A Novel Tomographic Reconstruction Method Based on the Robust Student's *t* Function For Suppressing Data Outliers

Daniil Kazantsev , Folkert Bleichrodt, Tristan van Leeuwen, Anders Kaestner, Philip J. Withers, Kees Joost Batenburg, *Member, IEEE*, and Peter D. Lee

Abstract—**Regularized iterative reconstruction methods in computed tomography can be effective when reconstructing from mildly inaccurate undersampled measurements. These approaches will fail, however, when more prominent data errors, or outliers, are present. These outliers are associated with various inaccuracies of the acquisition process: defective pixels or miscalibrated camera sensors, scattering, missing angles, etc. To account for such large outliers, robust data misfit functions, such as the generalized Huber function, have been applied successfully in the past. In conjunction with regularization techniques, these methods can overcome problems with both limited data and outliers. This paper proposes a novel reconstruction approach using a robust data fitting term which is based on the Student's *t* distribution. This misfit promises to be even more robust than the Huber misfit as it assigns a smaller penalty to large outliers. We include the total variation regularization term and automatic estimation of a scaling parameter that appears in the Student's *t* function. We demonstrate the effectiveness of the technique by using a realistic synthetic phantom and also apply it to a real neutron dataset.**

Index Terms—**Limited angle regularization, neutron tomography, proximal point, ring artifacts, robust statistics, X-ray CT, zingers.**
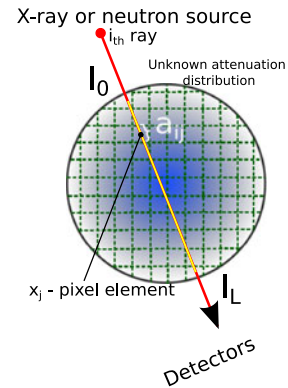
Fig. 1. Schematic representation of discrete forward projection, where a ray contributes to the i-th detector element. The registered contribution (physically related to the number of unabsorbed photons) is a sum of intersected pixel densities multiplied by their intersections lengths a_{ij}.

I. INTRODUCTION

TOMOGRAPHIC imaging provides an opportunity to explore the inner structure of materials in a non-destructive fashion using penetrating electromagnetic radiation (X-rays) or particle radiation (neutrons) [1]. The image formation relies on photoelectric absorption. The mathematical model we adopt in this paper is based on three basic assumptions: beam monochromaticity, absence of diffraction or refraction, and Beer's law [1], [2]:

$$\overline{\Lambda}_i = I_{0,i} \exp\left(-\int_{L_i} \mu(\mathbf{r})dl\right), \ i = 1, \ldots, m, \qquad (1)$$

where $\mathbf{r} \in \mathbb{R}^2$ is the spatial position, μ is the attenuation coefficient of the object under investigation, $I_{0,i}$ is the intensity of the incoming beam, $\overline{\Lambda}_i$ is the intensity of the beam at the detector of the i-th ray integrated along the line path L_i, and m is the total number of rays. The typical setup is illustrated in Fig. 1. Taking the negative log of the relative intensity leads to a linear relationship between the attenuation coefficient and the observations:

$$\int_{L_i} \mu(\mathbf{r})dl = -\ln\left(\frac{\overline{\Lambda}_i}{I_{0,i}}\right), \ i = 1, \ldots, m. \qquad (2)$$

For X-rays, the mentioned assumptions can be fulfilled to a degree; however for neutrons, the first two are usually not
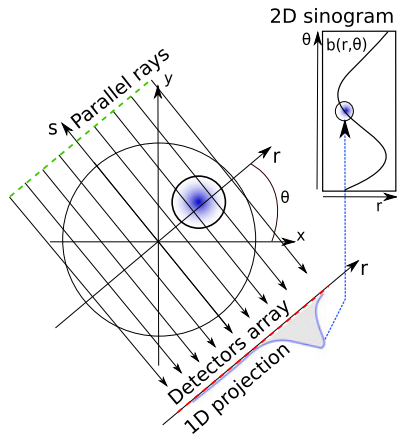
Fig. 2. Schematic figure of the parallel beam acquisition set-up to obtain a 2D sinogram from the set of 1D projections.

satisfied; the neutron beam is normally polychromatic and neutrons are scattered coherently and non-coherently.

Representing the attenuation coefficient on a square grid of n pixels: $\{\chi_j(\mathbf{r})\}_{i=1}^n$, i.e., $\mu(\mathbf{r}) = \sum_{j=1}^n x_j \chi_j(\mathbf{r})$, we obtain a discrete, linear, forward model $\mathbf{b} = \mathbf{Ax}$, where $\mathbf{x} \in \mathbb{R}^n$ is a vector with attenuation coefficients x_j, $\mathbf{b} \in \mathbb{R}^m$ is a vector with the log-corrected measurements, $b_i = -\ln\left(\frac{\bar{\Lambda}_i}{I_{0,i}}\right)$, and $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the projection matrix with elements $a_{ij} = \int_{L_i} \chi_j(\mathbf{r}) dl$ which captures the contribution of the i-th ray to the j-th pixel (see Fig. 1).

The tomographic acquisition process is based on collecting projections of an object at multiple angles. In the parallel-beam setup, multiple, parallel, rays are acquired at each rotation angle θ. Collecting these measurements for multiple angles leads to a two-dimensional *sinogram*, as illustrated in Fig. 2. The image reconstruction problem entails recovering the unknown attenuation coefficients x_j from the intensity measurements b_i.

When the measurements are densely sampled over a full angular range of π radians and not noisy, one can recover the attenuation coefficients using direct reconstruction methods such as Filtered Back Projection (FBP) [3]. When the angular range is not complete, the reconstruction problem is referred to as *limited angle* or a *missing wedge* [4]–[6]. The missing wedge reconstruction problem poses major difficulties due to non-uniqueness of the solution [2]. In addition, if the data are sparse the reconstruction is ill-posed. This relates to the issue of undersampled signal recovery in the frequency domain. Limited angle data normally result in aliasing streak artifacts and substantial loss of resolution in reconstructed images.

Iterative Image Reconstruction (IIR) techniques are better adapted to deal with noisy, under-sampled or incomplete data. Allowing more accurate modelling of the acquisition process one can incorporate *a priori* information (e.g. non-negativity, smoothness, sparsity, etc.) [1], [7], [8]. Regularized IIR methods are capable of producing images of high quality using fewer projections than required by the Nyquist sampling theorem [9]. A significant amount of research in the area of IIR is dedicated to the development of novel regularization methods, such as using the Total Variation (TV) penalty [10] and various generaliza-

tions of it [4], [5]. Much less attention, however, is given to the data misfit term. Although regularization is crucial to stabilize convergence and damp random noise in the reconstructed image, it cannot completely mitigate the effects of large or systematic errors present in the data. In other words, regularization alone will not solve problems related to typical acquisition artifacts, such as a streak (caused by an error in a random detector element) or a ring (a consistent error-offset due to a miscalibrated or defective detector). A traditional approach to eliminate those artifacts would be a selective filtering of radiographs, the sinogram [11]–[13] or the image itself [14]. However, no matter how successful the preprocessing approach is, it includes modification (filtration) of measurements. This can often result in a biased reconstructed image with introduced artifacts due to the filtering process [12]. In this case, the use of IIR is not recommended since the data have been modified unpredictably and do not represent real measurements. A better alternative would be to model various data inaccuracies within the IIR framework.

Using a Model Based IIR (MBIIR) approach, one can minimize data errors through more accurate modelling of the forward problem [15]–[18]. A good overview of MBIIR methods is given in [19]. Frequently used, the Least-Squares (LS) noise model assumes Gaussian noise in the measurements which is suitable in case of high radiation dose acquisitions. For lower dose imaging (e.g. emission tomography or dynamic tomography), the Poisson model [18], [20], [21], or statistically weighted LS model [15], [22], [23] are more appropriate. These models, however, are highly sensitive to severe errors (outliers) present in data. Recently, novel data models which are robust to outliers have been proposed. One group of methods is based on the generalized Huber penalty [15], [16], where outliers are removed by solving an auxiliary optimization problem. The Student's t penalty is a more robust (statistically) alternative to the Huber penalty [17], [24], [25].

In this work, we investigate the use of various robust data misfit penalties in conjunction with the TV penalty to minimize artifacts. In particular, we consider the weighted LS, Huber, and the Student's t penalties. The resulting optimization problem has a special structure; the objective is the sum of a smooth and a convex term. We use a proximal-gradient algorithm [26]–[29] to optimize the objective. By casting the various approaches into a common framework, we can easily compare the methods both from a theoretical as well as an experimental point-of-view. Extensive numerical experiments are performed with a synthetic phantom and real neutron measurements.

The paper is organized as follows: in Section II, the reconstruction problem with various robust penalties is formulated. Section III presents an optimization framework. Numerical results with a synthetic phantom are given in Section IV and real data results in V. Section VI concludes the manuscript.

II. PROBLEM FORMULATION AND ROBUST DATA MISFIT

The measured transmission data (1) normally contain noise due to the photon-counting process and Gaussian noise due to the electronic response [22]. With recent improvements in acquisition technology, Gaussian noise can be neglected and the

actual measured intensities Λ_i modelled as a Poisson random variable with parameter $\overline{\Lambda}_i$ [20], [23]:

$$\Lambda_i \sim \text{Poiss}\{\overline{\Lambda}_i\} = \text{Poiss}\{I_{0,i} e^{-[\mathbf{A}\mathbf{x}]_i}\}, \; i = 1, \ldots, m \quad (3)$$

One can approach the reconstruction problem from the Bayesian perspective by employing maximum *a posteriori* (MAP) estimate [15], [18], [20], [22], [23]:

$$\arg \min_{\mathbf{x}} \sum_{i=1}^{m} \left[I_{0,i}[\mathbf{A}\mathbf{x}]_i + \Lambda_i e^{-[\mathbf{A}\mathbf{x}]_i} \right] + \text{L}_{\text{prior}}(\mathbf{x}), \quad (4)$$

where L_{prior} is the negative log of the prior density function. In [23], a quadratic approximation to the likelihood function (4) was proposed, leading to

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2} (\mathbf{A}\mathbf{x} - \mathbf{b})^\top \mathbf{\Lambda} (\mathbf{A}\mathbf{x} - \mathbf{b}) + \text{L}_{\text{prior}}(\mathbf{x}), \quad (5)$$

where $\mathbf{\Lambda} \in \mathbb{R}^{m \times m}$ is a diagonal matrix with entries $\{\Lambda_i\}_{i=1}^m$. This so-called Penalized Weighted Least Squares (PWLS) model is a simple approximation to the Poisson model (3) which is used for many X-ray tomography cases [15], [22], [23]. The matrix $\mathbf{\Lambda}$ effectively gives a higher weight to projections with a higher photon-count. In the remainder of the paper we absorb these weights in the system of equations, i.e. redefining $\mathbf{A} \equiv \sqrt{\mathbf{\Lambda}}\mathbf{A}$ and $\mathbf{b} \equiv \sqrt{\mathbf{\Lambda}}\mathbf{b}$. We now express the corresponding regularized optimization problem in the general form:

$$\min_{\mathbf{x}} f(\mathbf{A}\mathbf{x} - \mathbf{b}) + \beta g(\mathbf{x}), \quad (6)$$

where $f : \mathbb{R}^m \to \mathbb{R}_+$ is a data misfit term, $g : \mathbb{R}^n \to \mathbb{R}_+$ is a convex penalty, and β is the regularization parameter. Next, we review several data misfit alternatives which can be used to replace a standard LS misfit of the form $f(\cdot) = \| \cdot \|_2^2$. Evaluation of some of these involves solving an auxiliary optimization problem that either has a closed-form solution or is easy to solve numerically. Next, we discuss various choices of the misfit function f.

A. Regularized Least Squares

Let $f(\cdot) = \| \cdot \|_2^2$ and $g(\cdot) = \| \cdot \|_{\mathbf{L}}^2$, which corresponds to a Gaussian prior distribution for $\mathbf{x}$ with mean zero and covariance $\mathbf{L}^\top \mathbf{L}$. The optimization problem (6) now has the following closed form solution $\hat{\mathbf{x}} = \mathbf{H}^{-1}\mathbf{A}^\top \mathbf{\Lambda}\mathbf{b}$, where $\mathbf{H} = [\mathbf{A}^\top \mathbf{\Lambda}\mathbf{A} + \beta \mathbf{L}^\top \mathbf{L}]$. It is infeasible to precompute the inverse of $\mathbf{H}$ exactly for large-scale problems, therefore various approximations can be used in practice [8]. The Toeplitz-block-Toeplitz nature of the matrix $\mathbf{H}$ can be also exploited by creating a deconvolution kernel which can be efficiently implemented through fast Fourier routines [23]. In this work, we will be using first-order optimization approaches which avoid explicit calculation of $\mathbf{H}^{-1}$ altogether [26]–[29].

B. Huber

A well-known robust misfit function is the Huber function [30] (see Fig. 3), which can be defined as the Moreau-Yosida envelope of the ℓ_1-norm [31]

$$f(\mathbf{r}) = \min_{\mathbf{s}} \frac{1}{2} \|\mathbf{s} - \mathbf{r}\|_2^2 + \lambda \|\mathbf{s}\|_1. \quad (7)$$
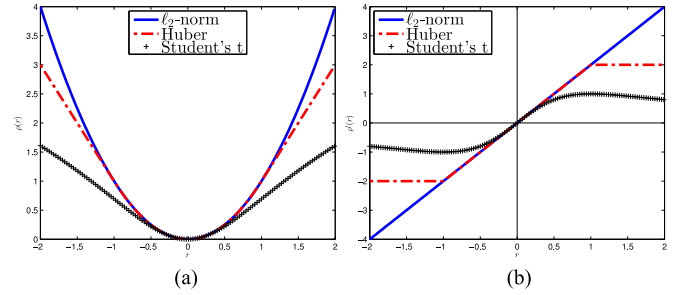

Fig. 3. LS, Huber and Student's t misfit functions (a) and their corresponding influence functions (b).

This optimization problem has a closed-form solution $\widehat{\mathbf{s}}$ with elements

$$\widehat{s}_i = \mathcal{S}_\lambda(r_i), \quad (8)$$

where $\mathcal{S}_\lambda$ is the (element-wise) soft-thresholding operator given by

$$\mathcal{S}_\lambda(r) = \begin{cases} r + \lambda & \text{if } r < -\lambda \\ 0 & \text{if } |r| \leq \lambda. \\ r - \lambda & \text{if } r > \lambda, \end{cases} \quad (9)$$

It is readily verified, using (8), that (7) is indeed equivalent to the Huber misfit function

$$f(\mathbf{r}) = \sum_{i=1}^{m} \rho_{\text{H}}(r_i), \quad (10)$$

where

$$\rho_{\text{H}}(r) = \begin{cases} \frac{1}{2} r^2 & \text{if } |r| \leq \lambda, \\ \lambda|r| - \frac{1}{2}\lambda^2 & \text{if } |r| > \lambda. \end{cases} \quad (11)$$

The gradient is given by

$$\nabla f(\mathbf{r}) = \mathbf{W}\mathbf{r},$$

where $\mathbf{W}$ is a diagonal matrix with elements

$$w_{ii} = \begin{cases} 1 & \text{if } |r_i| \leq \lambda, \\ \frac{\lambda}{|r_i|} & \text{if } |r_i| > \lambda. \end{cases} \quad (12)$$

Fig. 3(a) and (b) show the misfit function $\rho_{\text{H}}(r)$ and its derivative $\rho_{\text{H}}'(r)$. This plot clearly shows that the Huber misfit assigns a smaller weight to large residuals than the least-squares misfit.

C. Group-Huber (GH)

A specific form of the Huber misfit was proposed in [16]

$$f(\mathbf{r}) = \min_{\mathbf{s}} \frac{1}{2} \|\mathbf{s} - \mathbf{B}^\top \mathbf{r}\|_2^2 + \lambda \|\mathbf{s}\|_1, \quad (13)$$

where $\mathbf{B} = (\mathbf{I}_{m_1} \otimes \mathbf{1}_{m_2})/\sqrt{m_2}$ with $\mathbf{B}^\top \mathbf{B} = \mathbf{I}_{m_1}$. The reasoning behind this formulation[1] is that the outliers are likely to be correlated in such a way that the residuals can be grouped into m_1 groups with m_2 entries each. The matrix $\mathbf{B}$ sums the residuals in each group so that $\mathbf{s}$ is a vector of length m_2.

---

[1]The authors of [16] actually propose $f(\mathbf{r}) = \min_{\mathbf{s}} \frac{1}{2} \|\mathbf{B}\mathbf{s} - \mathbf{r}\|_2^2 + \lambda \|\mathbf{s}\|_1$, but both problems have the same solution.

We now find the following closed-form solution of (13)

$$\widehat{s}_i = \mathcal{S}_\lambda((\mathbf{B}^\top \mathbf{r})_i),$$

which can be written in terms of the conventional Huber misfit ρ_{H} as

$$f(\mathbf{r}) = \sum_{i=1}^m \rho_{\mathrm{H}}((\mathbf{B}^\top \mathbf{r})_i).$$

The gradient of this objective can be expressed as

$$\nabla f(\mathbf{r}) = \mathbf{B} \mathbf{W} \mathbf{B}^\top \mathbf{r}, \tag{14}$$

where $\mathbf{W}$ is defined as in (12).

D. Student's t Misfit

A more robust alternative to the Huber misfit is the Student's t misfit [17], [24], [25]. It arises when assuming that the noise follows a Student's t distribution. Even if the actual noise does not obey this assumption, the Student's t misfit has some desirable properties, such as insensitivity to large outliers (see Fig. 3). Taking the negative log of the univariate Student's t density function leads to

$$\rho_{\mathrm{StudentsT}}(r) = \log(1 + (r/\sigma)^2),$$

where σ is the noise variance parameter. It is a critical parameter since it determines the threshold beyond which residuals will have a diminishing influence. The objective and its gradient are now given by

$$f(\mathbf{r}) = \sum_{i=1}^m \log(1 + (r_i/\sigma)^2), \tag{15}$$

$$\nabla f(\mathbf{r}) = \mathbf{W} \mathbf{r}, \tag{16}$$

where $\mathbf{W}$ is a diagonal matrix with elements

$$w_{ii} = \frac{2}{\sigma^2 + r_i^2}. \tag{17}$$

A possible downside of this approach is that the misfit function is no longer convex. In practice, however, this appears not to be problematic (see Section IV).

Estimation of the variance parameter σ can be formulated as a scalar optimization problem [25]:

$$f(\mathbf{r}) = \min_\sigma \left\{ m \log(\pi\sigma) + \sum_{i=1}^m \log(1 + (r_i/\sigma)^2) \right\}. \tag{18}$$

For a given residual this problem is easily solved using a scalar minimization algorithm such as Newton's method or Golden-ratio search [32]. The resulting misfit function $f(\mathbf{r})$ is smooth and can be used in conjunction with any gradient-based optimization method [25]. The gradient of the objective w.r.t. $\mathbf{r}$ is given by $\mathbf{W}\mathbf{r}$, with $\mathbf{W}$ as in (17), evaluated at the optimal σ [25]. Algorithm 1 summarizes evaluation of the objective and its gradient.

Algorithm 1: Evaluation of the Student's t objective (15) and its gradient (17).

Input: residual $\mathbf{r}$
Output: $f(\mathbf{r})$ and $\nabla f(\mathbf{r})$
  compute
  $\widehat{\sigma} = \arg\min_\sigma m \log(\pi\sigma) + \sum_{i=1}^m \log(1 + (r_i/\sigma)^2)$
  $f(\mathbf{r}) = m \log(\pi\widehat{\sigma}) + \sum_{i=1}^m \log(1 + (r_i/\widehat{\sigma})^2)$
  $\nabla f(\mathbf{r}) = W\mathbf{r}$, with $w_{ii} = \frac{2}{\widehat{\sigma}^2 + r_i^2}$

Algorithm 2: FISTA-TV reconstruction scheme.

Input: $\mathbf{x}_0$, β, K,
Output: $\{\mathbf{x}_k\}_{k=1}^K$
  $t_1 = 1$
  for $k = 1$ to K **do**
    $\mathbf{y}_k = \mathrm{prox}_{L^{-1}\beta}[g]\left(\mathbf{x}_k - L^{-1}\mathbf{A}^\top \nabla f(\mathbf{A}\mathbf{x}_k - \mathbf{b})\right)$
    $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$
    $\mathbf{x}_{k+1} = \mathbf{y}_k + \left(\frac{t_k - 1}{t_{k+1}}\right)(\mathbf{y}_k - \mathbf{y}_{k-1})$
  end for

III. TOMOGRAPHIC IMAGE RECONSTRUCTION

A generic way to solve convex problems of the form (6) is by a proximal-gradient method [31]

$$\mathbf{x}_{k+1} = \mathrm{prox}_\beta[g]\left(\mathbf{x}_k - L^{-1} \cdot \mathbf{A}^\top \nabla f(\mathbf{A}\mathbf{x}_k - \mathbf{b})\right),$$

where $\mathrm{prox}_\beta[g](\mathbf{y}) = \arg\min_\mathbf{x} g(\mathbf{x}) + \frac{1}{2\beta}\|\mathbf{x} - \mathbf{y}\|_2^2$, and L is the Lipschitz constant of ∇f. In many cases the proximal operator is easily computed. A particularly suitable edge-preserving regularization for tomographic reconstruction is TV [10]. In this paper we employ an isotropic TV penalty: $g(\mathbf{x}) = \sum_{i=1}^m \sqrt{[D_1\mathbf{x}]_i^2 + [D_2\mathbf{x}]_i^2}$, where D_k is a discrete first derivative in direction k. The corresponding TV-proximal problem is solved using a projected gradient method [29].

Using the generic FISTA-framework given in Algorithm 2 we present four different methods: LS (when $\nabla f(\mathbf{r}) = \mathbf{r}$, $\beta = 0$), LS-TV ($\nabla f(\mathbf{r}) = \mathbf{r}$, $\beta > 0$), GH-TV (Group-Huber fidelity with $\nabla f(\mathbf{r})$ defined as in (14)), and Student-TV (Student's t fidelity with $\nabla f(\mathbf{r})$ evaluated by Algorithm 1). Notably, the statistical weighting $\mathbf{\Lambda}$ is included in all formulations, therefore the LS formulation is the PWLS one (5) (we employ the LS notation throughout the paper). We note that for the low-dose situations a true Poisson terms should be used [20] instead of approximations (5).

It should be noted that the convergence of FISTA cannot be guaranteed in general when using a non-convex penalty such as the Student's t. However, a minor modification of Algorithm 2 that includes an inertia term $\mathbf{x}_k - \mathbf{x}_{k-1}$ and some additional requirements on the line-search parameter can be shown to converge [33]. We do not expect the results to change dramatically as we did not encounter any convergence issues in the numerical experiments.
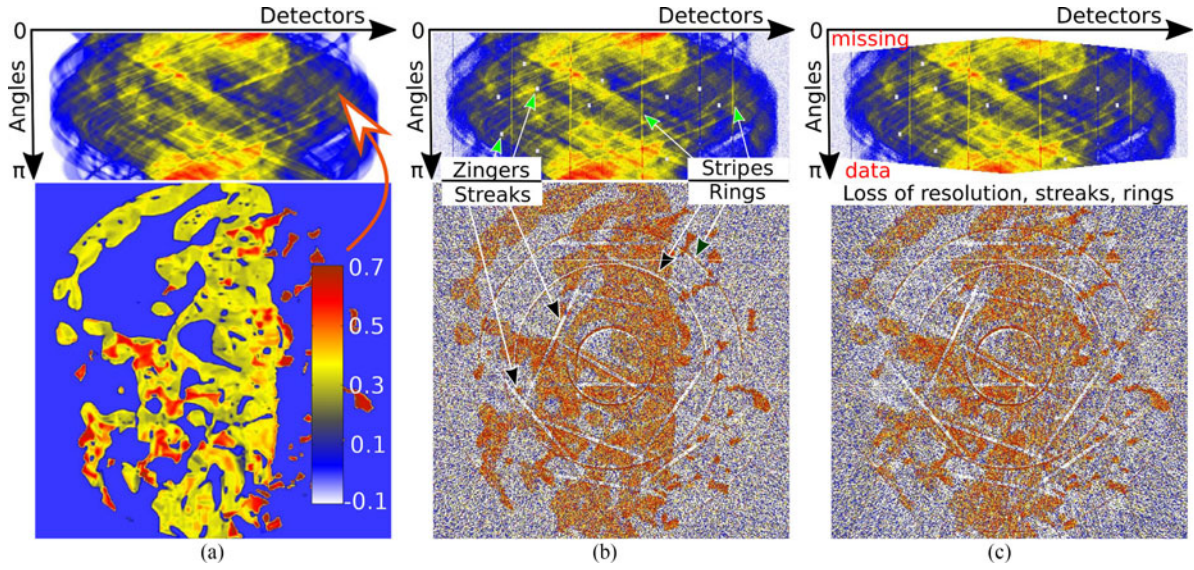
Fig. 4. (a) A numerical phantom of size 512×512 pixels and its noiseless sinogram (180 projections); (b) sinogram corrupted with Poisson noise, zingers, and vertical stripes. The corresponding image artifacts in FBP reconstruction are accentuated noise, streaks, and rings respectively; The errors: $\Delta_1 \backslash \Delta_2 = \mathbf{29 \backslash 0.10}$. (c) the sinogram (b) with the missing data (the result of limited angle acquisition). The sinogram's ramp-shaped cut-off is due to simulation of the partial shadowing effect. FBP reconstruction also contains rings, streaks, and the loss of information and contrast (blurred regions) can be seen. The errors: $\Delta_1 \backslash \Delta_2 = \mathbf{28 \backslash 0.08}$.

IV. NUMERICAL EXPERIMENTS

In this section we compare four iterative reconstruction methods: LS, LS-TV, GH-TV, and Student-TV. To test our techniques, we used a realistic phantom (see Fig. 4(a)) which was constructed from a high resolution X-ray scan of a bone-bioglass sample acquired at the DLS Diamond-Manchester Branchline [5]. To avoid the "inverse crime" [34], we used different pixel grids as well as different projection models (strip and linear) for data-generation and reconstruction. Poisson distributed noise was added to the projection data, assuming an incoming beam intensity of 5×10^3 (photon count).

To quantify the quality of the reconstructions we use three metrics of quality. The first metric is the Normalized Root Mean Square Error (NRMSE) Δ_1 given as:

$$\Delta_1(\hat{\mathbf{x}}, \mathbf{x}) = \frac{1}{|\text{ROI}|} \sum_{i \in \text{ROI}} (x_i - \hat{x}_i)^2 \times 100. \qquad (19)$$

where $\hat{\mathbf{x}}$ is an exact image and ROI is the region-of-interest. In the experiments, the ROI is selected as the phantom without the uniform zero background (see Fig. 4(a)). The second metric is the structural similarity index (SSIM) [35] which is given by:

$$\Delta_2(\hat{\mathbf{x}}, \mathbf{x}) = \frac{(2\mu_{\mathbf{x}}\mu_{\hat{\mathbf{x}}} + C_1) + (2\varsigma_{\mathbf{x}\hat{\mathbf{x}}} + C_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\hat{\mathbf{x}}}^2 + C_1)(\varsigma_{\mu_{\mathbf{x}}}^2 + \varsigma_{\mu_{\hat{\mathbf{x}}}}^2 + C_2)}, \qquad (20)$$

where μ and ς are the mean intensity and standard deviation of a patch centered around a pixel respectively (we used an 8×8 patch). The variable $\varsigma_{\mathbf{x}\hat{\mathbf{x}}}$ denotes the cross-correlation and $C_{1,2}$ are small constants to avoid division by zero [35]. The SSIM is different quality metric from the NRMSE (19), as it considers image degradation as a visually perceived change in the structural information. The SSIM value equals 1 if images are identical and -1 if they anti-correlated, i.e., $\Delta_2 \in [-1, 1]$. The third metric is the Coefficient of Variation (CoV) [21] and defined as:

$$\text{CoV}(\mathbf{x}_\omega) = \frac{1}{|\text{ROI}|} \sum_{i \in \text{ROI}} \sqrt{\frac{(x_{i,\omega} - \bar{x}_i)^2}{(\Omega - 1)\bar{x}_i^2}}, \qquad (21)$$

where $\bar{x}$ is the average over all noise realizations: $\Omega = 10$, $\omega = 1, \ldots, \Omega$.

For all of our experiments, we use $K = 300$ (the total number of outer FISTA iterations, see Algorithm 2). Iterations have been terminated earlier for the synthetic phantom when the error Δ_1 increases. To compute the proximal operator for the TV norm, we use a projected-gradient algorithm [29] with 20 (inner) iterations. TV-related iterations have been terminated when the residual becomes small $\|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 \leq \epsilon$, where $\epsilon = 1 \times 10^{-4}$. The Lipschitz constant L has been established using the Power method for the LS, LS-TV, GH-TV algorithms and manually for the Student-TV algorithm. One can establish it by selecting the lowest L which precludes fast divergence.

To avoid storing a large sparse projection matrix $\mathbf{A}$, we use on-the-fly forward and backward projection operations from the ASTRA toolbox [36]. Note that the ASTRA toolbox uses unmatched projectors, which formally means that the backward operator is not the transpose of the forward one. This can lead to an error in the gradient computation. However, these errors are typically much smaller than the desired tolerances, so we do not see their effect in practice [37]. The presented results can be reproduced using open-source codes [38].

A. Reconstruction of Data Corrupted by Zingers and Stripes

In this experiment we reconstruct a noisy sinogram with multiple zingers (single error pixels in the sinogram) and vertical stripes (consistent positive or negative offsets) (see Fig. 4(b), top)). We simulated some stripes to be of varying intensity and discontinuous. The corrupted sinogram has been generated from
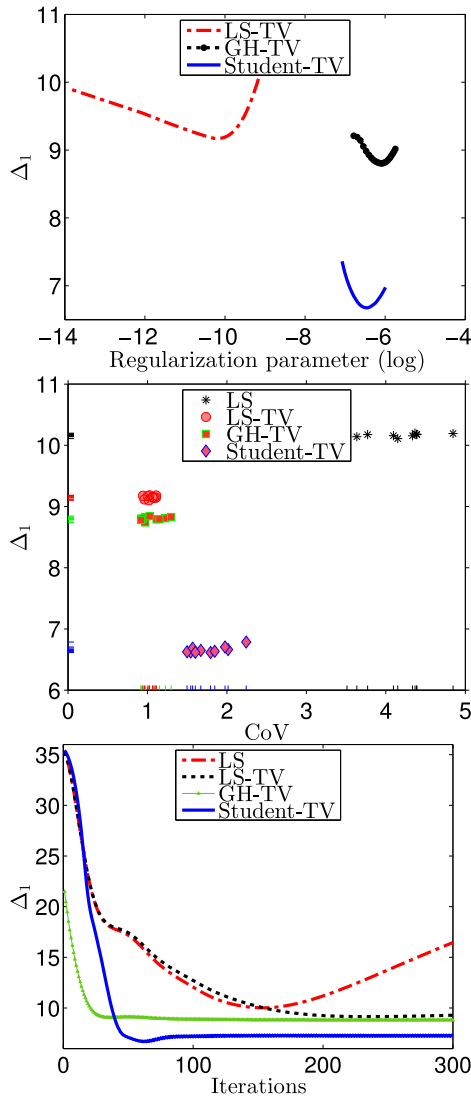
Fig. 5. **Top**: optimization curves to establish the optimal values for regularization parameters for each method. Each curve consists of 30 points of linearly progressing regularization parameter values. **Middle**: 10 Poisson noise realizations have been applied to the sinogram (see Fig. 4 (b)). Resulting values of Δ_1 against CoV are given for optimized regularization parameters. Notably, the Student-TV reconstructions have much lower bias in terms of Δ_1 error and slightly higher CoV than the LS-TV and the GH-TV methods. **Bottom**: Convergence plots for different IIR methods. The lowest error ($\Delta_1 = 6.6$) is obtained with the Student-TV method.

the noiseless version in Fig. 4(a, top) by adding noise and artifacts. In Fig. 4(b, bottom), the FBP reconstruction of the corrupted sinogram is shown. The reconstruction is exceedingly noisy and contains strong streaks and rings artifacts.

In order to equally compare all methods, we perform an empirical optimization to find the optimal regularization parameters. Each method (except LS, where $\beta = 0$) has been optimized for β with respect to NRMSE (Δ_1). Notably, regularization parameter λ for the GH-TV method has been found before the experiment.

In Fig. 5 (top) one can see how the optimal regularization parameter β (6) has been chosen based on Δ_1 values. Each curve consists of 30 points with linearly progressing regularization

parameter values. The global minimum of Δ_1 can be established for each method and the corresponding β value is selected. It can be seen that the minimum Δ_1 error belongs to the Student-TV method. We proceed with further examinations of the method.

Using the same technique as for the experiment in Fig. 5 (top), we perform an evaluation of all methods with respect to ten Poisson noise realizations (see Fig. 5 (middle)). For each noise realization, the optimal regularization parameters have been established. For each reconstruction, a minimum value of Δ_1 is obtained and plotted against the CoV value. One can see relatively compact distributions for all methods except the LS method. The values for the LS method are widespread due to absence of the regularization. The Student-TV method delivers reconstructions with much smaller Δ_1 error and only slightly larger CoV than the LS-TV and GH-TV methods. The larger variance of the Student-TV recovery can be also perceived from the reconstructed images in Fig. 6(d). One can summarize that the trade-off between bias and variance has been established by using Student's t data misfit.

To demonstrate the convergence behaviour of all methods we run 300 iterations for each method using established regularization parameters (see Fig. 5 (bottom)). One can see that the LS method triggers early divergence (minimal error $\Delta_1 = 10.1$), while the LS-TV method proceeds further reaching an error of $\Delta_1 = 9.2$. The convergence behaviour of GH-TV and Student-TV is similar. The Student-TV method has the lowest error of $\Delta_1 = 6.6$ compared to the GH-TV method $\Delta_1 = 8.8$.

In Fig. 6 (top row), we show reconstructed images for the optimal values of Δ_1 errors. In the middle row we show image residuals $(\hat{\mathbf{x}} - \mathbf{x}_k)^2$, which demonstrate the amount of structural information which differs from the ideal phantom given in Fig. 4(a, bottom). Therefore, a blank image residual (containing all zero values) would represent a perfect reconstruction; this, however, never happens in practice. In the bottom row, we demonstrate another residual in the sinogram space $(\mathbf{A}\mathbf{x}_k - \mathbf{b})^2$, where $\mathbf{b}$ is a corrupted noisy sinogram. Contrary to the image residual, if more inconsistent information appears in the sinogram residuals (e.g. offset pixels and stripes) then a smaller number of artifacts appears in the reconstructed image $\mathbf{x}_k$. Presented residuals in the image and sinogram spaces can help to explain the nature of methods and artifacts removal mechanisms.

The LS reconstruction is noisy and the corresponding image residual contains almost all artifacts introduced into the sinogram. The sinogram residual contains stripes of low intensity (i.e. the sinogram $\mathbf{A}\mathbf{x}_k$ contains strong stripes) which also confirms the presence of ring artifacts in the reconstructed image. Zingers can also be seen (as red clusters of pixels), however they are blurred (not so compact as originals in Fig. 4(d)), therefore streak artifacts are also present in the reconstructed image.

The minimal error for the LS-TV method (see Fig. 6(b)) is not significantly lower than for the LS method. This could be due to the fact that the LS-TV method oversmoothes the phantom, hence also lower CoV in Fig. 5 (middle). For the LS-TV method, the regularization parameter has been increased which helps to smooth streaks and some minor rings while the overall resolution is also lost (e.g. small pores have been
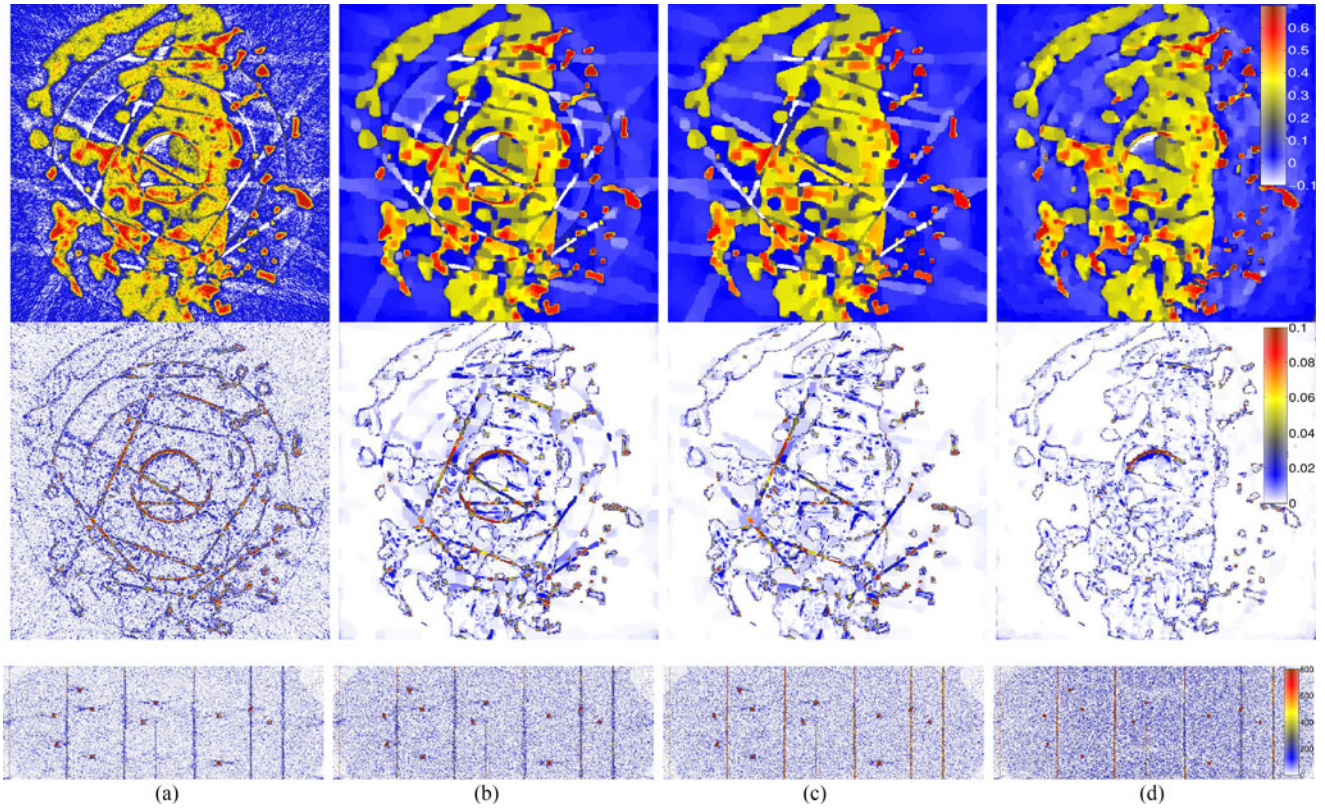
Fig. 6. **Top row**: reconstructed images for the LS, LS-TV, GH-TV, and the Student-TV methods using the corrupted noisy sinogram (see Fig. 4 (b, top)). **Middle**: residual $(\hat{\mathbf{x}} - \mathbf{x}_k)^2$ in image space. **Bottom**: residual $(\mathbf{A}\mathbf{x}_k - \mathbf{b})^2$ in sinogram space, where $\mathbf{b}$ is a corrupted noisy sinogram. Student-TV reconstruction provides the best resolution and the absence of streak and ring artifacts (the central strong ring artifact is only partially suppressed). (a) LS, $\Delta_1 \backslash \Delta_2 = \mathbf{10.1} \backslash \mathbf{0.51}$, (b) LS-TV, $\Delta_1 \backslash \Delta_2 = \mathbf{9.2} \backslash \mathbf{0.74}$, (c) GH-TV, $\Delta_1 \backslash \Delta_2 = \mathbf{8.8} \backslash \mathbf{0.79}$, (d) Stud-TV, $\Delta_1 \backslash \Delta_2 = \mathbf{6.6} \backslash \mathbf{0.85}$.

smoothed out). This observation is also confirmed by smaller values of CoV and higher bias for the LS-TV and GH-TV methods in Fig. 5 (middle). The LS-TV sinogram residual looks similar to the TV method, however the image residual is less noisy.

The GH-TV method delivers similar resolution but also effectively removes ring artifacts (the error is $\Delta_1 = 8.8$). In the image residual, some not very sharp ring contours are still present, however the streak artifacts have become more prominent. In the sinogram residual, the vertical stripes are sharp and of high intensity, which means that they have been removed from the reconstructed image. Zingers, however are not clustered similarly to the LS-TV method; streaks are present in the recovered image.

The Student-TV method provides the smallest reconstruction error ($\Delta_1 = 6.6$) among compared techniques and the absence of artifacts is evident (see Fig. 6(d)). The recovered image resolution is superior to the LS-TV and the GH-TV methods (e.g. note how the background is recovered inside smaller cavities). The image, however, appears slightly noisier and this is also confirmed by larger CoV values in Fig. 5 (middle). The degree of improvement is also clear from the corresponding image residual. It is less sharp than other image residuals, contains no streak artifacts, and only one remaining part of the central thick ring artifact. The sinogram residual shows sharp stripes of high intensity and also densely clustered zingers.

TABLE I
COMPUTATION TIMES GIVEN FOR ONE ITERATION OF ALGORITHM 2 FOR RECONSTRUCTION OF THE PHANTOM IN FIG. 4(B)

LS	LS-TV	GH-TV	Student-TV
0.03 sec.	0.06 sec.	0.07 sec.	0.16 sec.

Additionally, the SSIM values are given for all methods in Fig. 6. There is a big gap in Δ_2 values between the LS reconstruction ($\Delta_2 = 0.51$) and the LS-TV-type methods ($\Delta_2 = 0.75$-0.79). Notably the SSIM values for the LS-TV and the GH-TV methods are very close to each other. Indeed, the visual perception of these reconstructions is close to each other. Reconstruction with the Student-TV method gives a better (higher) SSIM value ($\Delta_2 = 0.85$).

In Table I we present the computation times for all methods to evaluate their computational efficiency. Here one can note that the Student-TV method is only 2 times slower than the LS-TV-type methods due to additional step involving the estimation of σ parameter in Algorithm 1.

In order to further explore the Student-TV method we apply it to limited angle tomographic data.

B. Reconstruction of the Missing Wedge Data

In this experiment, we will test our methods to reconstruct limited angle data. The incomplete sinogram is shown

in Fig. 4(c, top) where the missing data consist of 2 symmetric ramp-shaped cut-offs. FBP reconstruction of the missing wedge data is presented in Fig. 4(c, bottom). The image is very noisy, some streak artifacts are present due to missing information, and the loss of resolution in the central part of the phantom is evident. Due to the presence of abrupt features in the sinogram, streak and shadow artifacts are prominent in images [4]–[6].

The limited angle reconstruction problem is not uniquely defined due to the missing angles and also can be severely ill-conditioned due to limited projection data. Traditional regularization approaches are not able to fully alleviate all difficulties related to reconstruction of such data. In order to suppress aliasing artifacts, one can employ Fourier space filling procedures [6], directional TV regularization [4] or a combination of sparsifying penalties [5]. However, since directional inconsistency is present in the data term the isotropic regularization penalties are not suitable [4]. Alternatively, the negative effects of the missing wedge data can be minimized by using a data misfit that is more forgiving to severe data discrepancies.

Similar to the previous experiment, we estimated thoroughly all regularization parameters before performing comparisons and presenting results. In Fig. 7 (top) we show regularization parameters optimization curves. In Fig. 7 (middle) we performed an experiment similar to one in Fig. 5 where ten noise realizations were applied to the data to estimate Δ_1 errors with respect to CoV values. One can see that higher CoV values belong to the LS and Student-TV methods, while the LS-TV and GH-TV methods have lower CoV. However, the levels of Δ_1 error are higher for all methods except the Student-TV method. In Fig. 7 (bottom) we demonstrate convergence plots for the LS, LS-TV, and the Student-TV methods. One can see that the Student-TV method shows an impressive performance reaching the lowest error $\Delta_1 = 10.9$. However the convergence plot shows that the method must be stopped prematurely to avoid divergence.

In Fig. 8 (top row) image reconstructions from limited angle data are presented, the corresponding image residuals $(\hat{\mathbf{x}} - \mathbf{x}_k)^2$ are shown in the middle row, and in the bottom row the sinogram residuals are shown. The reconstructed image using the LS-TV method (see Fig. 8(a)) is severely contaminated by shadow (darker areas) and streak and rings artifacts. The LS-TV reconstruction is able to mitigate noise and some streak artifacts, however the shadowing remains and all other artifacts are present. The GH-TV reconstruction provides visually similar reconstruction to LS-TV except ring artifacts. The image reconstructed using the Student's t misfit (see Fig. 8(d)) has much less shadow influence in the central part of the phantom and other artifacts have been removed to a degree. Notably, the central vertical shadow which affects the LS-TV and GH-TV reconstructions is damped significantly with the Student-TV method, and the error is substantially lower. Additionally, the SSIM values are distinct for the methods with the significant advantage of the Student-TV method.

V. REAL DATA RECONSTRUCTION

In this section we apply the proposed methods to real neutron data obtained at the cold neutron imaging Beamline ICON at the
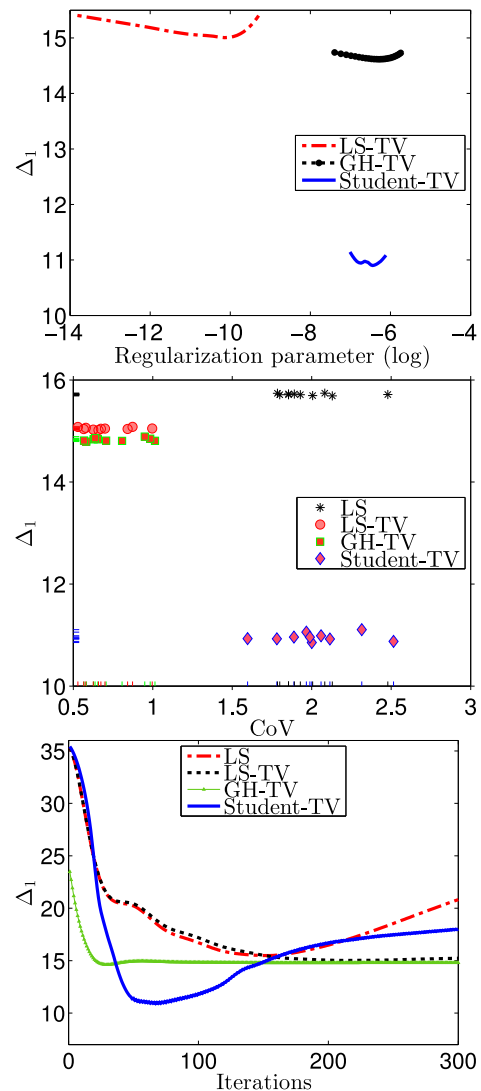

Fig. 7. **Top:** optimization curves to establish the optimal values for regularization parameters for each method for the missing wedge dataset (see Fig. 4 (c, top)). **Middle:** 10 realizations of Poisson noise have been applied to the missing wedge projection data. Resulting values of Δ_1 against CoV are given for the optimized regularization parameters. Notably, the Student-TV reconstructions have higher CoV that LS-TV and GH-TV, yet significantly lower bias in terms of Δ_1 error. **Bottom:** Convergence plots for different IIR methods using the missing wedge dataset (see Fig. 4(c, top)).

SINQ spallation neutron source at the Paul Scherrer Institute, Switzerland [39]. Neutron tomography (nCT) uses neutrons instead of X-rays by interacting with atomic nuclei, and the attenuation properties of materials are different compared with X-rays. For instance, water is highly attenuated and lead becomes highly translucent in a neutron beam.

We chose this particular nCT dataset because it contains different types of errors. In general, nCT has a worse spatial resolution than X-ray systems; the flux is lower and the nCT projections are strongly influenced by scatter. The zingers are caused by gammas (from neutron absorption) hitting the detector. The data are also significantly contaminated by ring artifacts. In Fig. 9 we show a noisy sinogram with various artifacts of porous basalt rock beads packed in a 25 mm aluminium
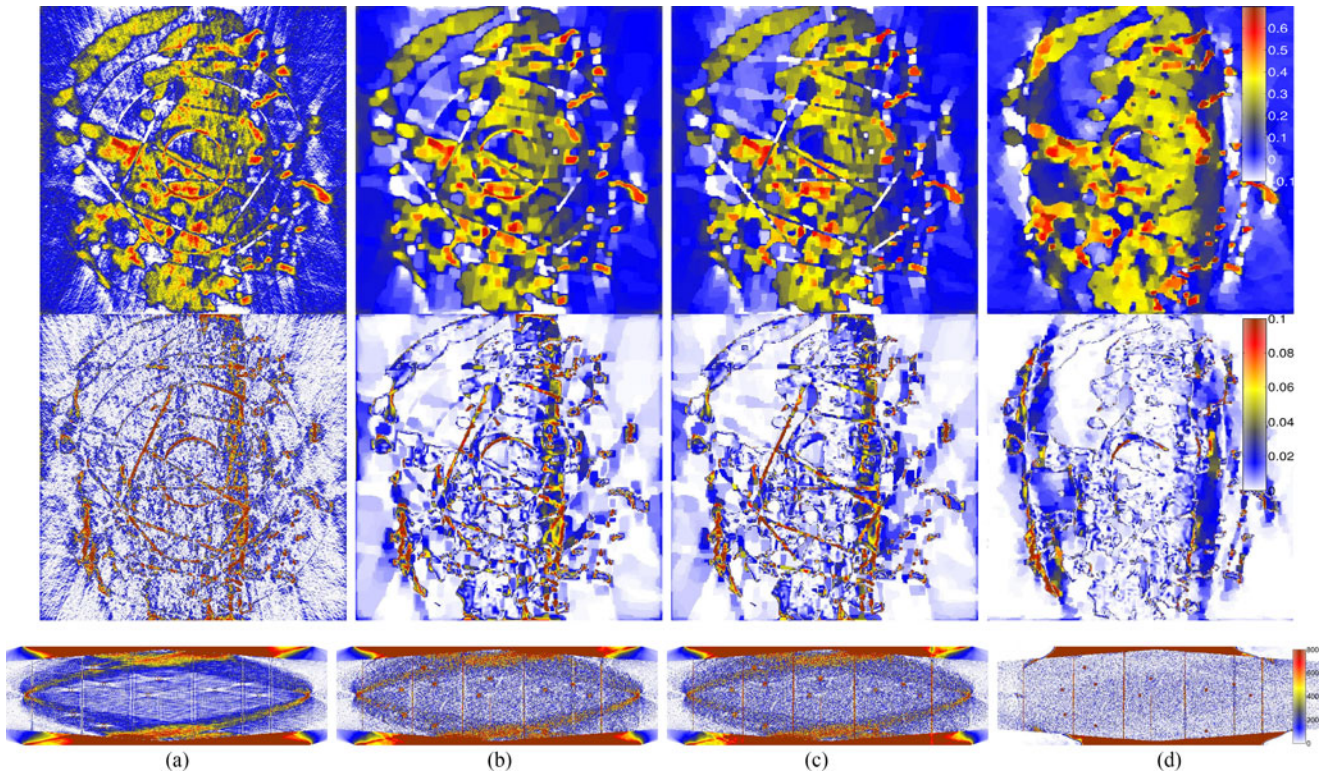
Fig. 8. **Top row:** reconstructed images for the LS, LS-TV, GH-TV and Student-TV methods using the corrupted by artifacts, noisy missing wedge sinogram (see Fig 4 (c, top)). **Middle:** residual $(\hat{\mathbf{x}} - \mathbf{x}_k)^2$ in image space. **Bottom:** residual $(\mathbf{A}\mathbf{x}_k - \mathbf{b})^2$ in sinogram space, where $\mathbf{b}$ is a corrupted sinogram. The Student-TV reconstruction provides the best resolution and the absence of streak and ring artifacts (the central strong ring again is only partially suppressed). (a) LS, $\Delta_1 \backslash \Delta_2 = \mathbf{15.8 \backslash 0.36}$, (b) LS-TV, $\Delta_1 \backslash \Delta_2 = \mathbf{15.3 \backslash 0.53}$, (c) GH-TV, $\Delta_1 \backslash \Delta_2 = \mathbf{15.1 \backslash 0.56}$, (d) Stud-TV, $\Delta_1 \backslash \Delta_2 = \mathbf{10.9 \backslash 0.67}$.
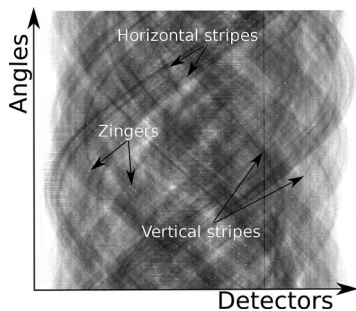

Fig. 9. Neutron source sinogram obtained from scanning a 25 mm aluminium cylindrical container filled with basalt rock beads [40]. Multiple artifacts are present.

cylindrical container (more information about the experiment in [40]). Data consists of 500 projections for reconstruction of 700×700 pixels image slice.

We reconstruct the sinogram in Fig. 9 using the LS, LS-TV, GH-TV, and the Student-TV methods. In order to establish optimal reconstruction parameters, we perform multiple reconstructions for each method and assess reconstructed images and their segmentations using Otsu's method [43] visually. Here we also run 300 outer iterations and 20 inner TV iterations.

In Fig. 10 we present a slice of the reconstructed volume (top row), the zoomed central regions of the reconstructed images are shown in the second row, the segmented images using Otsu's method [43] are in the third row, and the sinogram residuals

$(\mathbf{A}\mathbf{x}_k - \mathbf{b})^2$ are in the bottom row. As expected, the LS reconstruction is very noisy, contains many streaks and also several rings (one major ring is very prominent and sharp). Note that the sharp vertical stripe (related to the major ring artifact) is absent in the sinogram residual. The result of segmentation is unusable. Noise is significantly reduced by the LS-TV method, however all major artifacts remain. The sinogram residual shows a weak presence of the big stripe, therefore the corresponding ring artifact is still sharply defined in the reconstructed and segmented images. The ring modelling in the GH-TV algorithm helps to remove rings, however streaks are present in the image and in its segmentation. The sinogram residual shows more compact and sharp vertical stripes that have been modelled and removed from the reconstruction.

The Student-TV algorithm delivers the most promising results while completely removing streaks and major rings. Notably the sinogram residual for the Student-TV method can tell us more about the quality of the reconstructed image. Both vertical and horizontal stripes are sharply defined. Horizontal stripes may not directly affect the visual perception of the reconstruction, however they are errors and must be excluded. Additionally, streak artifacts have been removed from the reconstruction and can be seen as distinctive dots in the sinogram residual. The segmented image for the Student-TV method is visually satisfactory and the best among all compared methods. It does not contain any strong errors due to the rings and streaks as other segmented images do.
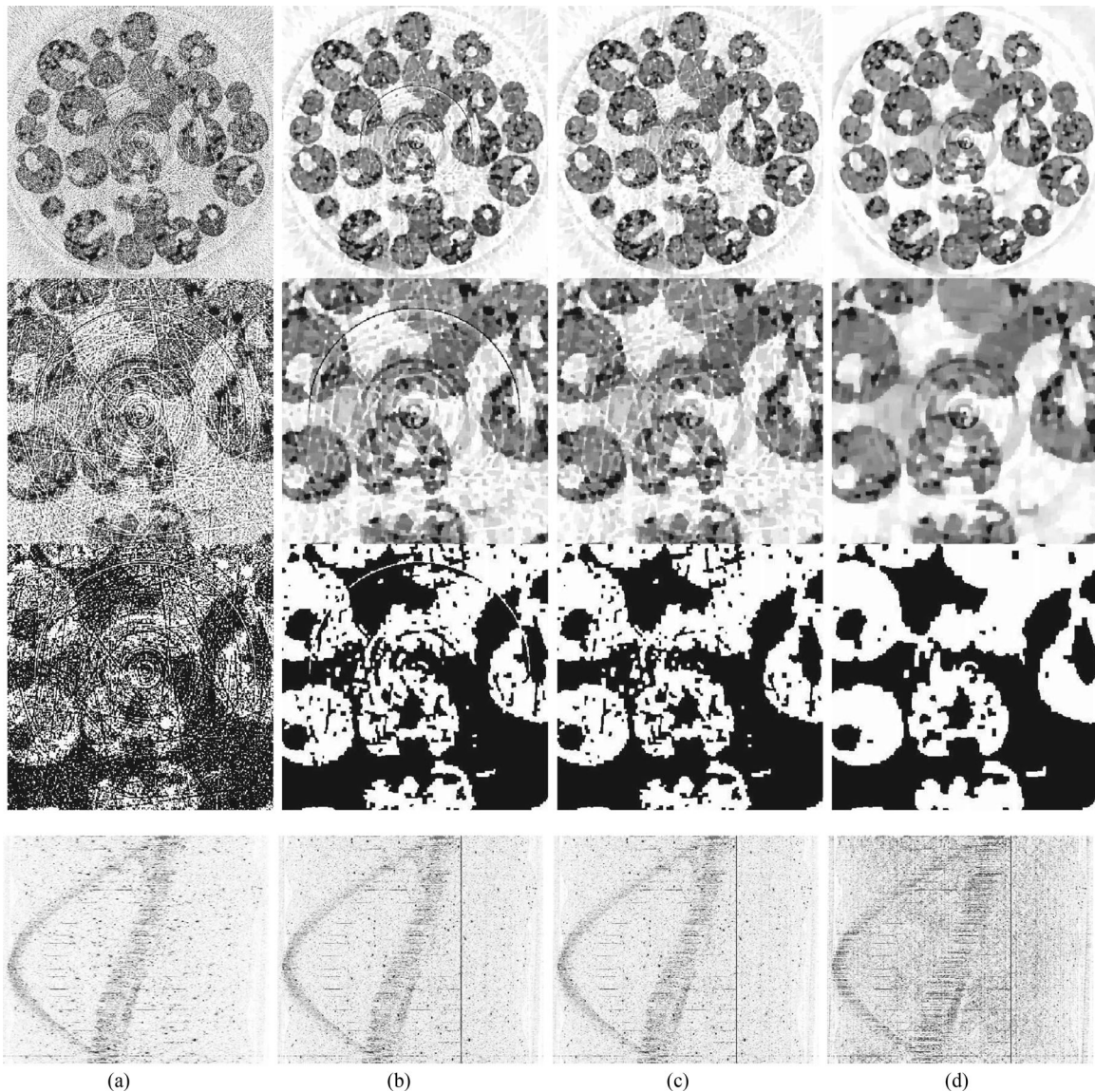
Fig. 10. First row: the reconstructed images for the LS (a), LS-TV (b), GH-TV (c) and Student-TV (d) methods using the sinogram in Fig 9. **Second:** zoomed into the central slice region of reconstructed images. **Third:** segmented images using global image threshold using Otsu's method. **Fourth:** residual $(\mathbf{A}\mathbf{x}_k - \mathbf{b})^2$ in the sinogram space. Note the amount of streak artifacts and rings in the LS reconstruction. The Student-TV method successfully removes artifacts.

VI. DISCUSSION AND CONCLUSION

In this manuscript, we present a novel MBIIR technique where the conventional data fidelity term is replaced by the Student's t distribution. The Student's t misfit is taken from the group of statistical estimators which provide a robust recovery of a signal against large errors (outliers) in data. For many imaging experiments in computed tomography (especially a synchrotron and neutron spallation sources), such outliers are inevitably present in the collected data. We demonstrated that the Student's t based method can handle various types of artifacts. The proposed data term is solely based on a robust loss function and therefore different from the existing model-based reconstruction algorithms [15], [16]. Since the proposed method and the GH-based method [16] can be put into one optimization framework, the comparison is straightforward. Comparison with the TIMBIR algorithm [15] is more complicated – due to differences in the optimization approaches, regularization terms, and projection models – and is left for future research.

We have shown that using TV regularization penalty with the Student's t misfit one can successfully reconstruct from under-sampled noisy data with various errors. The notorious limited-angle artifacts can be noticeably alleviated using the Student-TV algorithm. However, more advanced prior models can further improve quality of reconstructions, such as directional regularizers [4], higher-order terms [5], wavelet transforms [5], trained dictionaries [16], and patch-based priors [40].

In this work we did not attempt to deal with any possible effects of nonconvexity of the Student's t penalty. In practice we noticed a stable behaviour of the Student-TV algorithm converging to a (local) minimum. To alleviate some risks of converging to an undesirable local minimum, one can use a local quadratic approximation [41]. The proposed algorithm can be made more

robust to coherent outliers, such as rings, by applying the Student's t penalty to groups of data, in a similar fashion as the GH approach.

ACKNOWLEDGMENT

REFERENCES

[1] T. M. Buzug, *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT*. New York, NY, USA: Springer, 2008.

[2] C. L. Epstein, *Introduction to the Mathematics of Medical Imaging*. Philadelphia, PA, USA: SIAM, 2008.

[3] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. New York, NY, USA: IEEE Press, 1999.

[4] Z. Chen, X. Jin, L. Li, and G. Wang, "A limited-angle CT reconstruction method based on anisotropic TV minimization," *Phys. Med. Biol.*, vol. 58, no. 7, pp. 2119–2149, 2013.

[5] D. Kazantsev, E. Ovtchinnikov, P. J. Withers, W. R. B. Lionheart, and P. D. Lee, "Sparsity seeking total generalized variation for undersampled tomographic reconstruction," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2016, pp. 731–734.

[6] J. Lim *et al.*, "Comparative study of iterative reconstruction algorithms for missing cone problems in optical diffraction tomography," *Opt Express*, vol. 23, no. 13, pp. 16933–16948, 2015.

[7] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-posed Problems*. Great Falls, MT, USA: Winston, 1977.

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

[9] J. S. Jørgensen and E. Y. Sidky, "How little data is enough? Phase-diagram analysis of sparsity-regularized X-ray computed tomography," *Philos. Trans. A Math. Phys. Eng. Sci*, vol. 373, no. 2043, 2015, Art. no. 20140387.

[10] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1/4, pp. 259–68, 1992.

[11] J. Hsieh, "Adaptive streak artifact reduction in computed tomography resulting from excessive x-ray photon noise," *Med. Phys.*, vol. 25, no. 11, pp. 2139–2147, 1998.

[12] B. Münch, P. Trtik, F. Marone, and M. Stampanoni, "Stripe and ring artifact removal with combined wavelet-Fourier filtering," *Opt. Express*, vol. 17, no. 10, pp. 8567–8591, 2009.

[13] S. Titarenko, V. Titarenko, A. Kyrieleis, and P. J. Withers, "A ring artifact suppression algorithm based on a priori information," *Appl. Phys. Lett.*, vol. 95, no. 7, 2009, Art. no. 071113.

[14] J. Sijbers and A. Postnov, "Reduction of ring artefacts in high resolution micro-CT reconstructions," *Phys. Med. Biol.*, vol. 49, no. 14, pp. N247–N253, 2004.

[15] K. A. Mohan *et al.*, "TIMBIR: A method for time-space reconstruction from interlaced views," *IEEE Trans. Comput. Imag.*, vol. 1, no. 2, pp. 96–111, 2015.

[16] P. Paleo and A. Mirone, "Ring artifacts correction in compressed sensing tomographic reconstruction," *J. Synchrotron Radiat.*, vol. 22, no. 5, pp. 1268–1278, 2015.

[17] F. Bleichrodt, T. van Leeuwen, and K. J. Batenburg, "Robust artefact reduction in tomography using Student's t data fitting," in *Proc. Fully3D Conf.*, Newport, RI, USA, 2015, pp. 395–398.

[18] H. Erdogan and J. A Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–2851, 1999.

[19] M. Beister, D. Kolditz, and W. A. Kalender, "Iterative reconstruction methods in X-ray CT," *Phys. Med.*, vol. 28, no. 2, pp. 94–108, 2012.

[20] C. Melot, Y. Boursier, J. F. Aujol, and S. Anthoine, "Some proximal methods for Poisson intensity CBCT and PET," *Inverse Probl. Imag.*, vol. 6, no. 4, pp. 598–565, 2012.

[21] D. Kazantsev *et al.*, "A novel technique to incorporate structural prior information into multi-modal tomographic reconstruction," *Inverse Probl.*, vol. 30, no. 6, 2014, Art. no. 065004.

[22] J. B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "A recursive filter for noise reduction in statistical iterative tomographic imaging," *Proc. SPIE*, vol. 6065, 2006, Art. no. 60650X.

[23] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 534–548, Feb. 1993.

[24] K. L. Lange, R. J. Little, and J. M. Taylor, "Robust statistical modeling using the t distribution," *J. Amer. Statist. Assoc.*, vol. 84, no. 408, pp. 881–896, 1989.

[25] A. Y. Aravkin and T. van Leeuwen, "Estimating nuisance parameters in inverse problems," *Inverse Probl.*, vol. 28, no. 11, 2012, Art. no. 115016.

[26] P. L. Combettes and J. C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point Algorithms for Inverse Problems in Science and Engineering*. New York, NY, USA: Springer, 2011, pp. 185–212.

[27] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, 2011.

[28] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[29] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

[30] P. J. Huber, *Robust Statistics*. Berlin, Germany: Springer, 2011.

[31] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.

[32] J. Nocedal and S. Wright, *Numerical Optimization*. New York, NY, USA: Springer, 2006.

[33] P. Ochs, Y. Chen, T. Brox, and T. Pock, "iPiano: Inertial proximal algorithm for nonconvex optimization," *SIAM J. Imag. Sci.*, vol. 7, no. 2, pp. 1388–1419, 2014.

[34] J. Kaipio and E. Somersalo, "Statistical inverse problems: Discretization, model reduction and inverse crimes," *J. Comput. Appl. Math.*, vol. 198, no. 2, pp. 493–504, 2007.

[35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[36] W. J. Palenstijn, K. J. Batenburg, and J. Sijbers, "Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs)," *J. Struct. Biol.*, vol. 176, no. 2, pp. 250–253, 2011.

[37] F. Bleichrodt, T. van Leeuwen, W. J. Palenstijn, W. van Aarle, J. Sijbers, and K. J. Batenburg, "Easy implementation of advanced tomography algorithms using the ASTRA toolbox with Spot operators," *Numer. Algorithms*, vol. 71, no. 3, pp. 673–697, 2016.

[38] D. Kazantsev, "MATLAB code to reproduce presented results," 2017. [Online]. Available: https://ccpforge.cse.rl.ac.uk/gf/project/ccpi_itr/

[39] A. P. Kaestner *et al.*, "Recent developments in neutron imaging with applications for porous media research," *Solid Earth*, vol. 7, no. 5, pp. 1281–1292. 2016.

[40] D. Kazantsev *et al.*, "4D-CT reconstruction with unified spatial-temporal patch-based regularization," *Inverse Probl. Imag.*, vol. 9, pp. 447–467, 2015.

[41] R. F. Barber and E. Y. Sidky, "MOCCA: Mirrored convex/concave optimization for nonconvex composite functions," *J. Mach. Learn. Res.* vol. 17, pp. 1–51, 2016.

[42] A. Y. Aravkin, J. V. Burke, and G. Pillonetto, "Robust and trend-following Kalman smoothers using student's T," *IFAC Proc. Volumes*, vol. 45, no. 16, p. 1215–1220, 2012.

[43] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285/296, pp. 23–27, 1975.

Daniil Kazantsev received the Ph.D. degree from the Institute of Theoretical and Applied Mechanics, Novosibirsk, Russia, in 2008, while working in the Group of Computerized Tomography. As a PDRA, he was with the Centre for Medical Imaging Computing, University College London, London, U.K. He is currently a PDRA in the School of Materials (MXIF), University of Manchester, Manchester, U.K. He is interested in solving ill-posed inverse problems in computed tomography.
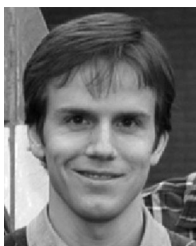
Folkert Bleichrodt received the Ph.D. degree in applied mathematics from the University of Leiden, Leiden, The Netherlands, in 2015. His Ph.D. research was carried out at the Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands. After his Ph.D., he was a Postdoctoral Researcher with the CWI. His main research interest focuses on tomographic reconstruction methods and algorithms for the reduction of reconstruction artifacts.

Philip J. Withers received the Ph.D. degree in metallurgy from University of Cambridge, Cambridge, U.K., where he held a lectureship until his appointment as a Professor in the University of Manchester in 1998. In 2008, he established the Henry Moseley X-ray Imaging Facility. In 2012, he became the Inaugural Director of the BP International Centre for Advanced Materials. His research interests include establishing correlative tomography linking x-ray and electron microscopy to provide multifaceted, multiscale information.

Tristan van Leeuwen received the B.Sc. and M.Sc. degrees in computational science from Utrecht University, Utrecht, The Netherlands, and the Ph.D. degree in geophysics from Delft University, Delft, The Netherlands, in 2010. After spending some time as a Postdoctoral Researcher with the University of British Columbia, Vancouver, BC, Canada, and the Centrum Wiskunde & Informatica, Amsterdam, the Netherlands, he returned to Utrecht University in 2014 as an Assistant Professor at the Mathematical Institute. His research interests include inverse problems, computational imaging, tomography, and numerical optimization.

Kees Joost Batenburg received the M.Sc. degree in mathematics and the M.Sc. degree in computer science from the University of Leiden, Leiden, The Netherlands, in 2002 and 2003, respectively, and the Ph.D. degree in mathematics in 2006. He is currently a Senior Researcher at Centrum Wiskunde & Informatica, Amsterdam, The Netherlands, where he performs research in the field of computational imaging. He also has a part-time professorship at the University of Leiden.

Anders Kaestner received the Ph.D. degree in signal processing from the Chalmers Institute of Technology, Gothenburg, Sweden. He is a Beamline Scientist at the Cold Neutron Imaging Beamline (ICON), Paul Scherrer Institute, Switzerland. In 2002, he started working on imaging and image processing in the Chalmers Institute of Technology. As a postdoc, he continued this topic working with image processing tasks for soil science at the Swiss Federal Institute of Technology, Zurich, Switzerland. His research interests include neutron imaging, computed tomography, and image enhancement.

Peter D. Lee is the Acting Director of the Research Complex at Harwell, Didcot, U.K., a center focused upon developing techniques at the physical-life interface. His research focuses on investigating the evolution of microstructures using x-ray imaging and computational simulation techniques, with applications from solidification to biomaterials.