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, K. 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—X-ray CT, neutron tomography, ring artifacts, zingers, limited angle regularization, proximal point, robust statistics

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]:

\[
\Lambda_i = I_{0,i} \exp \left( - \int_{L_i} \mu(r) dl \right), \ i = 1, \ldots, m, (1)
\]

where \( r \in \mathbb{R}^2 \) is the spatial position, \( \mu \) is the attenuation coefficient of the object under investigation, \( I_{0,i} \) is the intensity of the incoming beam, \( \Lambda_i \) 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:

\[
\int_{L_i} \mu(r) dl = - \ln \left( \frac{\Lambda_i}{I_{0,i}} \right), \ i = 1, \ldots, m. \quad (2)
\]

For X-rays, the mentioned assumptions can be fulfilled to a degree; however for neutrons, the first two are usually not satisfied; the neutron beam is normally polychromatic and neutrons are scattered coherently and non-coherently.

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 \( a_{ij} \).

Representing the attenuation coefficient on a square grid of \( n \) pixels: \( \{\chi_j(r)\}_{j=1}^n \), i.e., \( \mu(r) = \sum_{j=1}^n x_j \chi_j(r) \),
we obtain a discrete, linear, forward model $b = Ax$, where $x \in \mathbb{R}^n$ is a vector with attenuation coefficients $x_j$, $b \in \mathbb{R}^m$ is a vector with the log-corrected measurements, $b_i = -\ln \left( \frac{\sum_j x_{ij}}{b_{0j}} \right)$, and $A \in \mathbb{R}^{m \times n}$ is the projection matrix with elements $a_{ij} = \int_{L_j} \chi_j(r) dl$ 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 $\theta$. 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 $b_i$.

![Fig. 2: Schematic figure of the parallel beam acquisition set-up to obtain a 2D sinogram from the set of 1D projections.](image)

When the measurements are densely sampled over a full angular range of $\pi$ radians and not noisy, one can recover the attenuation coefficients using direct reconstruction methods such as Filtered Back Projection (FBP) [2]. 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 generalizations 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 these 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 IV the reconstruction problem with various robust penalties is formulated. Section V presents an optimization framework. Numerical results with a synthetic phantom are given in section VI and real data results in V. Section VII concludes the manuscript.
II. Problem formulation and robust data misfit

The measured transmission data \( I_0 \) 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 \( \Lambda_i \) modelled as a Poisson random variable with parameter \( \bar{\Lambda}_i \) \cite{20,23}:

\[
\Lambda_i \sim \text{Poiss}\{\bar{\Lambda}_i\} = \text{Poiss}\{I_0,e^{-|\Lambda_x|}\}, \quad i = 1, \ldots, m \quad (3)
\]

One can approach the reconstruction problem from the Bayesian perspective by employing maximum a posteriori (MAP) estimate \cite{15,18,20,22,23}:

\[
\arg\min_x \sum_{i=1}^{m} I_0,i[|\Lambda_x|]_i + \Lambda_i e^{-|\Lambda_x|}_i] + L_{\text{prior}}(x), \quad (4)
\]

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

\[
\hat{x} = \arg\min_x \frac{1}{2} (\Lambda x - b)^\top \Lambda (\Lambda x - b) + L_{\text{prior}}(x), \quad (5)
\]

where \( \Lambda \in \mathbb{R}^{m \times m} \) 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 \cite{3} which is used for many X-ray tomography cases \cite{15,22,23}. The matrix \( \Lambda \) effectively gives a higher weight to projections with a higher photon-count. In the remainder of the paper we absorb these weights in their corresponding regularized optimization problem in the general form:

\[
\min_x f(\Lambda x - b) + \beta g(x), \quad (6)
\]

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 \( \beta \) 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 \( g(\cdot) = \| \cdot \|_L^2 \), which corresponds to a Gaussian prior distribution for \( x \) with mean zero and covariance \( \mathbf{L}^\top \mathbf{L} \). The optimization problem \( (6) \) now has the following closed form solution \( \hat{x} = \mathbf{H}^{-1} \mathbf{A}^\top \mathbf{b} \), where \( \mathbf{H} = [\mathbf{A}^\top \mathbf{A} + \beta \mathbf{L}^\top \mathbf{L}] \). It is infeasible to precompute the inverse of \( \mathbf{H} \) exactly for large-scale problems, therefore various approximations can be used in practice \cite{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 \cite{23}. In this work, we will be using first-order optimization approaches which avoid explicit calculation of \( \mathbf{H}^{-1} \) altogether \cite{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 \( \ell_1 \)-norm \cite{31}:

\[
f(r) = \min_s \frac{1}{2} \|s - r\|_2^2 + \lambda \|s\|_1. \quad (7)
\]

This optimization problem has a closed-form solution \( \hat{s} \) with elements

\[
\hat{s}_i = \mathcal{S}_\lambda(r_i), \quad (8)
\]

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

\[
f(r) = \sum_{i=1}^{m} \rho_H(r_i), \quad (10)
\]

where

\[
\rho_H(r) = \begin{cases} 
  \lambda |r| - \frac{1}{2} \lambda^2 & \text{if } |r| \leq \lambda \\
  \frac{1}{2} r^2 & \text{if } |r| > \lambda.
\end{cases} \quad (11)
\]

The gradient is given by

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

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 \( \rho_H(r) \) and its derivative \( \rho_H'(r) \). 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]

\[ f(r) = \min_s \frac{1}{2} \| s - B^T r \|^2_2 + \lambda \| s \|_1, \tag{13} \]

where \( B = (\mathbf{I}_{m_1} \otimes 1_{m_2})/\sqrt{m_2} \) with \( B^T B = \mathbf{I}_{m_1} \). The reasoning behind this formulation 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 \( B \) sums the residuals in each group so that \( s \) is a vector of length \( m_2 \).

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

\[ \hat{s}_i = S_\lambda((B^T r)_i), \]

which can be written in terms of the conventional Huber misfit \( \rho_H \) as

\[ f(r) = \sum_{i=1}^m \rho_H((B^T r)_i). \tag{14} \]

The gradient of this objective can be expressed as

\[ \nabla f(r) = BWB^T r, \]

where \( W \) 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

\[ \rho_{\text{StudentsT}}(r) = \log(1 + (r/\sigma)^2), \]

where \( \sigma \) 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

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

\[ \nabla f(r) = W r, \tag{16} \]

where \( W \) is a diagonal matrix with elements

\[ w_{ii} = \frac{2}{\sigma^2 + r_{ii}^2}. \tag{17} \]

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 \( \sigma \) can be formulated as a scalar optimization problem [25]

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

For a given residual this problem is easily solved using a scalar minimization algorithm such as Newton’s method or Golden-ratio search [22]. The resulting misfit function \( f(r) \) is smooth and can be used in conjunction with any gradient-based optimization method [25]. The gradient of the objective w.r.t. \( r \) is given by \( W r \), with \( W \) as in (17), evaluated at the optimal \( \sigma \) [25]. Alg. 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 \( r \)
Output: \( f(r) \) and \( \nabla f(r) \)

compute \( \hat{\sigma} = \arg \min_\sigma \sum_{i=1}^m \log(1 + (r_i/\sigma)^2) \)

\[ f(r) = m \log(\pi \hat{\sigma}) + \sum_{i=1}^m \log(1 + (r_i/\hat{\sigma})^2) \]

\[ \nabla f(r) = Wr, \text{ with } w_{ii} = \frac{2}{\hat{\sigma}^2 + r_{ii}^2}. \]

III. Tomographic image reconstruction

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

\[ x_{k+1} = \text{prox}_{\beta g}(x_k - L^{-1} \cdot A^T \nabla f(A x_k - b)), \]

where \( \text{prox}_{\beta g}(y) = \arg \min_x g(x) + \frac{1}{2\beta} \| x - y \|^2_2 \), and \( L \) is the Lipschitz constant of \( \nabla f \). 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: \( g(x) = \sum_{i=1}^m \sqrt[D_1]{x_i^2} + [D_2 x_i]^2 \), 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 Alg. 2, we present four different methods: LS (when \( \nabla f(r) = r \), \( \beta = 0 \)), LS-TV (\( \nabla f(r) = r \), \( \beta > 0 \)), GH-TV (Group-Huber fidelity with \( \nabla f(r) \) defined as in (14)), and Student-TV (Student’s t fidelity with \( \nabla f(r) \) evaluated by Alg. 1). Notably, the statistical weighting \( A \) 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 Alg. 2 that includes an inertia term \( x_k - x_{k-1} \) 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.

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).
where $\hat{\text{SSIM}}$ is given by:

$$\text{SSIM} = \frac{(2\mu_x\mu_y + C_1)(\sigma_x\sigma_y + C_2)}{\mu_x^2 + \mu_y^2 + \sigma_x^2 + \sigma_y^2 + C_2}.$$  \hspace{1cm} (20)

The second metric is the structural similarity index (SSIM) \cite{Wang2004}. The SSIM value equals 1 if images are identical and visually perceived change in the structural information. The SSIM is different quality metric from the NRMSE, 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) \cite{Rohling2017} and defined as:

$$\text{CoV}(x_\omega) = \frac{1}{\text{ROI}} \sum_{i \in \text{ROI}} \sqrt{\frac{(x_{i,\omega} - \bar{x})^2}{(\Omega - 1)\bar{x}^2}}, \hspace{1cm} (21)$$

where $\mu$ and $\sigma$ are the average and the standard deviation of a patch centered around a pixel respectively (we used an $8 \times 8$ patch). The variable $\Omega$ denotes the cross-correlation and $C_{1,2}$ are small constants to avoid division by zero \cite{Vogel2002}. The SSIM is different quality metric from the NRMSE \cite{Laine2019}, 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) \cite{Rohling2017} and defined as:

$$\text{CoV}(x_\omega) = \frac{1}{\text{ROI}} \sum_{i \in \text{ROI}} \sqrt{\frac{(x_{i,\omega} - \bar{x})^2}{(\Omega - 1)\bar{x}^2}}, \hspace{1cm} (21)$$

where $\mu$ and $\sigma$ are the average and the standard deviation of a patch centered around a pixel respectively (we used an $8 \times 8$ patch). The variable $\Omega$ denotes the cross-correlation and $C_{1,2}$ are small constants to avoid division by zero \cite{Vogel2002}. The SSIM is different quality metric from the NRMSE \cite{Laine2019}, 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) \cite{Rohling2017} and defined as:

$$\text{CoV}(x_\omega) = \frac{1}{\text{ROI}} \sum_{i \in \text{ROI}} \sqrt{\frac{(x_{i,\omega} - \bar{x})^2}{(\Omega - 1)\bar{x}^2}}, \hspace{1cm} (21)$$

where $\mu$ and $\sigma$ are the average and the standard deviation of a patch centered around a pixel respectively (we used an $8 \times 8$ patch). The variable $\Omega$ denotes the cross-correlation and $C_{1,2}$ are small constants to avoid division by zero \cite{Vogel2002}. The SSIM is different quality metric from the NRMSE \cite{Laine2019}, 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) \cite{Rohling2017} and defined as:

$$\text{CoV}(x_\omega) = \frac{1}{\text{ROI}} \sum_{i \in \text{ROI}} \sqrt{\frac{(x_{i,\omega} - \bar{x})^2}{(\Omega - 1)\bar{x}^2}}, \hspace{1cm} (21)$$

where $\mu$ and $\sigma$ are the average and the standard deviation of a patch centered around a pixel respectively (we used an $8 \times 8$ patch). The variable $\Omega$ denotes the cross-correlation and $C_{1,2}$ are small constants to avoid division by zero \cite{Vogel2002}. The SSIM is different quality metric from the NRMSE \cite{Laine2019}, 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) \cite{Rohling2017} and defined as:

$$\text{CoV}(x_\omega) = \frac{1}{\text{ROI}} \sum_{i \in \text{ROI}} \sqrt{\frac{(x_{i,\omega} - \bar{x})^2}{(\Omega - 1)\bar{x}^2}}, \hspace{1cm} (21)$$
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 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 $\beta$ with respect to NRMSE ($\Delta_1$). Notably, regularization parameter $\lambda$ for the GH-TV method has been found before the experiment.

In Fig. 5 (top) one can see how the optimal regularization parameter $\beta$ has been chosen based on $\Delta_1$ values. Each curve consists of 30 points with linearly progressing regularization parameter values. The global minimum of $\Delta_1$ can be established for each method and the corresponding $\beta$ value is selected. It can be seen that the minimum $\Delta_1$ 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 $\Delta_1$ 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 $\Delta_1$ 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 $\Delta_1$ errors. In the middle row we show image residuals $\hat{x} - x_k$, 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 $(Ax_k - b)^2$, where $b$ 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
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 $(\hat{x} - x_k)^2$ in image space. **Bottom:** residual $(Ax_k - b)^2$ in sinogram space, where $b$ 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).

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.

Additionally, the SSIM values are given for all methods in Fig. 6. There is a big gap in $\Delta_2$ 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 $\sigma$ parameter in Alg. 1.

<table>
<thead>
<tr>
<th>Regularization parameter (log)</th>
<th>Iterations</th>
<th>CoV</th>
<th>$\Delta_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS</td>
<td>LS-TV</td>
<td>GH-TV</td>
<td>Student-TV</td>
</tr>
<tr>
<td>0.03 sec.</td>
<td>0.06 sec.</td>
<td>0.07 sec.</td>
<td>0.16 sec.</td>
</tr>
</tbody>
</table>

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. 5(middle) we performed an experiment similar to one in Fig. 5 where ten noise realizations were applied to the data to estimate $\Delta_1$ 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 $\Delta_1$ 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 $(\hat{x} - x_k)^2$ 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.
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 $(\hat{x} - x_k)^2$ in image space. **Bottom:** residual $(Ax_k - b)^2$ in sinogram space, where $b$ 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).

\[(a) \text{LS, } \Delta_1 \| \Delta_2 = 15.8 \pm 0.36 \quad (b) \text{LS-TV, } \Delta_1 \| \Delta_2 = 15.3 \pm 0.53 \quad (c) \text{GH-TV, } \Delta_1 \| \Delta_2 = 15.1 \pm 0.56 \quad (d) \text{Stud-TV, } \Delta_1 \| \Delta_2 = 10.9 \pm 0.67\]

\[\text{(d) has much less shadow influence in the central part of the phantom and other artifacts have been removed to a degree. Notably, the central vertical shadow which affects the LS-TV and GH-TV reconstructions is damped significantly with the Student-TV method, and the error is substantially lower. Additionally, the SSIM values are distinct for the methods with the significant advantage of the Student-TV method.}\]

\[\text{V. Real data reconstruction}\]

In this section we apply the proposed methods to real neutron data obtained at the cold neutron imaging Beamline ICON at the 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 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 $(\mathbf{A}x_k - \mathbf{b})^2$ 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.

VI. DISCUSSION AND CONCLUSION

In this manuscript, we present a novel MBIIIR 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 undersampled 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

This work was made possible by the facilities and support provided by the Research Complex at Harwell, funded in part by the EPSRC (EP/I02249X/1), as well as the CCPi initiative (EP/M022498/1). This research was also supported by the Netherlands Organisation for Scientific Research (NWO), programmes 639.072.005 and 639.073.506. Networking support was provided by the EXTREMA COST Action MP1207. All data created and used during this research are openly available from the CCPForge website funded by JISC and EPSRC.

REFERENCES


D. Kazantsev, “Matlab code to reproduce presented results”, https://ccpforge.cse.rl.ac.uk/gf/project/ccpi/itr/