Images Parole Signal Automatique, Institut Polytechnique de Grenoble, France. *Correspondence e-mail: pierre.paleo@esrf.fr, mirone@esrf.fr

A new method to reconstruct data acquired in a local tomography setup is proposed. This method uses an initial reconstruction and refines it by correcting the low-frequency artifacts, known as the cupping effect. A basis of Gaussian functions is used to correct the initial reconstruction. The coefficients of this basis are found by optimizing iteratively a fidelity term under the constraint of a known sub-region. Using a coarse basis reduces the degrees of freedom of the problem while actually correcting the cupping effect. Simulations show that the known region constraint yields an unbiased reconstruction, in accordance with uniqueness theorems stated in local tomography.

1. Introduction

In this section, we briefly recall the region of interest tomography problem and review the related work.

1.1. Region of interest tomography

Region of interest (ROI) tomography, also called local tomography, naturally arises when imaging objects that are too large for the detector field of view (FOV). It notably occurs in medical imaging, where only a small part of a body is imaged. Local tomography can also originate from a radiation dose concern in medical imaging.

Since the projection data do not cover the entire object, it is said to be *truncated* with respect to a scan that would cover the entire object. The aim is then to reconstruct the ROI from the 'truncated' data.

However, due to the nature of the tomography acquisition, the acquired data are not sufficient to reconstruct the ROI in general: for each angle, rays go through the entire object, not only the ROI. Thus, the data do not only contain information on the ROI but also contributions from the parts of the object external to the ROI. These contributions from the external parts actually preclude from reconstructing exactly the ROI from the acquired data in general.

The problem of reconstructing the interior of an object from truncated data is referred to as the *interior problem*. It is well known that the interior problem does not have a unique solution in general. If P denotes the projection operator, d the acquired data and x a solution of the problem $P(x) = d$, then x is defined up to a set of *ambiguity functions* u such that $P(x + u) = d$. An example is given by Clackdoyle & Defrise (2010), where u is non-zero in the ROI but $P(u) = 0$ in the detector zone corresponding to the ROI: two solutions differing by u would produce identical interior data. Wang
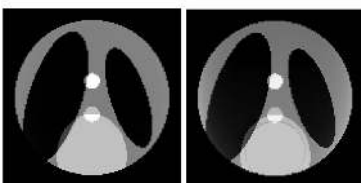
& Yu (2013) emphasize that the ambiguity is an *infinitely differentiable* function whose variation increases when moving outside the ROI. The non-uniqueness of the solution of the interior problem prevents quantitative analysis of the reconstructed slices.

Methods tackling the ROI tomography problem can mainly be classified into two categories. The first category methods aim at completing the data by extrapolating the sinogram. These are often oriented toward easy and practical use, although having no theoretical guarantees. The second category of methods rely on prior knowledge of the object. Many theoretical efforts have been made on these methods, providing for example uniqueness and stability results.

Other works use wavelets to localize the Radon transform (Rashid-Farrokhi *et al.*, 1997; Sastry & Das, 2005) or focus on the detection of discontinuities, the best known being probably Lambda-tomography (Yu & Wang, 2007).

1.2. Sinogram extrapolation methods

In a classical tomography acquisition, the whole object is imaged. If nothing is surrounding the object, the rays are not attenuated by the exterior of the object; thus the sinogram values for each angle go to zero on the left and right parts (after taking the negative logarithm of the normalized intensity). In a local tomography acquisition, however, the data are 'truncated' with respect to what would have been a whole scan.

The incompleteness of the data induces artifacts in the reconstructed image. The first obvious artifact is visible as a bright rim on the exterior of the image. This bright rim is the result of the abrupt transition in the truncated sinogram: the filtration process suffers from a Gibbs phenomenon. Another artifact is referred to as the *cupping effect*: an unwanted background appears in the reconstructed image, which makes further analysis such as segmentation challenging. These two artifacts occur simultaneously, but they have different causes. The bright rim comes from the truncation, while the cupping comes from the contribution of the external part.

Figs. 1 and 2 illustrate these artifacts. A synthetic slice is projected, and the resulting sinogram is truncated to simulate a ROI tomography setup. The filtering step enhances the transition between the ROI and the truncated part which is set to zero. The difference between the filtered whole sinogram and the filtered truncated sinogram also

shows the cupping effect, which appears as a low-frequency bias.

Sinogram extrapolation methods primarily aim at eliminating the bright rim resulting from the truncation by ensuring a smooth transition between the ROI and the external part. In addition, efforts have been put into estimating the missing data in order to reduce the cupping effect. These techniques are referred to as sinogram extrapolation methods: the external part is estimated from the truncated data with some extrapolating function.

The extrapolating function can be, for example, constant (the outermost left/right values are replicated), polynomial or $\cos^2$. Shuangren Zhao & Yang (2011) used a mixture of exponential and quadratic functions to estimate the external part, possibly iteratively. The projection of a circle, for which a closed-form formula is known, can also be used (Van Gompel, 2009). A common approach is using the values of the left/right part of the sinogram to estimate the external part, that is, replicating the borders values.

In general, sinogram extrapolation methods do not take into account the sinogram theoretical properties. For example, given an object being non-zero only inside a circle of a given radius, the sinogram decreases to zero at the left and right
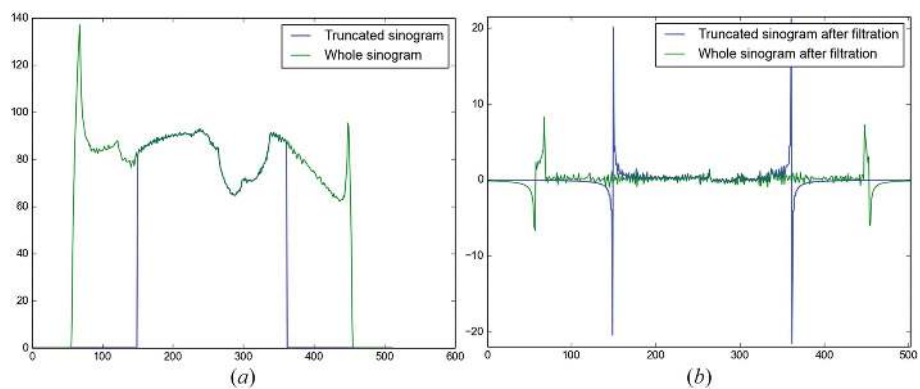
Figure 1
Illustration of the truncation artifacts on a line of the sinogram of the Shepp–Logan phantom. (*a*) Whole sinogram corresponding to a scan where all the object is imaged (green), and truncated sinogram (blue). (*b*) After ramp-filtering.
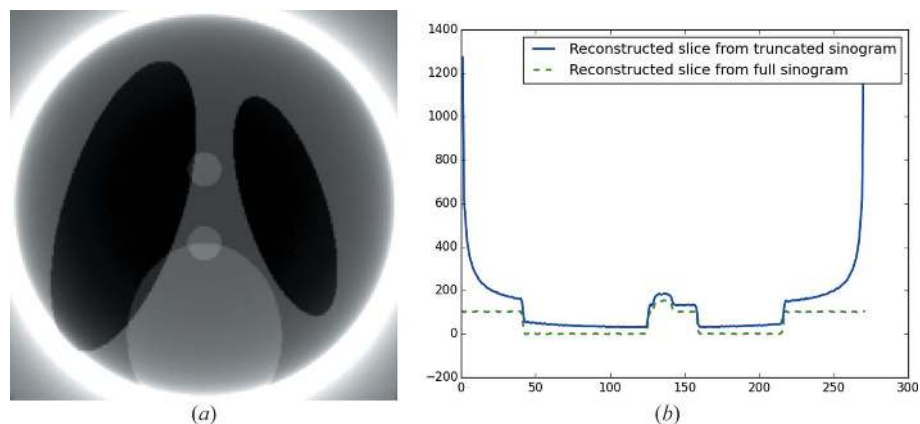
Figure 2
(*a*) Reconstruction of the truncated sinogram with filtered back-projection. The contrast has been modified to visualize the interior of the slice. (*b*) Line profile of the reconstruction.

boundaries. Generally speaking, a sinogram of complete measurements satisfies the Helgason–Ludwig consistency conditions (1) (Van Gompel, 2009):

$$H_n(\theta) = \int_{-\infty}^{\infty} s^n p(\theta, s)\, ds \qquad (1)$$

is a homogeneous polynomial of degree n in $\sin\theta$ and $\cos\theta$, for all $n \geq 0$. An alternative formulation is given by equation (2):

$$H_{n,k}(\theta) = \int_{0}^{\pi} \int_{-\infty}^{\infty} s^n \exp(jk\theta)\, p(\theta, s)\, ds\, d\theta = 0, \qquad (2)$$

for $k > n \geq 0$ and $k - n$ even. Van Gompel *et al.* (2006) used (2) as a quantitative measure of the sinogram consistency, and is optimized as an objective function.

For many applications, constant extrapolation provides acceptable results (Kyrieleis *et al.*, 2011), although the cupping artifact makes the segmentation challenging.

1.3. Prior-knowledge-based interior tomography

It was long believed that ROI tomography could be solved exactly, because of the nature of the Radon inversion through filtered back-projection (FBP): the reconstruction of each voxel requires knowledge of all the (complete) lines passing through this voxel.

However, in the last decade, it has been shown that multiple non-equivalent reconstruction formulas allow partial reconstruction from partial data in the two-dimensional case (Clackdoyle & Defrise, 2010). Alternatively to FBP reconstruction, which requires complete data, virtual fan beam (VFB) and differentiated back-projection (DBP) were developed based on the Hilbert projection equality (Clackdoyle *et al.*, 2009).

Moreover, uniqueness theorems based on analytical continuation of the Hilbert transform were stated and progressively refined (Noo *et al.*, 2002, 2004; Clackdoyle *et al.*, 2004; Zou *et al.*, 2005; Defrise *et al.*, 2006; Ye *et al.*, 2007; Kudo *et al.*, 2008; Tang *et al.*, 2012). These authors ensure an exact and stable reconstruction of the ROI given some assumptions. These assumptions can be of geometric nature, or in the form of prior knowledge.

Geometry-based prior knowledge is related to the acquisition geometry. For example, in DBP-based reconstruction, a point can be reconstructed if it lies on a line segment extending outside the object on both sides, and all lines crossing the segment are measured (Clackdoyle & Defrise, 2010). Similar results were obtained under less restrictive assumptions, for example the FOV extending the ROI on only one side (Ye *et al.*, 2007).

These geometry-based methods do not work, however, when the FOV does not extend the object. In this case, it has been shown (Courdurier *et al.*, 2008) that a prior knowledge of the function inside the ROI enables exact and stable reconstruction of the ROI. This knowledge can be in the form of the function values inside a sub-region of the ROI (Kudo *et al.*,

2008) or can be about the properties of the function to reconstruct, for example, sparsity in some domain.

This latest kind of knowledge has led to compressive sensing based ROI tomography. Yang *et al.* (2010), Yu & Wang (2009) and Lee *et al.* (2015) used the total variation method to reconstruct the ROI. Niinimäki *et al.* (2007) assumed the function to be sparse in the wavelet domain, and that a multi-resolution scheme reduces the number of unknowns by keeping only fine-scale wavelet coefficients inside the ROI. Klann *et al.* (2015) showed that piecewise constant functions are determined *everywhere* by their ROI data, the underlying hypothesis being formulated as sparsity in the Haar domain.

2. Low-frequencies artifacts correction with Gaussian blobs

Sinogram extrapolation usually copes well with the correction of discontinuities in the truncated sinogram, but does not correct the cupping effect in general. This cupping effect appears as a low-frequency bias in the reconstructed image. Sinogram extrapolation and other background correction techniques do not give guarantees that the low-frequency bias will actually be removed without distorting the reconstruction.

In this section, we describe a new method using prior knowledge of a sub-region of the reconstructed volume to eliminate the low-frequency cupping bias. The starting point of this method is an initial reconstruction, hereby denoted x_0, which can be obtained for example using the padded FBP method. This initial reconstruction is then refined with an additive correction term. This correction term uses the known sub-region as a constraint which should be sufficient, according to uniqueness theorems stated in the references given in §1.3, to accurately reconstruct the ROI.

As x_0 bears the high-frequencies features, the correction term is expressed as a linear combination of Gaussian functions to counterbalance the low-frequency artifacts. The representation in a basis of Gaussian blobs in tomography was first introduced by Lewitt (1992); we use it here for representing coarse features. The coefficients are optimized subject to knowledge of the sub-region, hereby denoted Ω. To constrain all the Gaussian coefficients by knowledge of the image values in Ω, a reduced set of coefficients is firstly computed inside Ω. Then, the Gaussian coefficients are iteratively optimized to fit the reconstruction error of the whole image, using the coefficients computed inside Ω as a constraint.

Let u_0 denote the 'true' object values in the known region Ω. The proposed method can be summarized as follows:

(i) The reconstruction error in Ω, denoted $e_{|\Omega} = [(x_0)_{|\Omega} - u_0]$, is expressed as a linear combination of two-dimensional Gaussians. The resulting Gaussian coefficients are denoted g_0.

(ii) The error in the whole image is iteratively fitted with Gaussians coefficients g, subject to $g_{|\Omega} = g_0$, to build a consistent reconstruction error in the whole image.

Details of each step are described in the following parts.

2.1. Capturing the low frequencies of the error in the known zone

The reconstruction error in Ω is first approximated as a linear combination of Gaussian functions. This function is chosen for computational convenience, more details are given in §2.2.3.

Equation (3) gives the expression of the approximation. The error estimation $\widehat{e_{|\Omega}}$ is a linear combination of translated Gaussians of weights $c_{i,j}$, with a spacing s,

$$\widehat{e_{|\Omega}} = \sum_{i,j} c_{i,j} \psi_\sigma(x - i \cdot s, y - j \cdot s),$$
$$\psi_\sigma(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \tag{3}$$

For simplicity, all the Gaussians have the same standard deviation σ and all have the same spacing s between them. Their location on the image grid is also fixed, so that the fit turns into a linear inverse problem:

$$g_0 = \underset{g}{\mathrm{argmin}}\left\{\frac{1}{2}\left\|Gg - \left[(x_0)_{|\Omega} - u_0\right]\right\|_2^2\right\}, \tag{4}$$

where G is the operator taking as an input the coefficients c_i, stacked in a vector g; and producing an image tiled with Gaussians (here in region Ω). The norm $\|\cdots\|_2^2$ is the squared Frobenius norm, that is, the sum of the squares of all components. Choosing the same standard deviation σ and the same spacing s for all Gaussians enables to implement G as a convolution. More precisely, given an image being zero everywhere, coefficients c_i are placed every s pixels in Ω and this image is convolved with two-dimensional ψ_σ.

The reconstruction error in Ω is thus estimated in the least-squares sense: g_0 is the vector of Gaussian coefficients giving the best estimation $\widehat{e_{|\Omega}}$ of the reconstruction error in the L2 sense. This vector g_0 will be used in the second part of the algorithm.

2.2. Correcting the reconstruction error in the image

2.2.1. Overview of the method.
Outside the known region Ω, the reconstruction error is not known. Like some other methods described in §1.3, this algorithm aims at using the known region information to accurately reconstruct the whole ROI. However, this approach focuses on *correcting* an initial reconstruction: the reconstruction error in Ω is fitted as a linear combination of Gaussians, then the whole image is corrected on a coarse Gaussian basis whose coefficients are constrained in the known sub-region.

The FBP with sinogram extrapolation is widely used in local tomography because it is both simple and gives satisfactory results in general (Kyrieleis *et al.*, 2011). Theoretical investigations found that FBP provides a reconstructed function bearing the same discontinuities as the reference function (Bilgot *et al.*, 2011), although the cupping effect can make the segmentation challenging. In this method, FBP with padding is used to obtain an initial estimate of the reconstruction; the aim is to correct the local tomography artifacts on this image using the prior knowledge. Equation (5) gives the expression of

the estimate x where x_0 is the initial reconstruction, G is the operator described in §2.1 and $\hat{g}$ is a linear combination of Gaussian functions aiming at counterbalancing the low-frequencies artifacts,

$$x = \tilde{x}_0 + G\hat{g}. \tag{5}$$

The vector $\hat{g}$ is found by minimizing an objective function which is built as follows. A new image $x = \tilde{x}_0 + Gg$, containing the initial reconstruction, is created. The image $\tilde{x}_0$ is an extension of the initial reconstruction x_0. This new image is projected with a projector P adapted to the bigger size. To compare with the acquired sinogram d, the computed sinogram $P(\tilde{x}_0 + Gg)$ is truncated by a cropping operator C. The data fidelity is then given by equation (6),

$$\frac{1}{2}\left\|CP(\tilde{x}_0 + Gg) - d\right\|_2^2. \tag{6}$$

We emphasize that this approach differs from the full estimation of the ROI based on a sub-region. The variables g are in a coarse basis while $\tilde{x}_0$ is fixed, which is notably reducing the degrees of freedom of the problem. The operation Gg has two goals: a coarse estimation of the exterior (outside the x_0 support) and a correction of the low-frequencies error inside the x_0 support.

As the minimization is on g, the initial estimate of the ROI x_0 is constant, and the data fidelity term (6) can be rewritten as in equation (7),

$$\frac{1}{2}\left\|CPGg - d_e\right\|_2^2, \tag{7}$$

where $d_e = d - \tilde{P}x_0$ is the difference between the acquired sinogram d and the projection of the initial reconstruction x_0, and $\tilde{P}$ is the projector adapted to the size of x_0. The optimization problem is given by equation (8),

$$\hat{g} = \underset{g}{\mathrm{argmin}}\left\{\frac{1}{2}\left\|CPGg - d_e\right\|_2^2 \quad \text{subject to} \quad g_{|\Omega_g} = g_0\right\}. \tag{8}$$

g_0 is the vector of Gaussian coefficients found in (4), such that Gg_0 approximates the error in the known zone. The set Ω_g denotes the subset of the Gaussian basis corresponding to Ω in the pixel basis: if a coefficient c_i of g lies in Ω_g in the Gaussian basis, then $(Gg)_i$ lies in Ω in the pixel basis. Equation (8) boils down to finding coefficients g minimizing the reconstruction error in the whole image, under the constraint that Gg should give the (known) reconstruction error in Ω. Alternatively, the coefficients g can be optimized with the constraint expressed in the *pixel* domain, yielding equation (9),

$$\hat{g} = \underset{g}{\mathrm{argmin}}\left\{\frac{1}{2}\left\|CPGg - d_e\right\|_2^2 \quad \text{subject to} \quad (Gg)_{|\Omega} = u_0\right\}. \tag{9}$$

The local constraint is propagated in all of the variables by the projection operator involved in the optimization process. Uniqueness theorems mentioned in §1.3 state that knowledge of a sub-region of the ROI is sufficient to yield an exact reconstruction. However, when using a pixel basis without space constraints, the number of degrees of freedom might be

too high; leading to a slow convergence. Using a coarse basis for correcting the low frequencies reduces this number of degrees of freedom.

2.2.2. Details of the involved operations. In this part, more details are given about the different steps of the algorithm. We start by computing a padded FBP reconstruction x_0 which gives an initial estimate of the ROI of size (N, N). This image is extended to a bigger image $\tilde{x_0}$ of size (N_2, N_2) where $N_2 > N$ and x_0 is placed in the center of the image.

At each iteration k, the image Gg_k, where g_k is the Gaussian coefficients vector at iteration k, is computed. The operator G consists of placing the coefficients g on a regular grid and convolving with the Gaussian kernel $(x, y) \mapsto (1/\sigma\sqrt{2\pi})\exp[-(x^2 + y^2)/2\sigma^2]$. Thus, the operator G can be written $G = C_\sigma U_s$ where C_σ is the convolution by the aforementioned Gaussian kernel of standard deviation σ, and U_s is an operator upsampling an image by a factor of s. As both are linear operators, G is a linear operator and $G^T = U_s^T C_\sigma^T$ where U_s^T is the s-downsampling operator and C_σ^T is a convolution by the matched Gaussian kernel, which is the same kernel due to symmetry. In our implementation, the Gaussian kernel has a size of $\lfloor 8\sigma + 1 \rfloor$, i.e. the Gaussian is truncated at 4σ. The resulting image is given by equation (10),

$$x(i_0, j_0) = (Gg)(i_0, j_0) = (C_\sigma U_s g)(i_0, j_0) = (C_\sigma z)(i_0, j_0)$$
$$= \sum_{i,j} z(i, j)\,\psi_\sigma(i_0 - i, j_0 - j), \qquad (10)$$

where ψ_σ is the discrete Gaussian kernel, z is the image containing the Gaussian coefficients g placed on the grid of size (N_2, N_2) after upsampling, that is, zeros almost everywhere except coefficients every s pixel. The summation in (10) is done on the convolution kernel support. If $s < 4\sigma$, the Gaussian functions supports can overlap once placed on the grid. In practice, these Gaussians should overlap to appropriately fit constant regions: for s close to σ, the Gaussians almost yield a partition of unity (Bale et al., 2002).

Once the coefficients are placed on a grid and convolved by the two-dimensional Gaussian function, the image x is projected. The projection operator adapted to the new geometry (the bigger image Gg_k) is denoted P. This is a standard Radon transform. This process is illustrated in Fig. 3.

As the new image x is bigger than x_0, the sinogram Px and the acquired data d cannot be directly compared. The computed sinogram Px is thus cropped to the region corresponding to the ROI. The cropping operator is denoted by C; it is also a linear operator whose transpose consists of extending the sinogram by inserting zeros on both sides.

The resulting sinogram aims at fitting the error between the acquired sinogram d and the (cropped) projection of the object. As the object is unknown except inside Ω, the reconstruction error is only known in Ω. The Gaussian coefficients g are constrained by those found by fitting the error inside Ω in §2.1. The projection operator involved in the process propagates the constraint to all the other coefficients.

Coefficients $\hat{g}$ from equation (8) are computed with an iterative solver. As this objective function is quadratic, efficient minimization algorithms like conjugate gradient can be

used. The final image is obtained with $\hat{x} = x_0 + G\hat{g}$ and is cropped to the ROI.

2.2.3. Computational aspects. Using Gaussians as functions to iteratively express the error has several computational advantages. Gradient-based algorithms for solving (8) involve the computation of the forward operator PG and its adjoint $G^T P^T$. They are usually the computationally expensive steps of iterative solvers. In this case, these operators can be computed in an efficient way.

The Gaussian kernel has an interesting property: it is the only (non-zero) one to be both rotationally invariant and separable (Kannappan & Sahoo, 1992). In our case, the convolution by a Gaussian followed by a projection (forward Radon transform) is equivalent to projecting first and convolving each line of the sinogram by the corresponding one-dimensional Gaussian, as illustrated in Fig. 3.

The first advantage is of a theoretical nature. In many implementations, the projector and backprojector pair are usually not adjoint of each other for performances reasons. Although giving satisfying results in most practical applications, this raises theoretical issues on convergence of algorithms using iteratively forward and backward operators (Zeng & Gullberg, 2000). By using a point-projector and a point-backprojector implementation, the pair can be exactly adjoint, besides giving more accurate results.

The second advantage is on the computational side. As the operator PG consists of projecting two-dimensional Gaussians disposed on the image, it is equivalent to placing one-pixel coefficients (Dirac functions in the continuous case) in the image on a grid denoted $\mathcal{I}$, projecting the image (with a point-projector) and convolving the sinogram by a one-dimensional Gaussian kernel. The same goes for the adjoint operator $G^T P^T$ consisting of retrieving the Gaussians coefficients from a sinogram. The standard way to compute this operator would
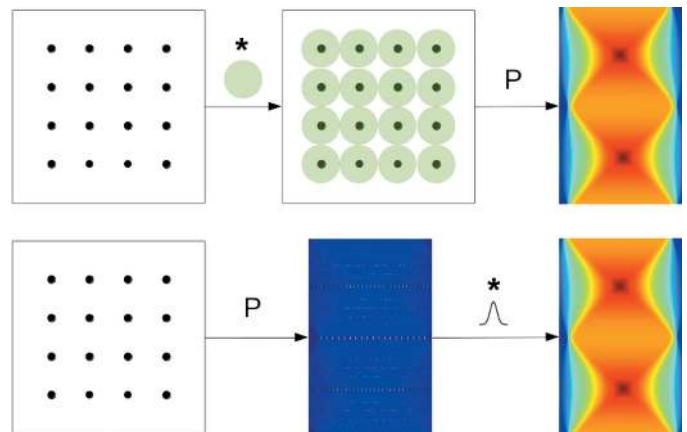
Figure 3
Description of the operator PG (first line). A grid of points is created; one coefficient is assigned to each point on the grid. The result is convolved by a two-dimensional Gaussian function (depicted as green circles), and projected to obtain a sinogram. An equivalent approach (second line) consists of first projecting the 'point coefficients' with an appropriate projector, and convolving each line of the sinogram by a one-dimensional Gaussian function.

be backprojecting the sinogram, convolving by the two-dimensional Gaussian (which is its own matched filter due to symmetry), and sampling the image on the grid $\mathcal{I}$ to obtain the coefficients. Here, the convolution can be first performed in one dimension along the sinogram lines. The sinogram is sampled at locations corresponding to points $\mathcal{I}$ in the image domain. The resulting sampled sinogram is then back-projected with a point-backprojector.

2.3. Pseudocode of the proposed method

In this section, the different steps of the proposed method are summarized in two algorithms. The first performs the fitting of the error in the known zone as described in §2.1, the second builds the resulting image as described in §2.2.

A complete implementation of the proposed method is available (Paleo, 2016). It contains comments on the different steps and can be tuned for various setups. This implementation relies on the *ASTRA* toolbox (Palenstijn *et al.*, 2011; van Aarle *et al.*, 2015); the point-projector scheme described in §2.2.3 is not implemented for readability, but this approach would be more suited to a production reconstruction algorithm where performances are an issue.

Algorithm 1. Known zone fitting.
d: acquired sinogram
Ω: location and size of the known zone
u_0: known values in the zone Ω

1: **procedure** FITKNOWN(d, Ω, u_0)
2:      $x_0 = \text{padded\_FBP}(d)$ ▷ Padded FBP of d
3:      $e_{|\Omega} = (x_0)_{|\Omega} - u_0$ ▷ Error in the known zone
4:      $g_0 = \underset{g}{\text{argmin}} \left\{ \left\| Gg - e_{|\Omega} \right\|_2^2 \right\}$ ▷ Fit the error with Gaussians
5:      $\tilde{x}_0 = \text{extend}(x_0)$ ▷ Extend to a bigger image
6:      **return** $\tilde{x}_0, g_0$
7: **end procedure**

The location of the known zone Ω can be simply implemented as a tuple of pixels (i_0, j_0) and a radius r for a circular zone.

Algorithm 2. Error correction
d: acquired sinogram
Ω: location and size of the known zone
u_0: known values in the zone Ω
N_2: size of the extended image
σ: standard deviation of the Gaussian functions
s: grid spacing

1: **procedure** LOCALTOMO$(d, \Omega, u_0, N_2, \sigma, s)$
2:      $\tilde{x}_0, g_0 = \text{FITKNOWN}(d, \Omega, u_0)$ ▷ Compute $\tilde{x}_0$ and g_0 with Algorithm 1
3:      $d_e = CP\tilde{x}_0 - d$ ▷ Difference between the cropped projection of x_0 and d
4:      $\hat{g} = \underset{g}{\text{argmin}} \left\{ \frac{1}{2} \| CPGg - d_e \|_2^2 \quad \text{s.t.} \quad g_{|\Omega_g} = g_0 \right\}$ ▷ Operators are described in
        2.2.2
5:      $\hat{x} = \tilde{x}_0 + G\hat{g}$
6:      **return** $\hat{x}$
7: **end procedure**

In practice, the final image $\hat{x}$ is cropped to the ROI. In Algorithm 2, the optimization (line 5) can be done with a gradient algorithm, as differentiating the quadratic error term requires only the operators and their adjoints. Another approach, expressing the known zone constraint in the pixel domain [equation (9)], only requires Algorithm 2 to be run.

3. Results and discussion

In this section, results and discussions on three test cases are presented. Synthetic sinograms are generated by projecting an object and truncating the sinogram to the radius of a given ROI in the image.

The following notations are used: σ is the standard deviation of the Gaussians of the basis, s is the grid spacing, N_2 is the size (width or height in pixels) of the extended image and R is the radius (in pixels) of the known region. In practice, the size of the 'original image' (which corresponds to the size of an image that would contain the whole object in practice) is unknown, hence N_2 is always chosen different from the width of the original test image.

In all cases, the input image is projected with a projector covering the entire object. The resulting synthetic sinogram is then truncated to the radius of the ROI. The truncated sinogram is the input of the methods. The proposed method is compared with the padded FBP. As the padded FBP is used as an initial reconstruction by the proposed method, the benchmark is mainly about checking that the cupping effect is actually removed, and that the correction does not induce distortion in the final image.

The first test involves the standard Shepp–Logan phantom (Fig. 4), 256×256 pixels. The ROI is embedded inside the 'absorbing outer material' (ellipse with the highest gray values) to simulate a local tomography acquisition. For an
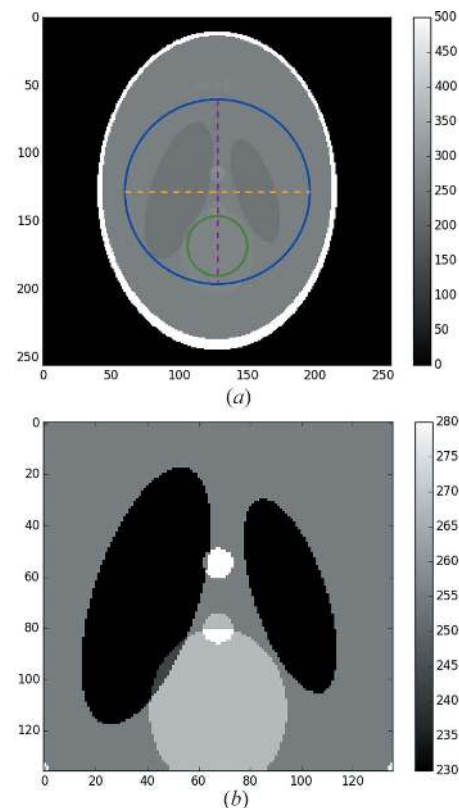

Figure 4
(*a*) Shepp–Logan phantom of size 256×256. The outer circle is the ROI, the inner circle is the known sub-region. The dashed lines indicate the profiles which are to be plotted in the reconstructed slice. (*b*) View of the ROI (with adapted contrast).
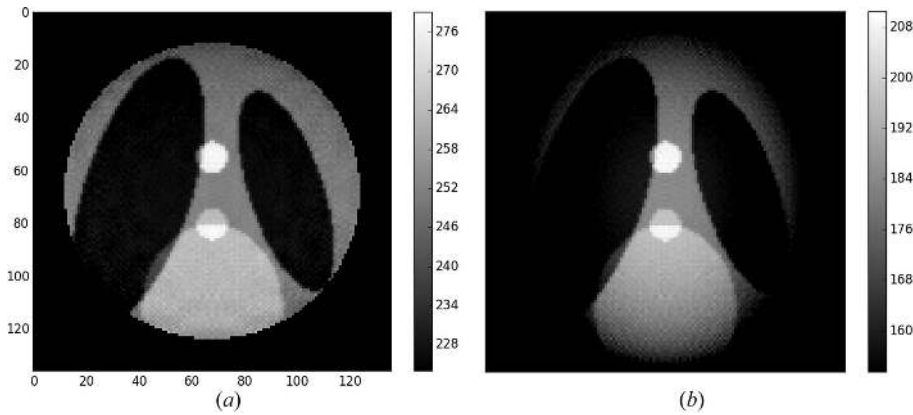
Figure 5
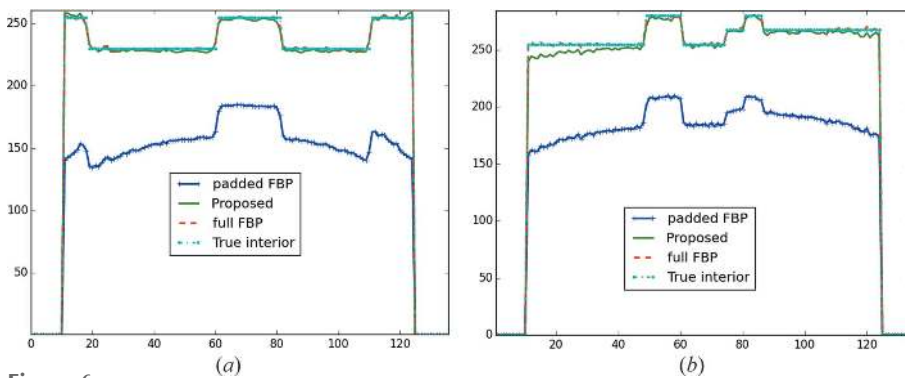Results of reconstructions with adapted contrast. (a) Proposed. (b) Padded FBP.

Figure 6
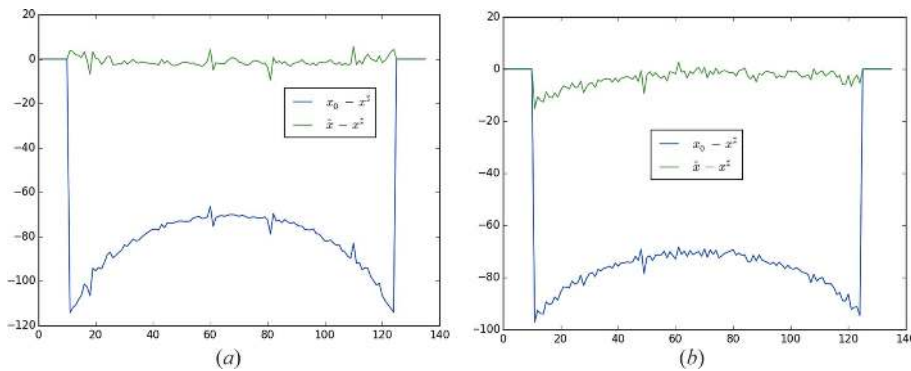Line profiles of reconstructions. (a) Middle line. (b) Middle column.

Figure 7
Difference between the reconstruction and the true volume $x^{\#}$. (a) Along the middle line, (b) along the middle column.

effect only affects few low-frequency components, and is thus less detrimental to the reconstruction quality.
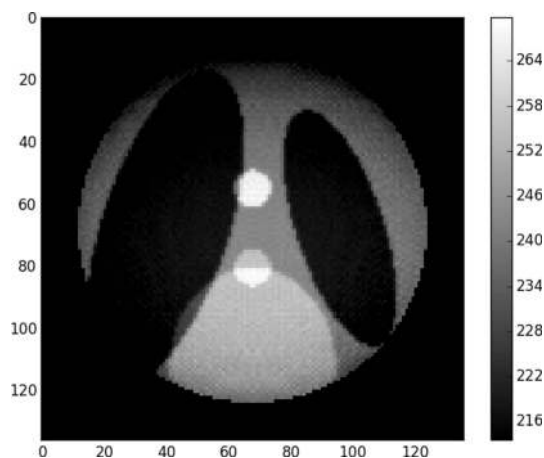
Fig. 5 shows the result of the reconstruction with padded FBP and with the proposed method. The Gaussian coefficients were computed with $\sigma = 4$ on a grid of spacing $s = 6$. The known region radius is $R = 20$ pixels, and the extended image width is $N_2 = 260$ pixels. On the padded FBP reconstruction, the cupping effect is relatively important due to the low contrast of the data.

By visual inspection, this method does not induce new artifacts in the reconstruction. Fig. 6 shows a line profile of this reconstruction. The cupping effect is visible for the padded FBP, and it has been removed using the proposed method. More importantly, the average reconstructed values are distributed around the true interior values. This provides an illustration of the uniqueness theorem: knowing the values of a sub-region of the ROI enables the ROI to be exactly reconstructed (up to numerical errors). The reconstruction with the proposed method bears the same high frequencies as the FBP with full data, which is a good indication that this method does not induce new artifacts. The fact that the reconstruction has the same mean values as the true interior could enable quantitative analysis of the reconstructed volume, which is not easily achievable in local tomography.

Fig. 7 shows the difference between the reconstructions and the interior values (denoted $x^{\#}$). As expected, the cupping effect is visible for padded FBP, while being almost entirely suppressed in the reconstruction with the proposed method.

The known zone constraint is important to remove the mean bias. As can be seen in Figs. 8 and 9, the cupping effect remains if no constraint is applied. In this case, the uniqueness theorem does not apply when there is no constraint, hence there are no guarantees that the method converges to an acceptable solution.

The second test involves the test image 'Lena', 512×512 pixels, bearing both smooth regions and high-frequencies textures. Ellipses with strong intensity values have been superimposed in the exterior of the ROI to simulate absorbing material.

easier interpretation of the line profiles in the final reconstructed images, the values of the standard phantom are multiplied by 250 so that all the values are between 0 and 500. The width of the extended for reconstruction image is $N_2 = 260$.

This phantom is the 'original' Shepp–Logan phantom, not the 'modified Shepp–Logan phantom' where the contrast is improved. In our case, low contrast is important for the tests, as the cupping effect is stronger and directly visible in the reconstructed slice. For high-contrast images, the cupping

Figure 8
Reconstruction of the Shepp–Logan phantom without known zone constraint.

Fig. 10 shows the test setup. The known region has been chosen as slowly varying as possible, as in real acquisitions the known region is likely to be air or coarse features. The width of the extended image is $N_2 = 520$.

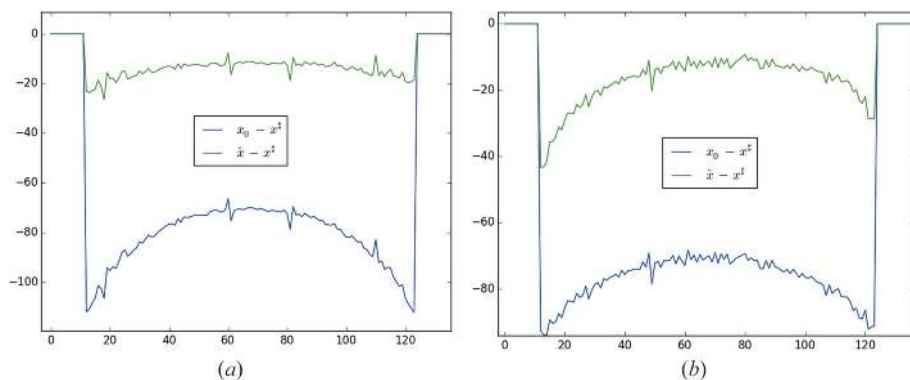Fig. 11 shows the difference between the true interior and the reconstruction with the proposed method, with and

Figure 9
Difference between the reconstruction and the true volume $x^{\#}$, when no known zone constraint is applied. (a) Along the middle line, (b) along the middle column.

Figure 10
(a) Phantom 'Lena'. Ellipses with high gray values have been added to accentuate the local tomography setup. The outer circle is the ROI, and the inner circle is the known region. The dashed lines indicate the profiles which are to be plotted in the reconstructed slice. (b) View of the ROI with adapted contrast.

without the known zone constraint. In this case, the differences between the proposed reconstruction and the true interior are not significant enough to display the curves of absolute line profiles, so only the difference profiles are shown. The parameter $\sigma = s = 3$ has been used for these reconstructions. Fig. 12 shows the reconstructed images. As the contrast is high, the cupping effect is not prominent on the reconstructions; however, an excessive brightness can be seen on the bottom side.

It is also interesting to visualize the reconstruction of the whole extended image. As can be seen in Fig. 13, the Gaussian basis even yields an approximation of the exterior. This approximation is actually important for modeling the contribution of the external part in the acquired sinogram. The bias correction is thus closely related to the modeling of the external part.
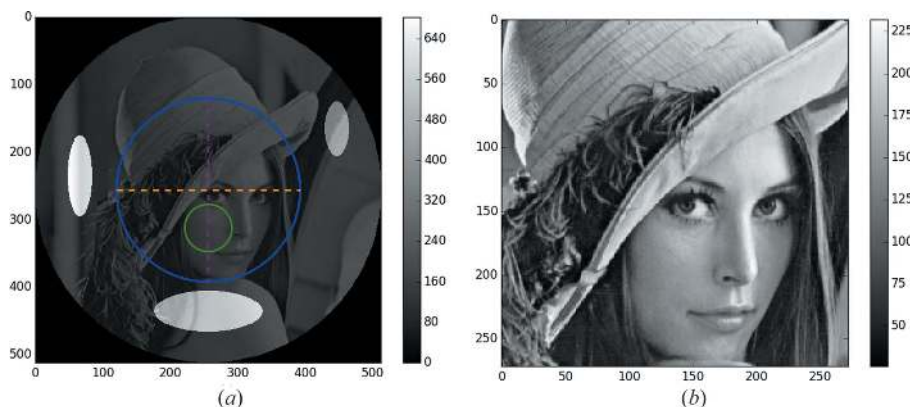
The third test involves the image of a pencil resulting from a scan at the ESRF ID19 beamline, 512×512 pixels, shown on Fig. 14. The width of the extended image is $N_2 = 520$.

Fig. 15 shows profiles of the difference between the reconstruction and the true interior. On this image, a greater radius also improves the cupping removal. The profile of a line through the reconstructed image is depicted in Fig. 16.

As a last remark, Fig. 17 shows the result of this method without using the known zone constraint, that is, without applying the constraint $g_{|\Omega_g} = g_0$ in (8). As expected, there is a not-null mean bias, even if it has been reduced with respect to padded FBP.

Beside visual inspection, reconstructions can be quantitatively compared with the true interior of the test image. Table 1 shows a comparison with two image metrics: peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM). As these metrics are indicators of an *average* distance between two images, we believe it is well suited for this purpose of evaluating how the low frequencies are corrected by the proposed method.

These results suggest that the proposed method yields better overall reconstruction quality than padded FBP. In particular, it does not induce spurious distortion in the reconstruction. For the 'Lena' test case, a similar reconstruction quality was obtained with $(\sigma, s) = (4, 6)$ with respect to $(\sigma, s) = (3, 3)$, which indicates that a coarser basis does not always yield worse reconstruction results.

For all the reconstructions with the proposed methods, the conjugate gradient optimization algorithm was used. The convergence is reached within

400 iterations. The prototype method given by Paleo (2016) takes the following execution times on a machine with an Intel Xeon CPU E5-1607 3.00 GHz CPU, and a Nvidia GTX 750 Ti GPU: 7 s for a 260 × 260 extended slice, 32 s for a 520 × 520 extended slice, and 152 s for a 1040 × 1040 extended slice. A high-performance implementation of this method, where the data are kept on GPU and the projector is implemented as described in §2.2.3, can probably achieve much shorter execution times.

The proposed method depends on some parameters. The first is the size of the extended image, which should be big enough to model the contribution of the external part. The other parameters are the Gaussian standard deviation σ and the spacing s of the grid. Both are related in a way that the Gaussian functions should slightly overlap to approximate constant functions: if the value of s is high, then σ should also be high, and *vice versa*. These parameters essentially tune how coarse the Gaussian basis is: high values would yield fast convergence but coarse result, while small values would yield slow convergence and fine result.

Using a Gaussian basis does not yield an exact correction of the error, as Gaussian functions defined in equation (3) do not form a basis. For example, Gaussian functions do not yield a partition of unity, although a very close approximation of this property can be achieved (Bale *et al.*, 2002). Thus, the final reconstruction cannot be *exact* due to the basis coarseness, but can provide results quite close to FBP with full data as seen in Figs. 6 and 16.

4. Advantages against the existing exact approaches

As mentioned in the *Introduction*, there are two main types of exact methods for local tomography. The first requires a special geometry: the FOV has to extend the object on at least one side (Clackdoyle & Defrise, 2010). This approach is not suited for interior tomography that we are focusing on. In the second type, a prior knowledge is used, either on a subregion or on the slice properties.

In the work of Kudo *et al.* (2008), an algorithm for exact local reconstruction is proposed, along with proof of uniqueness. This algorithm is based on differentiated backprojection and projection onto convex sets to invert the truncated Hilbert transform. Although this algorithm seems very promising, its implementation is not straightforward;
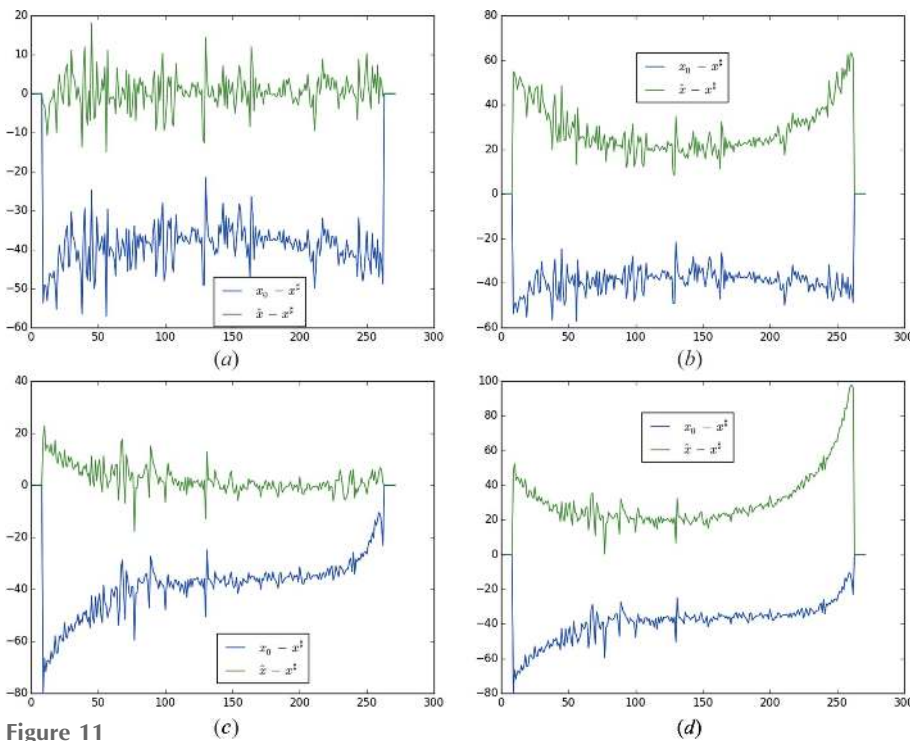
Figure 11
Profiles of difference between the reconstruction and the true interior for the Lena image. x_0, $\hat{x}$ and $x^{\#}$ are the padded FBP, the proposed reconstruction and the true interior, respectively. In blue: difference between the padded FBP and the true interior. In green: difference between the reconstruction with the proposed method with $\sigma = s = 3$ and the true interior. First row: profiles of the middle line of the image for (a) $R = 35$, (b) no constraint ($R = 0$). Second row: profiles of the middle column for (c) $R = 35$, (d) no constraint ($R = 0$).
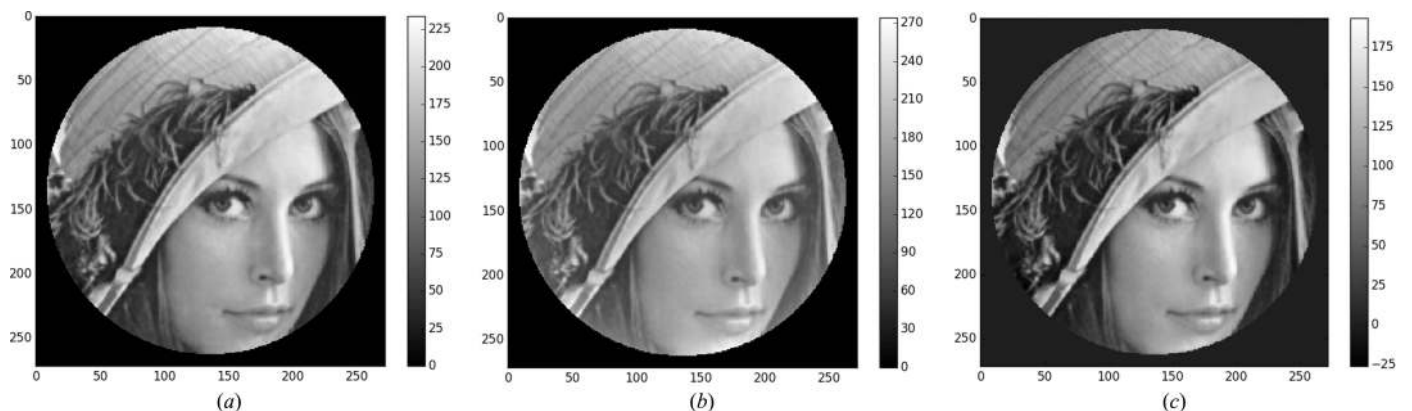
Figure 12
Reconstructions with (a) proposed method, known zone of radius $R = 35$, (b) proposed method, no known zone ($R = 0$), (c) padded FBP.
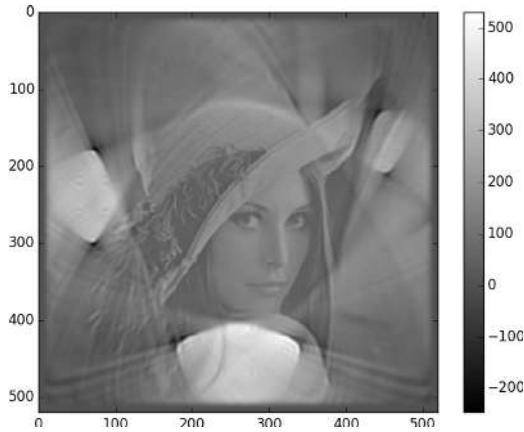
Figure 13
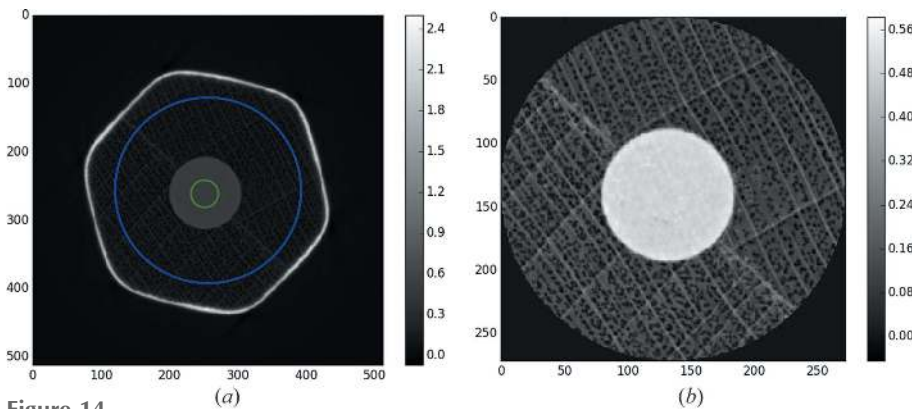Extended image after solving (8), without cropping to the ROI. The parameters used were $\sigma = s = 3$ and $R = 35$.

(a)

(b)

Figure 14
(*a*) Pencil test image. In red: ROI. In green: known sub-region. (*b*) View of the ROI.

for example, convex sets $C_1, \ldots, C_5$ may be cumbersome to define, and projections on these sets are not easy to compute. To our knowledge, no standard implementation of this method can be found.

On the other hand, several compressed-sensing methods were developed for interior tomography (Niinimäki *et al.*, 2007; Chaves Brandao~dos Santos *et al.*, 2014; Klann *et al.*, 2015). As these methods boil down to solving an optimization problem, the known region constraint can be integrated in these cases.
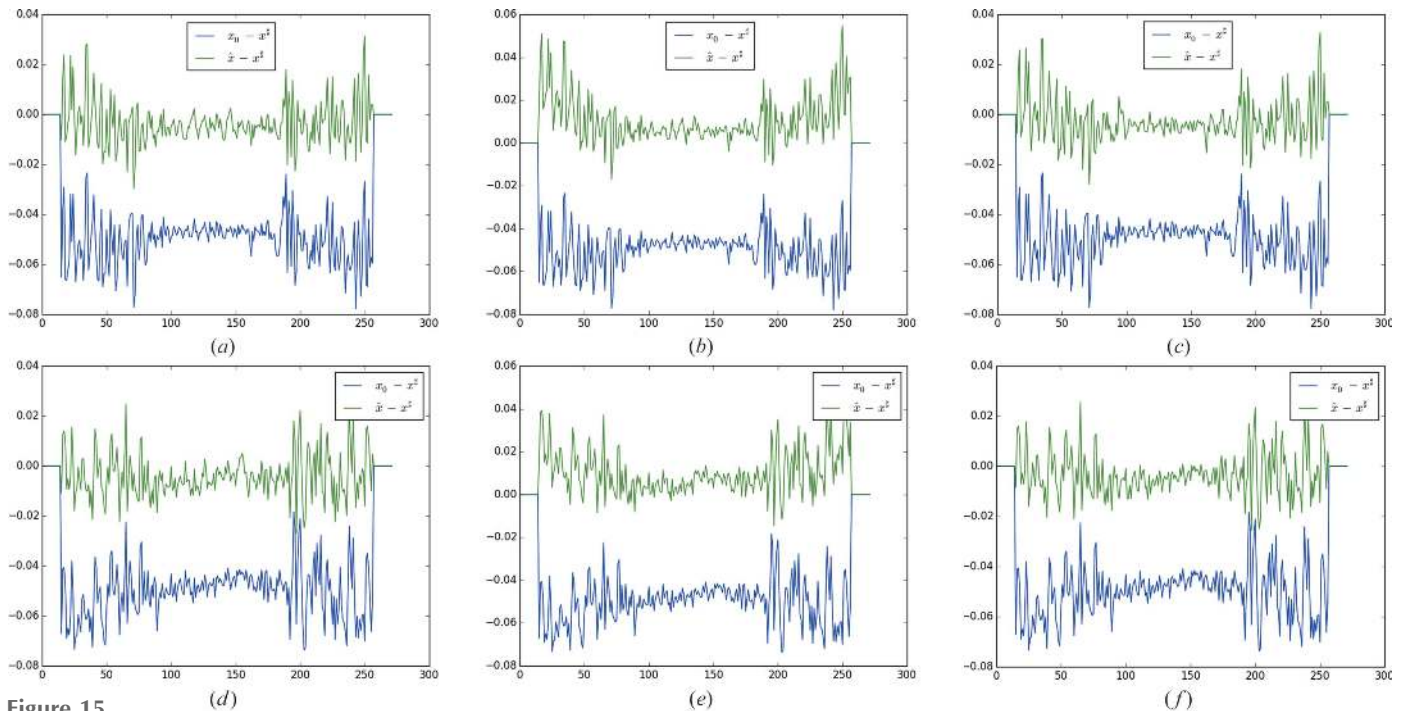
To highlight the advantage of the coarse basis representation with respect to a standard pixel basis, the proposed method is tested against a method solving

$$\arg\min_x \left\| CPx - d \right\|_2^2 + \lambda \left\| \nabla x \right\|_1 \quad \text{s.t.} \quad x_{|\Omega} = u_0 \qquad (11)$$

for x in the pixel space, meaning that all the variables are free, except those in the known region. The term $\|\nabla x\|_1$ is the total variation (TV) semi-norm which helps to reduce the ill-posedness of problem (11). This problem is a simple instance of the aforementioned compressed sensing methods, where the TV expresses the prior knowledge that the image is piecewise-constant. Related work on this method is given in §3 of Rashed & Kudo (2010). The proposed methods have two fundamental differences from this approach:
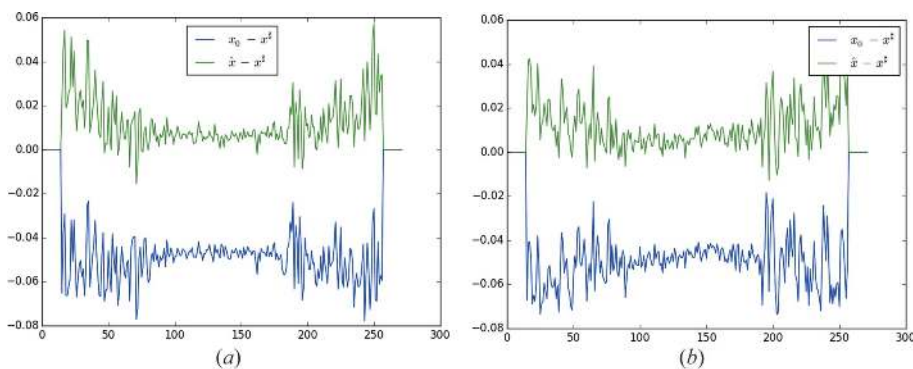

(a)

(b)

(c)

(d)

(e)

(f)

Figure 15
Profiles of difference between the reconstruction and the true interior for the pencil image. $x_0, \hat{x}$ and $x^{\#}$ are the padded FBP, the proposed reconstruction and the true interior, respectively. In blue: difference between the padded FBP and the true interior. In green: difference between the reconstruction with the proposed method with $\sigma = s = 3$ and the true interior. First row: profiles of the middle line of the image for (*a*) $R = 20$, (*b*) $R = 10$, (*c*) $R = 40$. Second row: profiles of the middle column for (*d*) $R = 20$, (*e*) $R = 10$, (*f*) $R = 40$.
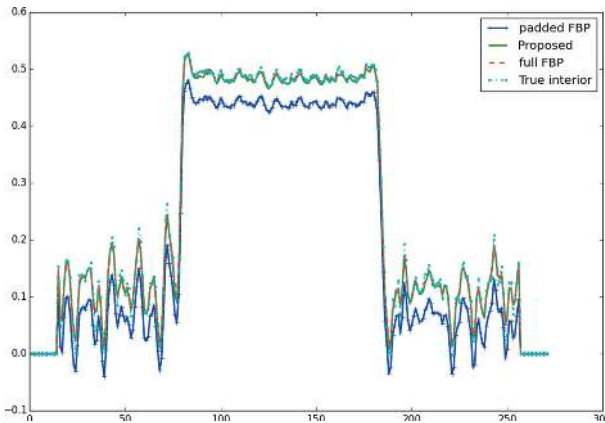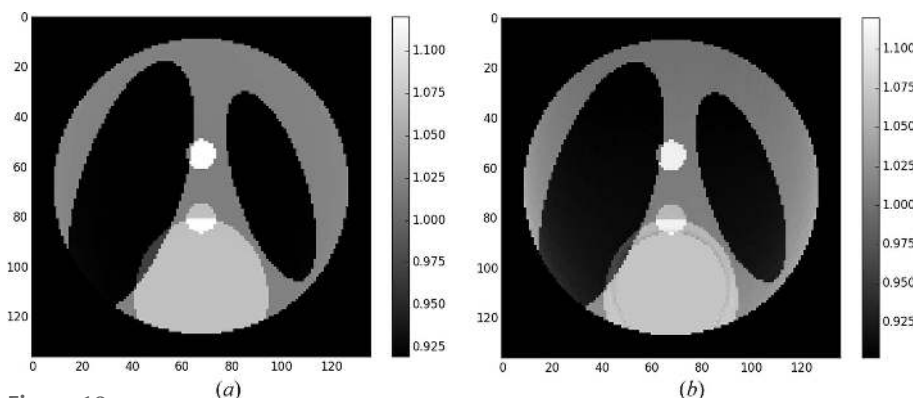
Figure 16
Line profiles for the pencil image. The proposed method were applied with parameters $\sigma = s = 3$ and $R = 40$.

Table 1
Metrics of reconstruction quality for the three test images, computed inside the reconstructed ROI.

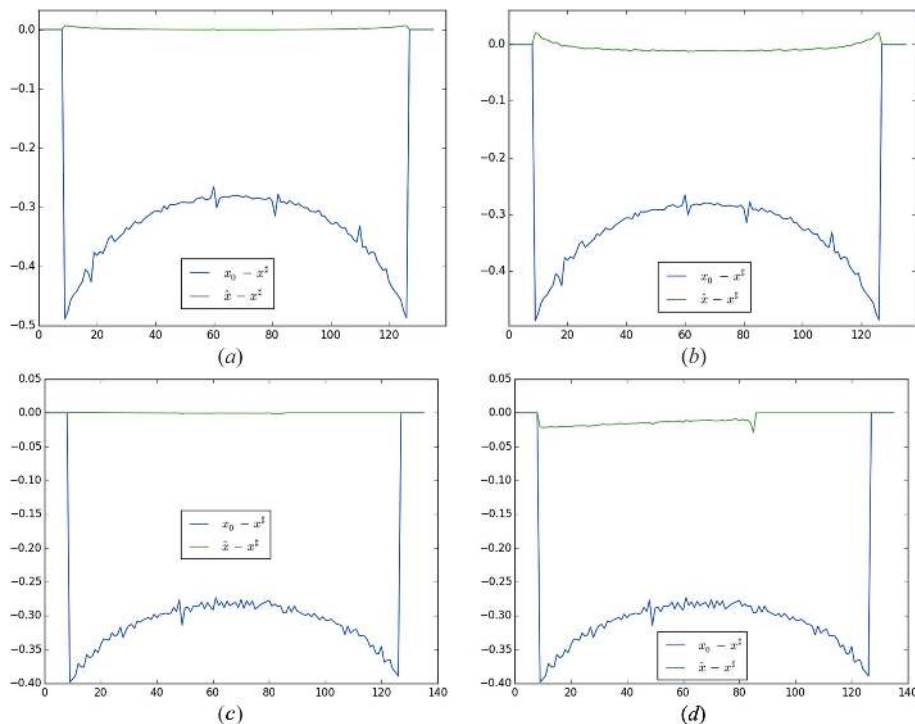| Image	Reconstruction method	Parameters	PSNR	SSIM
Shepp–Logan	Padded FBP		20.09	0.5751
Shepp–Logan	Proposed	$\sigma = 4, s = 6, R = 20$	38.40	0.6362
Shepp–Logan	Proposed	$\sigma = 5, s = 6, R = 20$	33.96	0.6360
Lena	Padded FBP		22.65	0.8417
Lena	Proposed	$\sigma = s = 3, R = 35$	35.89	0.9582
Lena	Proposed	$\sigma = 4, s = 6, R = 35$	33.80	0.9588
Pencil	Padded FBP		26.41	0.8542
Pencil	Proposed	$\sigma = s = 3, R = 10$	31.15	0.9840
Pencil	Proposed	$\sigma = 4, s = 6, R = 40$	31.91	0.9901
Pencil	Proposed	$\sigma = s = 3, R = 40$	34.22	0.9906

firstly, the variables are the coefficients of a coarse basis; secondly, the variables are optimized with respect to the *reconstruction error*.

Equation (11) was tested on the 256×256 Shepp–Logan phantom (Fig. 4) with the same conditions as previously. As may be seen from Figs. 18 and 19, it yields a near-perfect reconstruction (PSNR = 58.63, SSIM = 0.9999), as TV is adapted for this piecewise-constant phantom. However, it took of the order of 5000 iterations of the preconditioned

Chambolle–Pock optimization algorithm (Pock & Chambolle, 2011) to yield this result, as the convergence is extremely slow. An approximate result, with a slight cupping remaining, can be obtained with fewer iterations (Fig. 18b). This test case of 256×256 took approximately 67 s. The execution time for a 512×512 image, with 15000 iterations to converge, was measured to be 430 s. Thus, this method becomes impracticable, even with an efficient implementation, as it would take dozens of minutes for a single slice of modern datasets (2048×2048 or 4096×4096 pixels).

The advantage of the proposed method is to reduce the number of unknowns by modifying the forward model, expressing the variables in a coarse basis. On the implementation side, operators can be efficiently implemented according to §2.2.3, which is crucial when dealing with large image sizes. Results presented in this work were obtained with a prototype, which suggests that shorter execution times can be expected with an efficient implementation.

Figure 17
Profiles of difference between the reconstruction without known zone constraint and the true interior for the pencil image. (*a*) Line profile. (*b*) Column profile.

5. Conclusion

We have presented a new technique of local tomography reconstruction based on the knowledge of a zone of the ROI. This technique corrects the cupping effect in an initial reconstruction by expressing the error in a coarse basis of Gaussian functions. In accordance with local tomography uniqueness theorems, this method yields almost exact reconstructions, in spite of being only a correction with a coarse basis. Besides, practical considerations are given for an efficient implementation suitable for reconstruction of real data. A commented implementation of a prototype of this method is given by Paleo (2016).

Figure 18
Reconstruction with the exact pixel-domain method. (*a*) After 5000 iterations. (*b*) After 1000 iterations.

Pierre Paleo *et al.* · Practical local tomography reconstruction algorithm   **267**

Figure 19
Profiles of difference between the reconstruction and the true interior for the Shepp–Logan image. (*a*, *c*) Middle line and column after 5000 iterations. (*b*, *d*) Middle line and column after 1000 iterations.

References

Aarle, W. van, Palenstijn, W. J., De Beenhouwer, J., Altantzis, T., Bals, S., Batenburg, K. J. & Sijbers, J. (2015). *Ultramicroscopy*, **157**, 35–47.

Bale, R. A., Grossman, J. P., Margrave, G. F. & Lamoureux, M. P. (2002). *Multidimensional partitions of unity and Gaussian terrains.* CREWES Research Report, Vol. 14.

Bilgot, A., Desbat, L. & Perrier, V. (2011). *2011 IEEE Nucl. Sci. Symp. Conf. Rec.* pp. 4080-4085.

Chaves Brandao dos Santos, L., Gouillart, E. & Talbot, H. (2014). *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 1768–1772.

Clackdoyle, R. & Defrise, M. (2010). *IEEE Signal Process. Mag.* **27**, 60–80.

Clackdoyle, R., Noo, F., Junyu Guo, & Roberts, J. (2004). *IEEE Trans. Nucl. Sci.* **51**, 2570–2578.

Clackdoyle, R., Roy, D., Defrise, M., Mennessier, C. & Mohamed, M.-S. (2009). *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, pp. 3367–3371. IEEE.

Courdurier, M., Noo, F., Defrise, M. & Kudo, H. (2008). *Inverse Probl.* **24**, 065001.

Defrise, M., Noo, F., Clackdoyle, R. & Kudo, H. (2006). *Inverse Probl.* **22**, 1037–1053.

Kannappan, P. & Sahoo, P. K. (1992). *SIAM J. Math. Anal.* **23**, 1342–1351.

Klann, E., Quinto, E. T. & Ramlau, R. (2015). *Inverse Probl.* **31**, 025001.

Kudo, H., Courdurier, M., Noo, F. & Defrise, M. (2008). *Phys. Med. Biol.* **53**, 2207–2231.

Kyrieleis, A., Titarenko, V., Ibison, M., Connolley, T. & Withers, P. (2011). *J. Microsc.* **241**, 69–82.

Lee, M., Han, Y., Ward, J. P., Unser, M. & Ye, J. C. (2015). *SIAM J. Imaging Sci.* **8**, 2452–2486.

Lewitt, R. M. (1992). *Phys. Med. Biol.* **37**, 705–716.

Niinimäki, K., Siltanen, S. & Kolehmainen, V. (2007). *Phys. Med. Biol.* **52**, 6663–6678.

Noo, F., Clackdoyle, R. & Pack, J. D. (2004). *Phys. Med. Biol.* **49**, 3903–3923.

Noo, F., Defrise, M., Clackdoyle, R. & Kudo, H. (2002). *Phys. Med. Biol.* **47**, 2525–2546.

Palenstijn, W., Batenburg, K. & Sijbers, J. (2011). *J. Struct. Biol.* **176**, 250–253.

Paleo, P. (2016). *Implementation of a new method for local tomography reconstruction*, https://github.com/pierrepaleo/localtomo.

Pock, T. & Chambolle, A. (2011). *International Conference on Computer Vision*, pp. 1762–1769. IEEE.

Rashed, E. A. & Kudo, H. (2010). *Mathematical Programming in the 21st Century: Algorithms and Modeling*, **1676**, 145–156.

Rashid-Farrokhi, F., Liu, K., Berenstein, C. A. & Walnut, D. (1997). *IEEE Trans. Image Processing*, **6**, 1412–1430.

Sastry, C. S. & Das, P. (2005). *ANZIAM J.* **46**, 341–360.

Shuangren Zhao, K. Y. & Yang, X. (2011). *J. X-ray Sci. Technol.* **19**, 155.

Tang, S., Yang, Y. & Tang, X. (2012). *J. X-ray Sci. Technol.* **20**, 405.

Van Gompel, G. (2009). Phd thesis, University of Antwerp, Antwerp, Belgiium.

Van Gompel, G., Defrise, M. & Van Dyck, D. (2006). *Proc. SPIE*, **6142**, 61424B.

Wang, G. & Yu, H. (2013). *Phys. Med. Biol.* **58**, R161–R186.

Yang, J., Yu, H., Jiang, M. & Wang, G. (2010). *Inverse Probl.* **26**, 035013.

Ye, Y., Yu, H., Wei, Y. & Wang, G. (2007). *Int. J. Biomed. Imaging*, **2007**, 63634.

Yu, H. & Wang, G. (2007). *Int. J. Biomed. Imaging.* **2007**, 95295.

Yu, H. & Wang, G. (2009). *Phys. Med. Biol.* **54**, 2791–2805.

Zeng, G. L. & Gullberg, G. T. (2000). *IEEE Trans. Med. Imaging*, **19**, 548–555.

Zou, Y., Pan, X. & Sidky, E. Y. (2005). *Phys. Med. Biol.* **50**, 13–27.