TVR-DART: a more robust algorithm for discrete tomography from limited projection data with automated gray value estimation

Xiaodong Zhuge, Willem Jan Palenstijn and Kees Joost Batenburg

Abstract—In this paper, we present a novel iterative reconstruction algorithm for discrete tomography, named Total Variation Regularized Discrete Algebraic Reconstruction Technique (TVR-DART) with automated gray value estimation. This algorithm is more robust and automated than the original DART algorithm, and is aimed at imaging of objects consisting of only a few different material compositions, each corresponding to a different gray value in the reconstruction. By exploiting two types of prior knowledge of the scanned object simultaneously, TVR-DART solves the discrete reconstruction problem within an optimization framework inspired by compressive sensing to steer the current reconstruction toward a solution with the specified number of discrete gray values. The gray values and thresholds are estimated as the reconstruction improves through iterations. Extensive experiments from simulated data, experimental μCT and electron tomography datasets show that TVR-DART is capable of providing more accurate reconstruction than existing algorithms under noisy conditions from a small number of projection images and/or from a small angular range. Furthermore, the new algorithm requires less effort on parameter tuning compared to the original DART algorithm. With TVR-DART, we aim to provide the tomography society with an easy-to-use and robust algorithm for discrete tomography.

Index Terms—Discrete tomography, Total variation, compressive sensing, image reconstruction, sparse reconstruction, prior knowledge

I. INTRODUCTION

TOMOGRAPHY is a powerful technique for investigating the three-dimensional (3D) structures of objects by utilizing penetrating waves or particles, and has a wide range of applications such as computed tomography (CT) [1] and electron tomography (ET) [2]. Projection images of an object are acquired over a range of rotation angles, and a mathematical procedure known as tomographic reconstruction is required to recover the 3D object information from its 2D projections. Due to its large influence on the outcome of the complete tomography experiment, reconstruction algorithms have been a subject under intensive research [3, 4].

In most practical applications, it is extremely advantageous if the reconstruction algorithm can still produce accurate results using a small number of projection images under moderate/high noise levels. For example, medical CT uses ionizing radiation and reducing the dose to the patient is of high importance. In transmission electron tomography, the electron beam causes damage to the sample during acquisition and cannot penetrate the sample section at high tilt angles. These practical aspects limit the number of projection images, acquired image quality (low dose leads to high noise level), and/or available angular range for reconstruction. Under such conditions, the tomography reconstruction problem is highly underdetermined and there is no unique solution to the inverse problem based on only the acquired data.

This necessitates the full utilization of the prior knowledge we have on the unknown object. Compressive sensing (CS) is one of the concepts under intensive research in recent years [5, 6]. It proves that if an image is sparse in a certain domain, it can be recovered accurately from a small number of measurements with high probability when the measurements satisfy certain randomization properties [7]. Total Variation Minimization (TVmin) can be seen as a special case of CS when the boundary of the object is sparse within the image [8-11]. Discrete tomography (DT) considers another type of prior knowledge where the object is known to consist of a limited number of materials, each producing a constant gray value in the reconstruction [12]. The Discrete Algebraic Reconstruction Technique (DART) is one of the practical algorithms that exploits the discrete nature of the object by alternating iteratively between discretization steps of segmentation based on gray values, and continuous steps of reconstruction on the boundary of the segmented image [13]. DART has been successfully used for reconstructing samples from applications in CT [14] and ET [15, 16].

Despite its superior performance demonstrated in some cases, applying DART in practice is still a challenging and time consuming process. This difficulty is related to several issues: First, DART is a heuristic procedure that requires many parameters to be specified by the user. Figuring out the optimal choices for these parameters for experimental data requires

substantial effort and time for manual tuning. Second, one of the important concepts behind DART is to reduce the number of unknowns by fixing and removing the pixels/voxels within the flat regions of the segmented image from the system of linear equations. This makes the system better determined as the reconstruction improves through iterations. However, fixing the interior regions of the segmented image can misassign a substantial number of pixels in the reconstruction especially when the reconstruction is still not accurate in the early stage. This can push the solution in the wrong direction when the limited projection data also contains a certain mismatch from the model (e.g., image alignment error, nonlinearity in image formation) and/or moderate to high level of noise. Third, DART applies a smoothing filter to the free pixels as a way to even out the fluctuations over the boundary pixels and combat the influence of noise in the projection data. Although this strategy makes sense intuitively, it is hard to predict its effect on the reconstruction under noisy conditions. Last but not least, the reconstruction can be not entirely discrete in practice due to imperfection of the imaging system or the object itself. The hard segmentation in the discrete step of DART imposes a strong constraint on the solution, making it difficult to cope with practical complications.

Efforts have been made to deal with some of the above mentioned problems. For example, an algorithm was proposed that couples DART with a search-based algorithm in order to find the right gray values for the reconstruction [17]. An algorithm known as SDART was proposed to deal with noisy projections by spreading the noise across the whole image domain using a penalty matrix [18].

In this paper, we propose a new iterative reconstruction algorithm, TVR-DART, which can produce more accurate reconstructions than DART under noisy conditions from limited projection data. TVR-DART takes the key concept of DART in terms of steering the solution toward discrete gray values, and incorporates this strategy within a automated optimization framework of compressive sensing. We replace the hard segmentation step of DART with a soft segmentation function which is described with a sum of logistic functions. This smoothes the objective function and allows us to solve the discrete reconstruction and gray value estimation problems alternatingly in a non-convex optimization framework. A total variation term applied on the segmented reconstruction is added to the objective function to combat noise and regularize the reconstruction under extremely limited data conditions. The core idea of TVR-DART is to gently push the gray values of the reconstruction using the soft segmentation function at each iteration, and to continuously update the entire image so that the reconstruction after applying the soft segmentation better matches the projection data and is at the same time sparse in its boundaries between regions with constant discrete gray values. Due to the fact that the ℓ^1-norm is applied on the segmented reconstruction, TVR-DART produces accurate and sharp boundaries in the reconstruction without blurring the solution. The algorithm has been implemented using the open source ASTRA Tomography Toolbox [26], in which the basic forward and backward projection operations are efficiently computed using a Graphical Processing Unit (GPU). A thorough experimental study is provided in this paper to investigate the ability of TVR-DART to reconstruct images from noisy data and from a small number of projection images, and to compare with existing algorithms including Simultaneous Iterative Reconstruction Technique (SIRT), TVmin, and DART. The results show that TVR-DART can produce more accurate reconstructions under difficult practical conditions. Furthermore, we show that the limited number of parameters of the algorithm can be easily configured under different conditions.

This paper is organized as follows. In Section II the mathematical notation and the DART algorithm are briefly introduced. The TVR-DART algorithm is described in detail in Section III. In Section IV, we present the numerical simulations and comparison studies. Results from experimental datasets are discussed in Section V. Section VI concludes this paper.

II. MATHEMATICAL CONCEPTS AND DART ALGORITHM

A. Problem Definition

We treat the tomographic reconstruction problem as a system of linear equations. The formulation is generic and covers 2D and 3D imaging in parallel/fan/cone beam geometries. Let $x = (x_j) \in \mathbb{R}^n$ denote a vector containing the discretized pixel/voxel values of the object being imaged, while $p = (p_i) \in \mathbb{R}^m$ represents the measured projection values from all detectors collapsed into a single vector. Then the forward image formation process can be modeled using a projection operator $W = (w_{ij}) \in \mathbb{R}^{m \times n}$ which maps the object to the measured projection data:

$$p = Wx \qquad (1)$$

In discrete tomography, we intend to solve the following problem: Let G>1 be the number of gray values and $R = \{\rho_1, \dots, \rho_G\}$ denote the set of gray values. Find $x \in R^n$ such that $Wx = p$. In practice a solution can be seen as the result of the following optimization:

$$\min_{x \in \{\rho_1, \dots, \rho_G\}^n} ||Wx - p|| \qquad (2)$$

where $||\cdot||$ represents a certain norm. Notice that because the solution space R^n is not a convex set, many algorithms from convex optimization cannot be directly applied for the reconstruction problem in discrete tomography.

B. Original DART algorithm

Here we provide a brief overview of the DART algorithm. More details can be found in [13]. DART alternates between a continuous algebraic reconstruction step and a discrete segmentation step, and uses a well-designed procedure to gradually improve the segmented image.

The algorithm starts with an initial continuous reconstruction using an algebraic reconstruction method (ARM). As the first step, a hard segmentation is applied on the reconstruction, rounding all pixels to the nearest gray value in R. In the second step, *boundary* pixels, that have at least one neighbor pixel with a different gray value, and some randomly chosen pixels from the non-boundary region are selected as *free* pixels. The

probability of a *non-boundary* pixel to be classified as *free* pixel is $1 - p$, with p known as the *fix probability*. In the third step, the unknowns corresponding to the pixels other than *free* pixels, known as *fixed* pixels, are removed from the system of linear equations in (1). Specifically the corresponding columns in W and rows in x are removed while the projections of the *fixed* pixels are subtracted from p. Then the *free* pixels are updated with a continuous reconstruction step with the ARM. In the fourth step, the free pixels are smoothed with a Gaussian filter with a small kernel as a way to regulate strong fluctuations. The whole procedure iterates between step 1 and 4 until a termination criterion is met.

III. FORMULATIONS FOR TVR-DART

In TVR-DART we combine the concept of solution steering of DART with TV regularization, and solve the discrete tomography problem within an automated optimization framework. The objective function F consists of two parts: a data fit term F_{fit} incorporating the discrete prior and a regularization term F_{reg} ensuring the sparsity of image gradients:

$$F(x, \bar{R}) = F_{\text{fit}}(x, \bar{R}) + \lambda \cdot F_{\text{reg}}(x, \bar{R}) \quad (3)$$

$$F_{\text{fit}} = \|WS(x, \bar{R}) - p\|_2^2 \quad (4)$$

$$F_{\text{reg}} = \sum_j M_\epsilon \left(\left(\nabla S(\tilde{x}, \bar{R}) \right)_j \right) \quad (5)$$

where λ is the weight for controlling the trade-off between the two parts of the objective function. $S(x, \bar{R})$ represents the Soft Segmentation Function that smoothly pushes the gray values toward discrete solutions, and $M_\epsilon(\cdot)$ is the Huber norm function [19]

$$M_\epsilon(r) = \begin{cases} r^2/2\epsilon & 0 \le |r| \le \epsilon \\ |r| - \epsilon/2 & |r| > \epsilon \end{cases} \quad (6)$$

which interpolates between smooth ℓ^2 treatment of small residuals and robust ℓ^1 treatment of large residuals with ϵ as the threshold between the two types of norms. $\bar{R}$ is the set of gray values augmented with the set of thresholds $\{\tau_1, ..., \tau_G\}$, $\tilde{x}$ is the spatial image representation of vector x, and ∇ represents the discrete gradient operator.

The inclusion of the soft segmentation in the objective function applies a soft push on the pixel values that encourages discrete solutions. The Huber norm of the image gradient is applied to steer towards sparse solutions and it yields a differentiable objective function. TVR-DART aims to minimize the objective function (3) over both the reconstruction x and the soft segmentation parameters $\bar{R}$:

$$\min_{x \in \mathbb{R}^n, \bar{R} = \{ \substack{\rho_1, ..., \rho_G \\ \tau_1, ..., \tau_G} \}} F(x, \bar{R}) \quad (7)$$

where G denotes the prior knowledge in terms of the total number of discrete gray values in the reconstruction.

The objective function in (3)-(5) means that TVR-DART pursues a reconstruction that matches well with the projection data in a least square sense after applying the segmentation
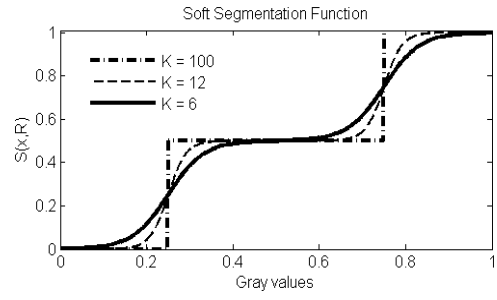

Fig. 1. Illustration of the soft segmentation function $S(x, R)$ under different values of K. The gray values are $(0, 0.5, 1.0)$ with thresholds at $(0.25, 0.75)$.

function, and the segmented solution is preferred to also exhibit sparse boundaries. Therefore, the solution we are looking for in the end is not x but $S(x, \bar{R})$ which is optimized to fit with the projection data. By changing the formulation of discrete tomography from (2) to (7), we manage to switch the problem of a convex objective function on a non-convex set, to a problem of a non-convex objective function over a convex set, which can be solved using non-convex optimization techniques. Previous attempts of limited data discrete tomography using non-convex optimization techniques can be found in [20-21]. In CT, non-convex prior models were also proposed for reconstruction from nonlinear X-ray measurements [22-24].

In order to have a smooth objective function for the optimization, we define the Soft Segmentation Function, $S(x, \bar{R})$, as a sum of logistic functions:

$$S(x, \bar{R}) = \sum_{g=2}^{G} \left(\rho_g - \rho_{g-1} \right) u(x - \tau_g, k_g) \quad (8)$$

where ρ_g are the discrete gray values with $\rho_1 = 0$, assuming the background is zero in the reconstruction, and τ_g represent the corresponding thresholds, with $\tau_1 = 0$. The function $u(x, k_g)$ is the logistic function defined as:

$$u(x, k_g) = \frac{1}{1 + e^{-2k_g x}} \quad (9)$$

where k_g is different for each step between consecutive gray values:

$$k_g = \frac{K}{\rho_g - \rho_{g-1}} \quad g = 2, ..., G \quad (10)$$

K is the transition constant for controlling the sharpness of transitions in the segmentation curve. The reason to define k_g as a function of the gray value difference is to make sure the maximum gradient of S is the same for every transition step. This also prevents the need to change the value of K when the gray values of the reconstruction fall in different value ranges. The soft segmentation function for different values of K is illustrated in Figure 1. As we can see, $S(x, \bar{R})$ approximates the staircase function of a hard segmentation when k_g increases. Because the function is monotonically increasing, it acts as a 'compression/expansion' operator on the solution space. This causes the local gradients to be compressed/expanded, which effectively encourages discrete solutions.
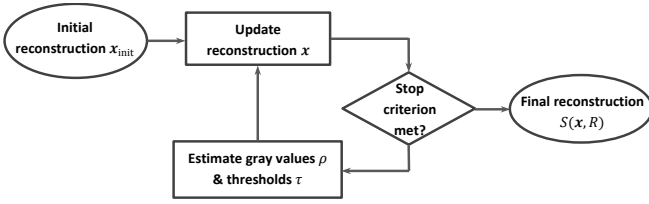
Fig. 2. Flowchart of the TVR-DART algorithm.

The flowchart of the TVR-DART algorithm is shown in Figure 2. The algorithm initializes x with a continuous reconstruction using TVmin, and it operates iteratively in a manner of alternating minimization. At first, the parameters for the segmentation function are estimated based on the current reconstruction. Then the reconstruction is updated based on the current gray values and thresholds. The whole process iterates until the stop criterion is met. By following this process, TVR-DART utilizes a simplified and automated procedure requiring only a small amount of parameter tuning. In the following three subsections, we will derive update equations that aim to minimize the defined objective function and provide more details on the initialization and optimization procedures.

A. Reconstruction step

Since the objective function $F(x, \bar{R})$ is differentiable and almost everywhere twice differentiable due to the choice of a smooth soft segmentation function and Huber norm, we use the second order approximation to find the next iteration of the reconstruction. The second order Taylor series of $F(x, \bar{R})$ around the current iteration x^t is

$$F(x) = F(x^t) + (x - x^t)^T J(x^t) + \frac{1}{2}(x - x^t)^T H(x^t)(x - x^t) \quad (11)$$

where $J(x^t)$ denotes the gradient and $H(x^t)$ is the Hessian matrix. Newton's approach would suggest we set the gradient of (11) to zero, which directly leads us to the stationary point of the quadratic function:

$$x = x^t - H(x^t)^{-1} J(x^t) \quad (12)$$

But this would require the estimation of the inverse of an extremely large Hessian matrix (n × n with n being the number of pixels in the reconstruction). Even quasi-Newton methods that approximate the inverse of the Hessian matrix would require manipulation of such large matrices, which is prohibitively expensive for the tomography application in practice. Therefore, we pursued a solution that replaces the Hessian matrix with a diagonal matrix $\bar{H}(x^t)$, whose inverse is straightforward to compute. We show that this forms an auxiliary function that lies above the original objective function around the current estimate. And taking a step to the minimum of the auxiliary function will lead us to the next estimate that will decrease the objective function. The details of this approach are provided in the following.

The gradient of both F_{fit} and F_{reg} over x are derived as:

$$J_{\text{fit}} = 2(W^T W S - W^T p) \circ \nabla_x S \quad (13)$$
$$J_{\text{reg}} = \nabla_S F_{\text{reg}} \circ \nabla_x S \quad (14)$$

with

$$\nabla_x S = 2K \sum_{g=2}^{G} u_g (1 - u_g) \quad (15)$$

where S is an abbreviation for $S(x^t, \overline{R^t})$, u_g is an abbreviation for $u(x - \tau_g, k_g)$, $\nabla_x S$ represents the partial derivative of S over the reconstruction x, and $\nabla_S F_{\text{reg}}$ is the partial derivative of F_{reg} with respect to the segmented reconstruction S. The formulation of $\nabla_S F_{\text{reg}}$ is derived in Appendix II. Here '$\circ$' represents element-wise multiplication.

The Hessian matrix of the data fit term F_{fit} is formulated as:

$$H_{\text{fit}} = 2\{\text{diag}(\nabla_x S) W^T W \text{diag}(\nabla_x S) + \text{diag}[(W^T W S - W^T p) \circ \nabla_{xx}^2 S]\} \quad (16)$$

with

$$\nabla_{xx}^2 S = 4K \sum_{g=2}^{G} k_g u_g (1 - u_g)(1 - 2u_g) \quad (17)$$

The second term of (16) is a diagonal matrix. And the first term is a full matrix of which the complete inverse would be expensive to compute. Here we replace the first term of H_{fit} with a diagonal matrix η_{fit}:

$$\bar{H}_{\text{fit}} = 2\{\eta_{\text{fit}} + \text{diag}[(W^T W S - W^T p) \circ \nabla_{xx}^2 S]\} \quad (18)$$

where

$$\eta_{\text{fit}} = \text{diag}[\text{diag}(\nabla_x S) W^T W \text{diag}(\nabla_x S) e] \quad (19)$$

Here $e = (e_j = 1) \in \mathbb{R}^n$ denotes a vector containing all ones. Note that η_{fit} can be effectively computed as:

$$\eta_{\text{fit}} = \text{diag}(\nabla_x S \circ W^T W \nabla_x S) \quad (20)$$

Because $\bar{H}_{\text{fit}} - H_{\text{fit}}$ is a positive semi-definite matrix (shown in Appendix I), we have

$$(x - x^t)^T [\bar{H}_{\text{fit}} - H_{\text{fit}}](x - x^t) \geq 0 \quad (21)$$

Therefore, replacing H_{fit} with $\bar{H}_{\text{fit}}$ forms an auxiliary function above the quadratic approximation of the data fit term of the objective function in a neighborhood of the current estimate x^t. Taking a step towards the minimum of the auxiliary function gives us the next estimate. We will show in later sections that this strategy almost always lowers the objective function.

The Hessian matrix of the regularization term F_{reg} is formulated as

$$H_{\text{reg}} = \text{diag}(\nabla_x S)[J(\nabla_S F_{\text{reg}})(S)]\text{diag}(\nabla_x S) + \text{diag}[\nabla_S F_{\text{reg}} \circ \nabla_{xx}^2 S] \quad (22)$$

The second term of H_{reg} is a diagonal matrix, and the term $J(\nabla_S F_{\text{reg}})(S)$ is the Hessian matrix of the Huber norm with respect to S. $J(\nabla_S F_{\text{reg}})(S)$ is a large but very sparse matrix and its formulation is provided in Appendix II. We replace the first term of H_{reg} in (22) with a diagonal matrix η_{reg}:

$$\bar{H}_{\text{reg}} = \eta_{\text{reg}} + \text{diag}[\nabla_S F_{\text{reg}} \circ \nabla_{xx}^2 S] \quad (23)$$

with

$$\eta_{\text{reg}} = \text{diag}\{\tilde{Q} e\} \quad (24)$$

and $Q = (q_{ij}) = \text{diag}(\nabla_x S)[J(\nabla_S F_{\text{reg}})(S)]\text{diag}(\nabla_x S)$, $\tilde{Q} = (|q_{ij}|)$. We show in Appendix I that $\bar{H}_{\text{reg}} - H_{\text{reg}}$ is positive semi-definite, so:

$$(x - x^t)^T [\bar{H}_{\text{reg}} - H_{\text{reg}}](x - x^t) \geq 0 \quad (25)$$

Thus replacing H_{reg} with $\bar{H}_{\text{reg}}$ forms an auxiliary function

above the quadratic approximation of the regularization term of the objective function in a neighborhood of the current estimate x^t. Note that this is just one possible choice of auxiliary function, and other choices may work just as well.

Substituting (18)(23) in (12) results in the following update:

$$x^{t+1} = x^t - \bar{H}(x^t)^{-1} J(x^t) \tag{26}$$

where

$$J(x^t) = J_{\text{fit}} + \lambda \cdot J_{\text{reg}} \tag{27}$$

$$\bar{H}(x^t) = \bar{H}_{\text{fit}} + \lambda \cdot \bar{H}_{\text{reg}} \tag{28}$$

The advantage of the proposed algorithm using the auxiliary functions is that it is computationally efficient and easy to implement. There is no need to adjust the step size as in gradient descent methods, and the algorithm achieves faster convergence using the curvature information. This is beneficial for large-scale tomography applications in practice.

B. Gray value and threshold estimation steps

In this step, we estimate the gray values and thresholds based on the current estimate of the reconstruction using Newton's method. From the objective function in (3), the gradients of F_{fit} and F_{reg} over either the gray value or threshold, denoted by $\bar{R}_g$, are derived as follows:

$$J_{\text{fit}}^{\bar{R}_g} = 2(WS - p)^T W \nabla_{\bar{R}_g} S \tag{29}$$

$$J_{\text{reg}}^{\bar{R}_g} = [\nabla_S F_{\text{reg}}]^T \nabla_{\bar{R}_g} S \tag{30}$$

And the Hessian matrices are written as:

$$H_{\text{fit}}^{\bar{R}_g} = 2 \left\{ \left(W \nabla_{\bar{R}_g} S\right)^T \left(W \nabla_{\bar{R}_g} S\right) + (WS - p)^T \left(W \nabla_{\bar{R}_g}^2 S\right) \right\} \tag{31}$$

$$H_{\text{reg}}^{\bar{R}_g} = \sum \left\{ \text{diag}(\nabla_{\bar{R}_g} S)[J(\nabla_S F_{\text{reg}})(S)]\text{diag}(\nabla_{\bar{R}_g} S) + \text{diag}\left[\nabla_S F_{\text{reg}} \circ \nabla_{\bar{R}_g \bar{R}_g}^2 S\right] \right\} \tag{32}$$

where $\nabla_{\bar{R}_g} S$ and $\nabla_{\bar{R}_g}^2 S$ denote the first and second order partial derivatives of S over either the gray value ρ_g or the threshold τ_g, respectively. Their analytical expressions are as follows:

$$\nabla_{\rho_g} S =$$
$$\begin{cases} u_g \cdot [1 - 2k_g(x - \tau_g)(1 - u_g)] \\ \quad - u_{g+1} \cdot [1 - 2k_{g+1}(x - \tau_{g+1})(1 - u_{g+1})], & g \neq G \\ u_g \cdot [1 - 2k_g(x - \tau_g)(1 - u_g)], & g = G \end{cases} \tag{33}$$

$$\nabla_{\rho_g}^2 S = \begin{cases} \frac{4K^2(x - \tau_g)^2 u_g(1 - u_g)(1 - 2u_g)}{(\rho_g - \rho_{g-1})^3} \\ \quad + \frac{4K^2(x - \tau_{g+1})^2 u_{g+1}(1 - u_{g+1})(1 - 2u_{g+1})}{(\rho_{g+1} - \rho_g)^3}, & g \neq G \\ \frac{4K^2(x - \tau_g)^2 u_g(1 - u_g)(1 - 2u_g)}{(\rho_g - \rho_{g-1})^3}, & g = G \end{cases} \tag{34}$$

$$\nabla_{\tau_g} S = -2K u_g (1 - u_g) \tag{35}$$

$$\nabla_{\tau_g}^2 S = 4K k_g u_g (1 - u_g)(1 - 2u_g) \tag{36}$$

with $u_g = u(x - \tau_g, k_g)$ and $u_{g+1} = u(x - \tau_{g+1}, k_{g+1})$.

These lead to the update for gray values and thresholds:

$$\rho_g^{t+1} = \rho_g^t - J^{\rho_g}/H^{\rho_g} \tag{37}$$

$$\tau_g^{t+1} = \tau_g^t - J^{\tau_g}/H^{\tau_g} \tag{38}$$

where

$$J^{\rho_g} = J_{\text{fit}}^{\rho_g} + \lambda \cdot J_{\text{reg}}^{\rho_g} \tag{39}$$

$$H^{\rho_g} = H_{\text{fit}}^{\rho_g} + \lambda \cdot H_{\text{reg}}^{\rho_g} \tag{40}$$

$$J^{\tau_g} = J_{\text{fit}}^{\tau_g} + \lambda \cdot J_{\text{reg}}^{\tau_g} \tag{41}$$

$$H^{\tau_g} = H_{\text{fit}}^{\tau_g} + \lambda \cdot H_{\text{reg}}^{\tau_g} \tag{42}$$

In each update step of gray values and thresholds, (37) and (38) are applied to obtain a better estimation. Please note that the proposed algorithm might still diverge under certain data conditions and initialization. But as we show in the numerical and experimental sections of the paper, the proposed technique leads to faithful reconstructions and estimation of gray values for a wide range of conditions and datasets.

C. Optimization procedure, initialization and convergence criteria

With the derived formulas in the two previous subsections, we have the necessary update equations to take steps towards a solution. However, there are still a few issues regarding initialization, the optimization procedure and convergence that deserve further discussions.

The proposed algorithm requires to be initialized with a reconstruction x^0 with gray values between the minimum and maximum gray values of the soft segmentation function. As we mentioned, the inclusion of the soft segmentation function in the objective function causes the local gradients to be compressed/expanded. As a consequence, if the initial estimation only contains gray values within the flat regions of the soft segmentation function (e.g., all pixels are below 0 or larger than the maximum gray value), the solution will only move in very small steps due to the small gradients in these regions. We suggest initializing the reconstruction algorithm with a solution from TVmin as suggested in [25], and setting the initial maximum gray value of the soft segmentation function to the maximum pixel value of the initial reconstruction. The rest of the gray values are then uniformly distributed between the minimum and maximum values.

The procedure of the complete TVR-DART algorithm is illustrated in Figure 3 with the necessary equations highlighted. After initialization, the algorithm updates both reconstruction and gray values/thresholds in a manner of alternating minimization. Based on the initial reconstruction, both gray values and thresholds are updated. Then the reconstruction is improved using the updated soft segmentation parameters. The optimization is ended when the convergence criteria is reached:

$$\frac{|S^t - S^{t-1}|_1}{|S^{t-1}|_1} \leq \gamma \tag{43}$$

where $|\cdot|_1$ denotes the ℓ^1 norm, and γ denotes the convergence threshold. In practice, we have opted to set $\gamma = 10^{-5}$.

Compute an initial reconstruction x^0 using TVmin;

$t := 0;$

while **(stop criterion in (43) is not met)** do

begin

  $t := t + 1;$

  Compute the soft segmented image S and its partial derivatives;

  /**Gray value and threshold estimation step**/

  Prepare the Gradient and Hessian matrices using (29)-(32)

  Update $\bar{R}^{t+1} = \min\limits_{\bar{R}=\{\rho_1,...,\rho_G, \tau_1,...,\tau_G\}} F$ using (37) and (38);

  Update $S, \nabla_{\bar{R}_g} S$, and $\nabla^2_{\bar{R}_g} S$;

  /**Reconstruction step**/

  Prepare the auxiliary function using (13)(14) and (18)(23)

  Update $x^{t+1} = \min\limits_{x\in\mathbb{R}^n} F$ using (26);

  Update $S, \nabla_x S$, and $\nabla^2_{xx} S$;

end

Fig. 3. Procedure of the TVR-DART algorithm.

IV. NUMERICAL INVESTIGATION

In this Section, numerically simulated data are used to investigate the behavior of TVR-DART and evaluate its performance. The complete algorithm is implemented in Matlab using the open source ASTRA Tomography Toolbox [26], in which the basic forward and backward projection operations are efficiently computed in parallel using Graphical Processing Unit (GPU).

The simulation experiments use five 2D Phantoms, shown in Figure 4. All Phantoms are 512×512 pixels in size. Phantom 1 to 3 are binary images with increasing shape complexity. Phantom 4 and 5 are images containing 3 and 6 discrete gray values. The projection data (sinogram) is generated assuming a parallel beam geometry. In our experiments, the length of the detector array is equal to the width of the image, and the detector spacing is the same as the pixel size of the Phantom. Thus the projection for each angle consists of 512 detector values. The Radon transform of the images is first computed resulting in a sinogram with linear integrals of attenuation coefficients. Then the noise-less projection data is generated by taking the exponential of the negative values of the sinogram. The projection data is further perturbed with Poisson noise assuming a maximum number of photon counts I_0 at the detectors. In our experiments, I_0 varied from 5×10^3 (high noise) up to 10^6 (little noise) to study the algorithm's performance at different noise levels. The final sinogram used for reconstruction is obtained by taking the negative logarithm of the rescaled noisy projection data.

Three algorithms are used for comparison: SIRT, TVmin, and DART. For SIRT, 200 iterations are applied to ensure convergence is reached. Chambolle's algorithm described in [27] is used for TVmin. In total, 200 iterations are applied while the weight for the continuous TVmin is chosen between 2-5 for all the experiments. For the DART implementation, the SIRT algorithm is utilized as the ARM step. The algorithm iterates for 100 iterations to ensure convergence. Within each main iteration, 20 sub-iterations of SIRT are performed on the free pixels. For noisy data, the fix probability is specified at $p = 0.5$ in order to minimize the effect of noise on the reconstruction of boundaries as described in [13].

For quantification of the reconstruction performances, we opted for the Relative Mean Error (RME) that is given by:

$$\text{RME} = \frac{\sum_j |x_j^o - x_j^r|}{\sum_j |x_j^o|} \tag{44}$$

where x_j^o and x_j^r denote the *j*-th pixel of the original and the reconstructed image, respectively. Thus the RME value gives the error in proportion to the object pixels in the original image.

A. Noisy and Small Number of Projections

We first test the reconstruction performance of TVR-DART using a small number of projections that are also contaminated with a relatively high level of Poisson noise. It is assumed in this experiment that the projection angles are equally spaced between 0 and 180 degrees. Figure 5 shows TVR-DART reconstructions of all five Phantoms using various numbers of projections. The number of projections for each Phantom is chosen to obtain a satisfactory reconstruction, and depends on the complexity of the morphologies and the number of gray values of the Phantom. Corresponding SIRT, TVmin, and DART reconstructions are shown for comparison. For DART and TVR-DART in this experiment, we assume the gray values are known a priori in order to evaluate the reconstruction part of the algorithms separately. As we can clearly see, TVR-DART is able to consistently generate more accurate reconstructions than SIRT, TVmin and DART. It is able to combine the merits of both discrete tomography and total variation minimization. Similar to DART, TVR-DART is capable of recovering structures that are otherwise difficult to reconstruct by conventional algorithms. But unlike DART, it handles noise well benefiting from its TV regularization.

Figure 6 demonstrates the reconstruction of TVR-DART with simultaneous estimation of gray values and thresholds under the same conditions as in Figure 5. For all Phantoms, the initial gray values are specified by setting the maximum gray values to the maximum pixel value of the initial continuous reconstructions from TVmin, and the intermediate gray values uniformly distributed between 0 and the maximum. The starting values of the thresholds are chosen as the average values between adjacent gray values. We observe that the algorithm performs equally well for Phantom 1 to 4 while providing a degraded reconstruction for Phantom 5. This is due to the need to estimate a much larger number of gray values in
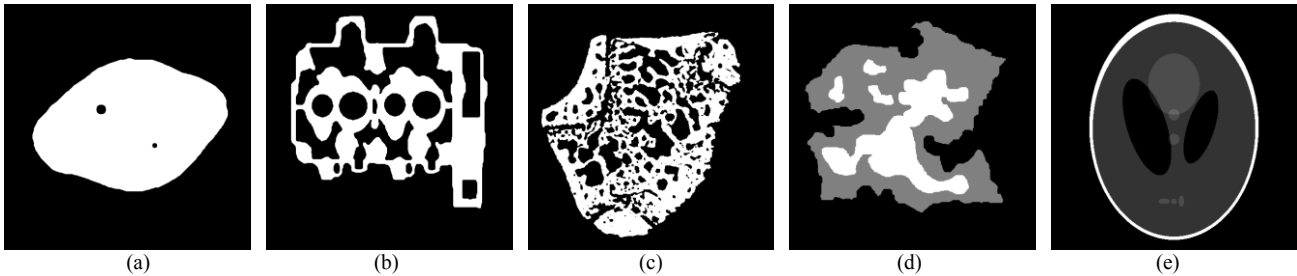
Fig. 4. Phantoms used in the numerical simulation. (a) Phantom 1. (b) Phantom 2. (c) Phantom 3. (d) Phantom 4. (e) Phantom 5.
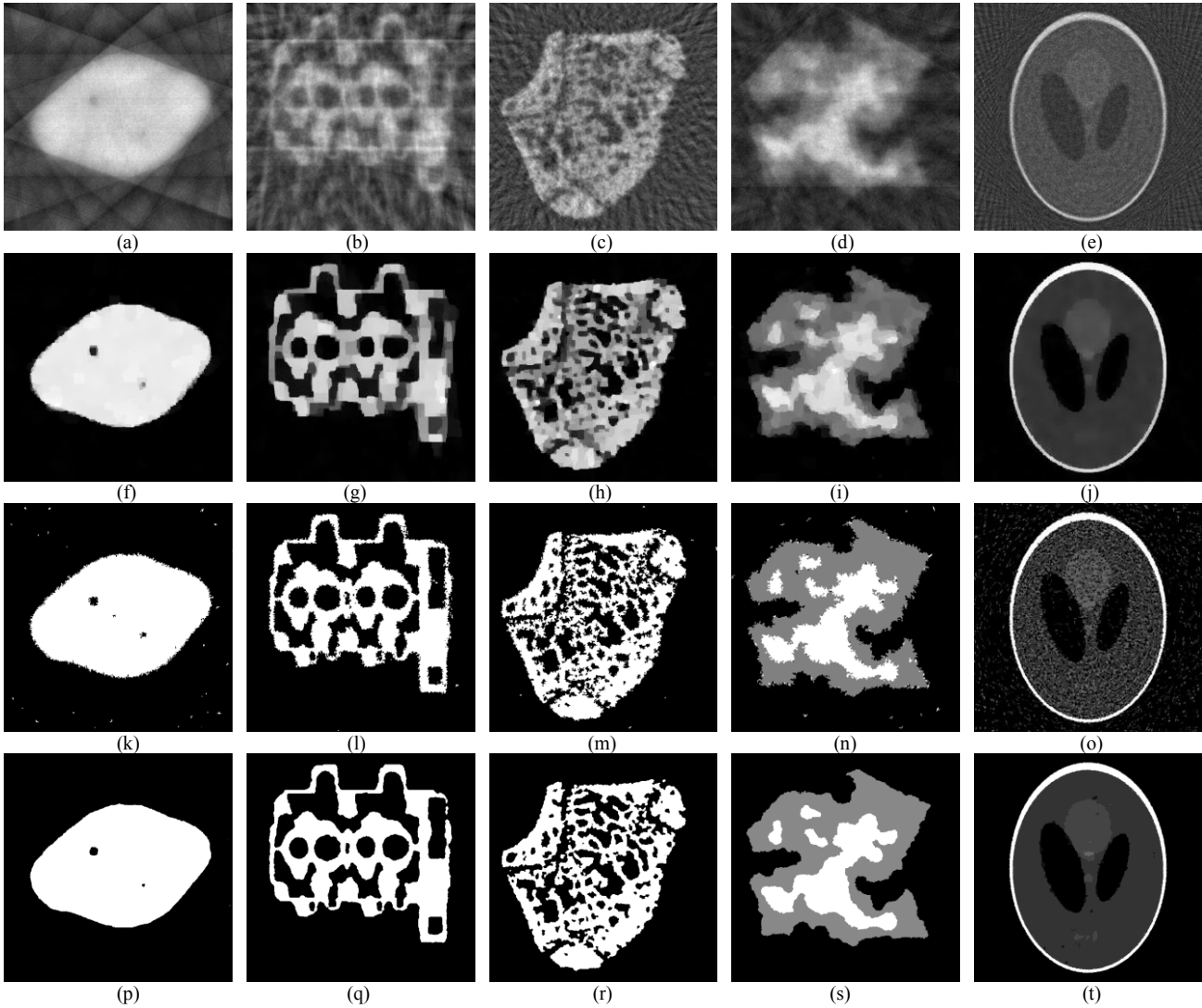

Fig. 5. Comparison of reconstructions using noisy and limited number of projections, from SIRT (row 1), TVmin (row 2), DART (row 3), and TVR-DART (row 4) for Phantom 1 using 9 projections (column 1), Phantom 2 using 15 projections (column 2), Phantom 3 using 30 projections (column 3), Phantom 4 using 12 projections (column 4), and Phantom 5 using 60 projections (column 5), respectively. (a)–(e) SIRT. (f)–(j) TV minimization. (k)–(o) DART. (p)–(t) Proposed TVR-DART with gray values fixed.

Phantom 5 that led to less accurate estimation of gray values and in turn a less accurate reconstruction. Therefore, although the proposed algorithm can help converge towards more accurate estimation of gray values, a less ideal initial guess of these values under limited data conditions could still lead to worse image quality compared to the case where all gray values are known in advance. Figure 6 (f) and (g) show the convergence of the algorithm by both the objective function and the sum of the absolute errors of the estimated gray values. Despite limited and noisy data, the TVR-DART algorithm shows a steadily decreasing objective function for all Phantoms. Figure 6 (h) further lists the RME for all
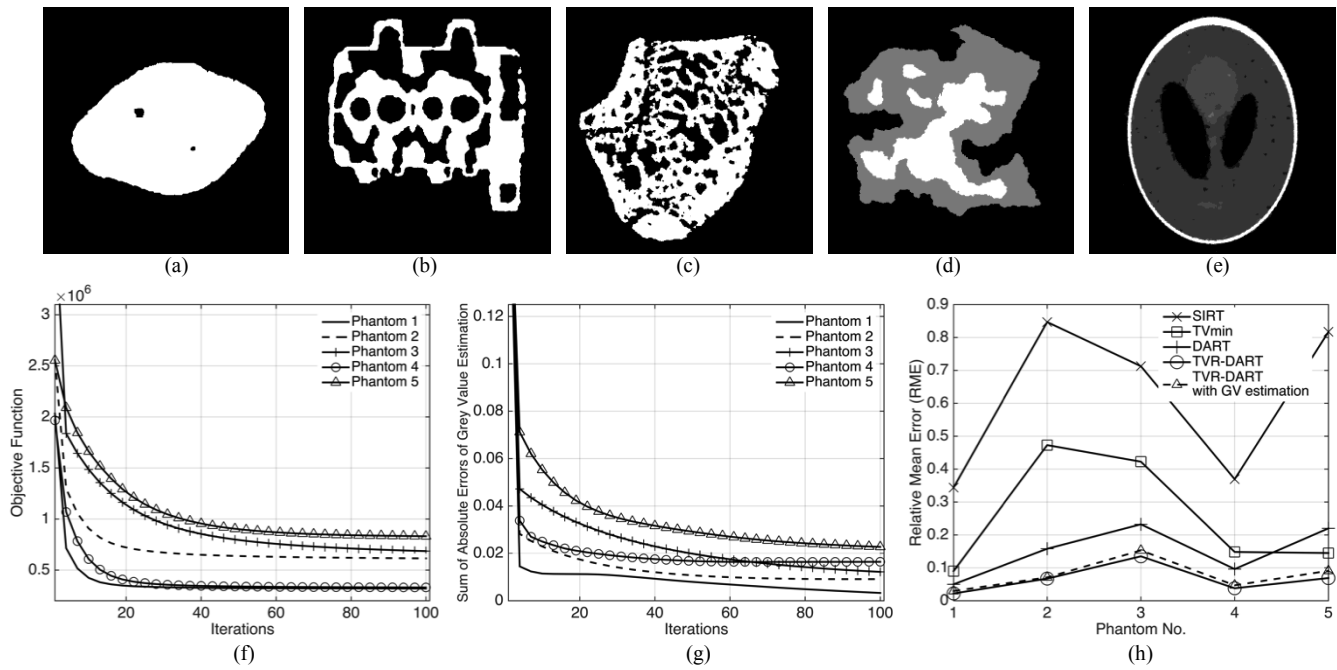
Fig. 6. Reconstruction with simultaneous estimation of gray values using TVR-DART under the same noise level and number of projection images as in Fig 5. (a) - (e) Reconstruction of phantom 1 to 5. (f) Convergence of the objective function through iterations. (g) Convergence of gray value estimation as sum of absolute errors of the estimated gray values through iterations. (h) Comparison of the Relative Mean Error (RME) of reconstructions among SIRT, TVmin, DART, TVR-DART with and without gray value estimation for all 5 phantoms.
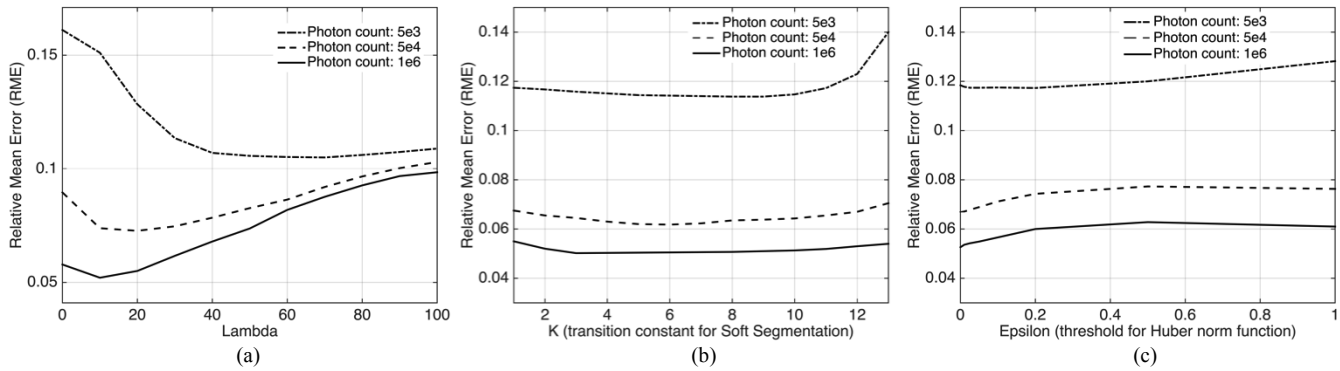

Fig. 7. Performance of TVR-DART under different choice of parameters and noise levels. (a) Relative Mean Error (RME) of reconstruction under different λ weight for the regularization, (b) RME under different transition constant K for the Soft Segmentation Function, (c) RME under different choice of ϵ the threshold for the Huber norm function. Phantom 3 is used for plotting these curves.

reconstructions shown in Figure 5 and 6. We see that the RME from TVR-DART is consistently lower than all other techniques. When the gray values are unknown and need to be estimated, slight degradation of results can be seen for Phantom 3 and Phantom 5, due to their higher morphological complexity and larger number of gray values. In these situations, a useful strategy is to first estimate both gray values and the reconstruction, and then reinitiate the reconstruction while keeping the estimated gray value fixed. This is a common strategy in alternating minimization.

B. Optimum Choice of Parameters

Figure 7 demonstrates the RME of the reconstruction of Phantom 3 as a function of the three main parameters of TVR-DART: the total variation weighting λ, the transition constant K, and the threshold ϵ for the Huber norm function. Here a full tilt angle range of 180 degrees with 6 degree steps (30 projections) is assumed. As expected, the optimum choice of λ increases with higher noise level in the projection data. For data with a relatively small amount of noise, a value range of $\lambda \in [10, 20]$ gives the optimal results while a higher value of $\lambda \in [50, 100]$ is needed for extremely noisy conditions.

The choice of the transition constant K is less influenced by the data. While a lower value of K makes the segmentation function too smooth by offering not enough push to the discrete gray values, setting the value too high pushes the solution too fast toward discrete values. The threshold for the Huber norm also exhibits a relative weak influence on the reconstruction accuracy. Choosing a small value of ϵ ensures that large
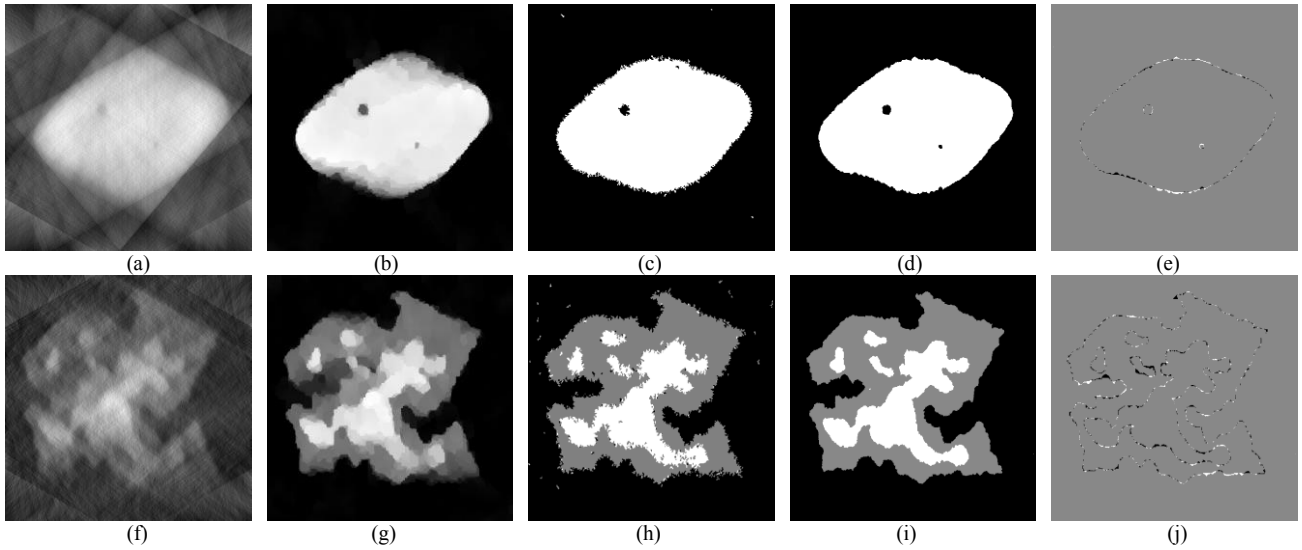
Fig. 8. Comparison of reconstructions under limited angular range (-60 to 60 degree) using noisy (5×10^3 photon count) and limited number of projections, from SIRT (a, f), TVmin (b, g), DART (c, h), and TVR-DART with simultaneous estimation of gray values (d, i) for phantom 1 using 12 projections (row 1), phantom 4 using 21 projections (row 2). (e, j) shows the difference between the ground truth and the TVR-DART reconstructions.

regions of the reconstruction are treated with the edge-preserving ℓ^1 norm, which leads to a sharper and more accurate solution. For the following studies, we keep $K = 6$ and $\epsilon = 0.02$ while λ is selected following the summarized guidelines.

C. Limited Angular Range

In many practical applications such as electron tomography or certain medical applications, it is impossible to acquire projection images of the object from all tilt angles, and the data can be only acquired over a limited angular range. In such a situation, conventional algorithms such as FBP or SIRT cannot solve the reconstruction problem accurately even using a sufficiently large number of projections. Here we test TVR-DART using simulated data from a typical angular range in electron tomography, from -60 to 60 degrees with a limited number of projections. The projection data is further perturbed by a high level of noise (5×10^3 photon count) to mimic practical conditions.

Figure 8 shows the reconstructions for Phantom 1 and 4 using SIRT, TVmin, DART, and TVR-DART. As we can see, the advantage of TVR-DART is even more significant under limited angular range conditions. The proposed algorithm yields much more accurate reconstructions than the other techniques used in this comparison. This is further verified by the difference between the reconstruction and original Phantom. By imposing prior knowledge of both discrete gray values and sparse boundaries, TVR-DART is shown to be capable of dealing with limited angular range problem even under sparse and noisy data conditions.

V. EXPERIMENTAL VERIFICATION

To further evaluate the proposed algorithm in practical conditions, we applied TVR-DART on experimental datasets from two different imaging modalities: X-ray Micro-CT and Electron Tomography.

A. X-ray Micro-CT Experiment

For this Micro-CT experiment, a hardware Phantom is used. The Phantom has a convex shape like a bean and the body is made of Plexiglass with three holes drilled through in the vertical direction. The bigger hole on the left is filled with white spirit while the two smaller holes are filled with air. The projection data of the Phantom is acquired by a Skyscan 1172 microtomography X-ray scanner. A total of 600 projection images are used in this experiment over a tilt range of 174 degree. The tilt angle step between projections is 0.6˚. The dataset is first pre-processed to reduce some of the physical artifacts such as beam hardening, and make the data applicable for discrete tomography. We applied the algorithms on two subsets of the full scan within the complete angular range: one with 30 projections of 6˚ steps, and one with 10 projections of 18˚ steps. This tests the algorithms under highly limited data conditions. A Segmented SIRT reconstruction from all 600 projections is used to generate an image to be used as ground truth to evaluate the reconstructions under limited data conditions. The scanner has a cone beam geometry. Because the Phantom is the same along the vertical direction, the central slice of the data is reconstructed in this experiment using a fan beam geometry. The pixel size of the projection data is 25 μm with 1000 pixels over the width of the detector.

For TVR-DART, assuming that there are only three different materials, we first estimated the corresponding three gray values over the first 50 iterations, then the reconstruction is reinitiated to obtain the final reconstruction over 200 iterations. The weight of the total variation is specified as $\lambda = 60$ for the 30 projection scenario and $\lambda = 40$ for the 10 projection case, while the transition constant of the soft segmentation function is kept at $K = 6$ for both experiments.
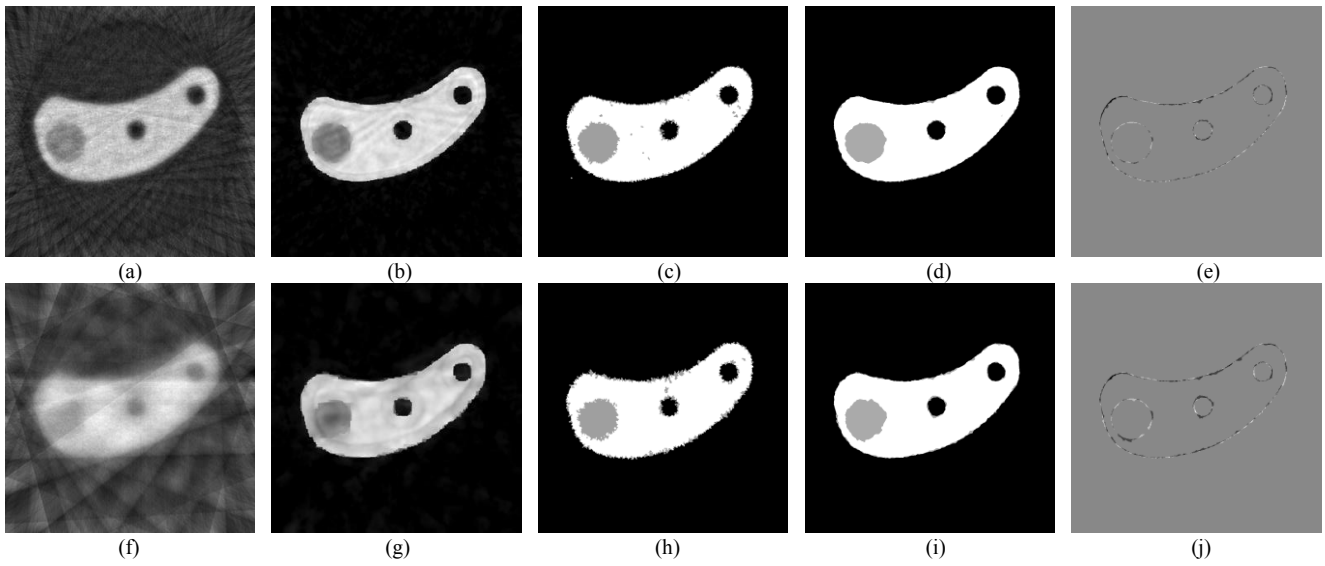
Fig. 9. Experimental verification using X-ray Micro-CT (SkyScan 1172) data of a physical bean-shaped phantom. Reconstruction from 30 projections (row 1) and 10 projections (row 2) in fan beam geometry obtained using (a)-(f) SIRT, (b)-(g) TVmin, (c)-(h) DART, (d)-(i) TVR-DART, and (e)-(j) the difference between TVR-DART reconstructions and the ground truth of the phantom. The body of the physical phantom is made of Plexiglass with the larger hole on the left filled with white spirit and the two small holes on the middle and right side filled with air.

Figure 9 shows the reconstructions of the Phantom using SIRT, TVmin, DART, and TVR-DART from 30 and 10 projections. The pixel error maps between the TVR-DART reconstructions and the ground truth of the Phantom are also presented. SIRT reconstructions are much noisier than other techniques and degrade the most when the number of projections is further reduced. TVmin provides a good alternative to SIRT in terms of reducing noise and artifacts in the reconstruction. However, the boundaries of TVmin are less defined, especially around the region of the large hole with white spirit, which has less contrast to the surrounding material. DART provides more defined boundaries than TVmin, but fails to handle the noise in the original data. This problem mostly appears around the boundaries of the object, which makes the reconstruction less useful. Among all, TVR-DART is able to provide the best reconstructions even using only 10 projections. The degradation of quality is also the smallest among algorithms comparing the results from 30 and 10 projections. This shows the clear advantage of combining the prior knowledge of both discreteness and sparsity of the object in the reconstruction technique.

B. Electron Tomography Experiment

An electron tomography experiment of nanomaterials is further carried out to test the proposed algorithm. The sample consists of PbSe-CdSe core shell nanoparticles with an average diameter of 9nm. The nanoparticles are dispersed on a carbon support. A tilt series of the particle assembly is acquired with an angular range from $-55°$ to $55°$ with $1°$ increments in high angle annular dark field scanning transmission electron microscopy (HAADF-STEM) mode. The acquisition was performed at a FEI TITAN microscope using a Fischione tomography tilt holder operated by the FEI Xplore3D software. The pixel size used in the acquisition equals 1.36 nm. The acquired images are

first aligned correcting slight image rotation, magnification and shifts over the tilt series.

Reconstructions are performed on the full data and on a subset of data with half of the projections over the same angular range. The results from SIRT, TVmin, DART, and TVR-DART are presented in Figure 10. For TVR-DART, the number of materials and its sparsity of object boundaries are used as prior information for the reconstruction. We can observe that all particles are visible in the SIRT and TVmin reconstructions from the complete data but less so when only half of projections are used. DART does a good job identifying all small particles, but also generates noisy boundaries and some inconsistent holes within the inner regions of the nanoparticles. Among the four tested algorithms, TVR-DART yields the most accurate reconstructions with clear recovery of the small particles and boundaries of the shape of the large ones while maintaining a high image quality even using only 55 images under a severe missing angle condition.

VI. CONCLUSIONS

In this paper, we have presented a new discrete tomography reconstruction algorithm, TVR-DART, that exploits two types of prior information of the object simultaneously in terms of the sample's discreteness in materials and sparsity of its boundaries. The algorithm is aimed at tomographic reconstruction of objects consisting of a few different material compositions, each approximately corresponding to a constant gray value in the reconstruction. By defining a soft segmentation function within the objective function of the reconstruction algorithm, TVR-DART smoothly steers the solution toward discrete gray values while minimizing the total variation of the boundaries of the discrete solution. Since it is very difficult to know the exact gray values in most practical
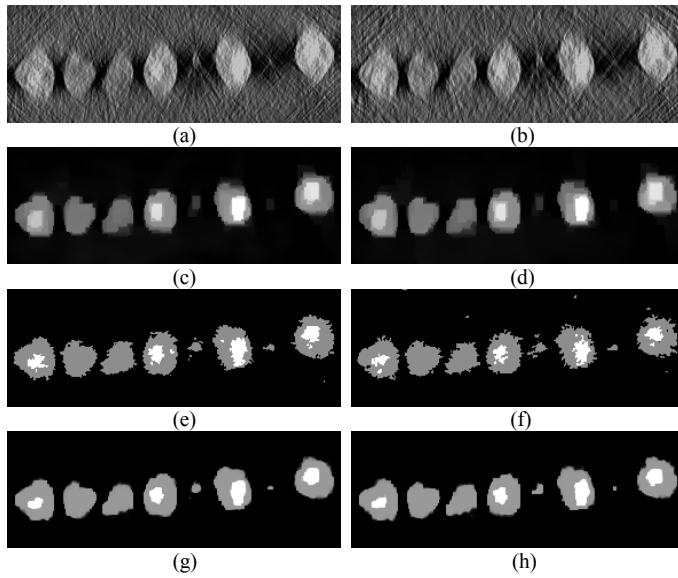
Fig. 10. Experimental verification using Electron Tomography of PbSe-CdSe core/shell nanoparticles with an average diameter of 9nm. The tilt angle ranges from -55 to 55 degrees. Comparison of reconstructions from 111 projections (column 1) and 55 projections (column 2) obtained using (a)(b) SIRT, (c)(d) TVmin, (e)(f) DART, (g)(h) TVR-DART with simultaneous estimation of gray values.

applications, the gray values and thresholds of the segmentation function are automatically estimated in an alternating manner with the reconstruction assuming the total number of gray values is known. Extensive numerical simulations were carried out under a small number of projections corrupted with a high level of Poisson noise and further within a limited angular range. Experimental data from two imaging modalities of micro-CT and Electron Tomography were further used to test the algorithm in practical applications. The accuracy of TVR-DART is compared with SIRT, TV minimization, and DART. All the experimental results demonstrate that the proposed TVR-DART algorithm is able to take the advantage of the two types of priors, and is capable of consistently producing accurate reconstructions under noisy and limited data conditions. With the key computation steps of the algorithm readily implemented using GPU for fast and efficient computations, we aim to provide the experimental fields with a robust and automated technique for discrete tomography.

APPENDIX I

This Section gives the proof of the positive semi-definite results referred to in Section III.A. Specifically, proving (21) and (25) requires the following lemma:

Lemma 1: Let $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ be a square symmetric n × n matrix, $\tilde{A} = (|a_{ij}|) \in \mathbb{R}^{n\times n}$ the matrix with absolute values of A as its elements, and e a n × 1 vector with all entries equal to one, then $B = \mathrm{diag}(\tilde{A}e) - A$ is a positive semi-definite matrix.

Proof: Assume that $t = [t_1, t_2, \ldots, t_n]^T \neq 0_{n\times 1}$, Then

$$t^T(B)t = t^T(\mathrm{diag}(\tilde{A}e) - A)t \tag{45}$$

$$= \sum_{ij}(t_i^2|a_{ij}| - t_i a_{ij} t_j)$$

$$= \sum_{ij}\left(\frac{1}{2}(t_i^2 + t_j^2)|a_{ij}| - t_i a_{ij} t_j\right) \quad \text{(Due to symmetry)}$$

Each individual term is now non-negative, since:
If $a_{ij} < 0$, we have

$$\left(-\frac{1}{2}t_i^2 - t_i t_j - \frac{1}{2}t_j^2\right)a_{ij} = -a_{ij}\frac{1}{2}(t_i + t_j)^2 \geq 0 \tag{46}$$

If $a_{ij} \geq 0$, we have

$$\left(\frac{1}{2}t_i^2 - t_i t_j + \frac{1}{2}t_j^2\right)a_{ij} = a_{ij}\frac{1}{2}(t_i - t_j)^2 \geq 0 \tag{47}$$

Therefore, $t^T(B)t \geq 0$, thus $B = \mathrm{diag}(\tilde{A}e) - A$ is positive semi-definite.

It is obvious that the first term of the Hessian matrix H_{fit} in (16) is a symmetric matrix:

$$\mathrm{diag}(\nabla_x S)W^T W \mathrm{diag}(\nabla_x S)$$
$$= [W\mathrm{diag}(\nabla_x S)]^T W \mathrm{diag}(\nabla_x S) \tag{48}$$

Because the soft segmentation function is non-decreasing, all the entries of $\nabla_x S$ are non-negative. Also all the elements within the projection matrix W are non-negative. Thus applying Lemma 1 shows that

$$\bar{H}_{\mathrm{fit}} - H_{\mathrm{fit}}$$
$$= \mathrm{diag}[\mathrm{diag}(\nabla_x S)W^T W \mathrm{diag}(\nabla_x S)e]$$
$$- \mathrm{diag}(\nabla_x S)W^T W \mathrm{diag}(\nabla_x S) \tag{49}$$

is positive semi-definite, therefore (21) is true.

For (25), the term $J(\nabla_S F_{\mathrm{reg}})(S)$ within the Hessian matrix H_{reg} in (22) is a symmetric matrix because it is the Hessian matrix of the regularization term of the objective function with respect to S. Thus the first term of H_{reg} is symmetric:

$$[\mathrm{diag}(\nabla_x S)[J(\nabla_S F_{\mathrm{reg}})(S)]\mathrm{diag}(\nabla_x S)]^T$$
$$= \mathrm{diag}(\nabla_x S)[J(\nabla_S F_{\mathrm{reg}})(S)]\mathrm{diag}(\nabla_x S) \tag{50}$$

Using Lemma 1, we can prove that

$$\bar{H}_{\mathrm{reg}} - H_{\mathrm{reg}} = \mathrm{diag}\{\tilde{Q}e\} - Q \tag{51}$$

is positive semi-definite, thus (25) is valid. With both (21) and (25), we arrive at

$$(x - x^t)^T[\bar{H} - H](x - x^t) \geq 0 \tag{52}$$

Therefore, replacing $H(x^t)$ with $\bar{H}(x^t)$ in (11) forms an auxiliary function $G(x, x^t)$:

$$G(x, x^t) = F(x^t) + (x - x^t)^T J(x^t)$$
$$+ \frac{1}{2}(x - x^t)^T \bar{H}(x^t)(x - x^t) \tag{53}$$

which satisfies the following conditions:

$$G(x, x^t) \geq F(x), \quad G(x^t, x^t) = F(x^t) \tag{54}$$

This in turn shows that the objective function F is non-increasing if a step minimizing the auxiliary function is taken in each iteration:

$$x^{t+1} = \arg\min_{x \in \mathbb{R}^n} G(x, x^t) \tag{55}$$

due to the fact that

$$F(x^{t+1}) \leq G(x^{t+1}, x^t) \leq G(x^t, x^t) = F(x^t) \tag{56}$$

Therefore, the proposed algorithm is capable of decreasing the objective function.

APPENDIX II

In this Section we derive analytical formulas for the gradient and Hessian matrices of the Huber norm cost function F_{reg} with respect to the segmented reconstruction S : $\nabla_S F_{\text{reg}}$ and $J(\nabla_S F_{\text{reg}})(S)$, which are needed for estimation of (14), (23), (30) and (32). Let $\tilde{S} = (\tilde{S}_{u,v}) \in \mathbb{R}^{U \times V}$ denote the two-dimensional soft segmented reconstruction matrix, with S being its vectorized version. We define the discrete gradient operator ∇ as

$$(\nabla \tilde{S})_{u,v} = \left((\nabla \tilde{S})_{u,v}^1, (\nabla \tilde{S})_{u,v}^2 \right) \tag{57}$$

with

$$(\nabla \tilde{S})_{u,v}^1 = \begin{cases} \tilde{S}_{u+1,v} - \tilde{S}_{u,v} & u < U \\ 0 & u = U \end{cases} \tag{58}$$

$$(\nabla \tilde{S})_{u,v}^2 = \begin{cases} \tilde{S}_{u,v+1} - \tilde{S}_{u,v} & v < V \\ 0 & v = V \end{cases} \tag{59}$$

And the $|\cdot|$ within the Huber norm function in (6) is defined as

$$\left| (\nabla \tilde{S})_{u,v} \right| = \sqrt{ \left((\nabla \tilde{S})_{u,v}^1 \right)^2 + \left((\nabla \tilde{S})_{u,v}^2 \right)^2 } \tag{60}$$

The partial derivative of F_{reg} over $\tilde{S}_{u,v}$ depends on three adjacent pixels $\tilde{S}_{u,v}$, $\tilde{S}_{u,v-1}$, and $\tilde{S}_{u-1,v}$:

$$\frac{\partial F_{\text{reg}}}{\partial \tilde{S}_{u,v}} =$$

$$\frac{\partial}{\partial \tilde{S}_{u,v}} \left\{ \frac{1}{2\epsilon} \left((\nabla \tilde{S})_{u,v} \right)^2 \cdot M_{u,v} + \left(\left| (\nabla \tilde{S})_{u,v} \right| - \frac{\epsilon}{2} \right) \cdot \bar{M}_{u,v} + \right.$$
$$\frac{1}{2\epsilon} \left((\nabla \tilde{S})_{u,v-1} \right)^2 \cdot M_{u,v-1} + \left(\left| (\nabla \tilde{S})_{u,v-1} \right| - \frac{\epsilon}{2} \right) \cdot \bar{M}_{u,v-1} +$$
$$\left. \frac{1}{2\epsilon} \left((\nabla \tilde{S})_{u-1,v} \right)^2 \cdot M_{u-1,v} + \left(\left| (\nabla \tilde{S})_{u-1,v} \right| - \frac{\epsilon}{2} \right) \cdot \bar{M}_{u-1,v} \right\} \tag{61}$$

where $\mathbf{M} = (M_{u,v}) \in \mathbb{R}^{U \times V}$ is the mask of pixels treated with ℓ^2 norm while $\bar{\mathbf{M}} = (\bar{M}_{u,v} \in \mathbb{R}^{U \times V})$ is the mask for pixels treated with ℓ^1 norm:

$$(\mathbf{M})_{u,v} = \begin{cases} 1 & 0 \leq \left| (\nabla \tilde{S})_{u,v} \right| \leq \epsilon \\ 0 & \left| (\nabla \tilde{S})_{u,v} \right| > \epsilon \end{cases} \tag{62}$$

$$(\bar{\mathbf{M}})_{u,v} = \begin{cases} 0 & 0 \leq \left| (\nabla \tilde{S})_{u,v} \right| \leq \epsilon \\ 1 & \left| (\nabla \tilde{S})_{u,v} \right| > \epsilon \end{cases} \tag{63}$$

Therefore:

$$\frac{\partial F_{\text{reg}}}{\partial \tilde{S}_{u,v}} = -\frac{1}{\epsilon} \left[(\tilde{S}_{u+1,v} - \tilde{S}_{u,v}) + (\tilde{S}_{u,v+1} - \tilde{S}_{u,v}) \right] \cdot M_{u,v}$$
$$- \frac{(\tilde{S}_{u+1,v} - \tilde{S}_{u,v}) + (\tilde{S}_{u,v+1} - \tilde{S}_{u,v})}{\sqrt{(\tilde{S}_{u+1,v} - \tilde{S}_{u,v})^2 + (\tilde{S}_{u,v+1} - \tilde{S}_{u,v})^2}} \cdot \bar{M}_{u,v}$$
$$+ \frac{1}{\epsilon} (\tilde{S}_{u,v} - \tilde{S}_{u,v-1}) \cdot M_{u,v-1}$$
$$+ \frac{(\tilde{S}_{u,v} - \tilde{S}_{u,v-1})}{\sqrt{(\tilde{S}_{u+1,v-1} - \tilde{S}_{u,v-1})^2 + (\tilde{S}_{u,v} - \tilde{S}_{u,v-1})^2}} \cdot \bar{M}_{u,v-1}$$
$$+ \frac{1}{\epsilon} (\tilde{S}_{u,v} - \tilde{S}_{u-1,v}) \cdot M_{u-1,v}$$
$$+ \frac{(\tilde{S}_{u,v} - \tilde{S}_{u-1,v})}{\sqrt{(\tilde{S}_{u-1,v+1} - \tilde{S}_{u-1,v})^2 + (\tilde{S}_{u,v} - \tilde{S}_{u-1,v})^2}} \cdot \bar{M}_{u-1,v} \tag{64}$$

And $\nabla_S F_{\text{reg}}$ is the vectorized version of $\frac{\partial F_{\text{reg}}}{\partial \tilde{S}_{u,v}}$.

The Hessian matrix of the Huber norm, $J(\nabla_S F_{\text{reg}})(S)$, is a large $n \times n$ sparse matrix with $7 \times n$ non-zero values. These seven 2nd order partial derivatives over S are listed as follows:

$$\frac{\partial^2 F_{\text{reg}}}{\partial \tilde{S}_{u,v} \partial \tilde{S}_{u,v}} = \frac{2}{\epsilon} \cdot M_{u,v} + \frac{1}{\epsilon} \cdot M_{u,v-1} + \frac{1}{\epsilon} \cdot M_{u-1,v}$$
$$+ \frac{(\tilde{S}_{u+1,v} - \tilde{S}_{u,v+1})^2}{\left[(\tilde{S}_{u+1,v} - \tilde{S}_{u,v})^2 + (\tilde{S}_{u,v+1} - \tilde{S}_{u,v})^2 \right]^{\frac{2}{3}}} \cdot \bar{M}_{u,v}$$
$$+ \frac{(\tilde{S}_{u+1,v-1} - \tilde{S}_{u,v-1})^2}{\left[(\tilde{S}_{u+1,v-1} - \tilde{S}_{u,v-1})^2 + (\tilde{S}_{u,v} - \tilde{S}_{u,v-1})^2 \right]^{\frac{2}{3}}} \cdot \bar{M}_{u,v-1}$$
$$+ \frac{(\tilde{S}_{u-1,v+1} - \tilde{S}_{u-1,v})^2}{\left[(\tilde{S}_{u-1,v+1} - \tilde{S}_{u-1,v})^2 + (\tilde{S}_{u,v} - \tilde{S}_{u-1,v})^2 \right]^{\frac{2}{3}}} \cdot \bar{M}_{u-1,v} \tag{65}$$

$$\frac{\partial^2 F_{\text{reg}}}{\partial \tilde{S}_{u,v} \partial \tilde{S}_{u+1,v}} = -\frac{1}{\epsilon} \cdot M_{u,v}$$
$$- \frac{(\tilde{S}_{u,v+1} - \tilde{S}_{u,v})(\tilde{S}_{u,v+1} - \tilde{S}_{u+1,v})}{\left[(\tilde{S}_{u+1,v} - \tilde{S}_{u,v})^2 + (\tilde{S}_{u,v+1} - \tilde{S}_{u,v})^2 \right]^{\frac{2}{3}}} \cdot \bar{M}_{u,v} \tag{66}$$

$$\frac{\partial^2 F_{\text{reg}}}{\partial \tilde{S}_{u,v} \partial \tilde{S}_{u,v+1}} = -\frac{1}{\epsilon} \cdot M_{u,v}$$
$$- \frac{(\tilde{S}_{u+1,v} - \tilde{S}_{u,v})(\tilde{S}_{u+1,v} - \tilde{S}_{u,v+1})}{\left[(\tilde{S}_{u+1,v} - \tilde{S}_{u,v})^2 + (\tilde{S}_{u,v+1} - \tilde{S}_{u,v})^2 \right]^{\frac{2}{3}}} \cdot \bar{M}_{u,v} \tag{67}$$

$$\frac{\partial^2 F_{\text{reg}}}{\partial \tilde{S}_{u,v} \partial \tilde{S}_{u-1,v}} = -\frac{1}{\epsilon} \cdot M_{u-1,v}$$

$$-\frac{(\tilde{S}_{u-1,v+1}-\tilde{S}_{u-1,v})(\tilde{S}_{u-1,v+1}-\tilde{S}_{u,v})}{\left[(\tilde{S}_{u-1,v+1}-\tilde{S}_{u-1,v})^2+(\tilde{S}_{u,v}-\tilde{S}_{u-1,v})^2\right]^{\frac{2}{3}}}\cdot\overline{M}_{u-1,v}$$

$$(68)$$

$$\frac{\partial^2 F_{\text{reg}}}{\partial\tilde{S}_{u,v}\,\partial\tilde{S}_{u,v-1}}=-\frac{1}{\epsilon}\cdot M_{u,v-1}$$

$$-\frac{(\tilde{S}_{u+1,v-1}-\tilde{S}_{u,v-1})(\tilde{S}_{u+1,v-1}-\tilde{S}_{u,v})}{\left[(\tilde{S}_{u+1,v-1}-\tilde{S}_{u,v-1})^2+(\tilde{S}_{u,v}-\tilde{S}_{u,v-1})^2\right]^{\frac{2}{3}}}\cdot\overline{M}_{u,v-1}$$

$$(69)$$

$$\frac{\partial^2 F_{\text{reg}}}{\partial\tilde{S}_{u,v}\,\partial\tilde{S}_{u-1,v+1}}=$$

$$-\frac{(\tilde{S}_{u,v}-\tilde{S}_{u-1,v})(\tilde{S}_{u-1,v+1}-\tilde{S}_{u-1,v})}{\left[(\tilde{S}_{u-1,v+1}-\tilde{S}_{u-1,v})^2+(\tilde{S}_{u,v}-\tilde{S}_{u-1,v})^2\right]^{\frac{2}{3}}}\cdot\overline{M}_{u-1,v}$$

$$(70)$$

$$\frac{\partial^2 F_{\text{reg}}}{\partial\tilde{S}_{u,v}\,\partial\tilde{S}_{u+1,v-1}}=$$

$$-\frac{(\tilde{S}_{u+1,v-1}-\tilde{S}_{u,v-1})(\tilde{S}_{u,v}-\tilde{S}_{u,v-1})}{\left[(\tilde{S}_{u+1,v-1}-\tilde{S}_{u,v-1})^2+(\tilde{S}_{u,v}-\tilde{S}_{u,v-1})^2\right]^{\frac{2}{3}}}\cdot\overline{M}_{u,v-1}$$

$$(71)$$

In practice, these seven partial derivatives are estimated in vector form and together form the sparse Hessian matrix.

REFERENCES

[1] G. T. Herman, Fundamentals of computerized tomography: Image reconstruction from projection, 2nd ed, Springer, 2009

[2] P. A. Midgley, R. E. Dunin-Borkowski, Electron tomography and holography in materials science, Nat. Mater. 8 (2009) 271-280.

[3] F. Natterer, *The Mathematics of Computerized Tomography*, Society for Industrial and Applied Mathematics, 2001.

[4] A. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, Society for Industrial and Applied Mathematics, 2001.

[5] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

[6] G. –H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. Phys.*, vol. 35, no. 2, pp. 660-663, 2008.

[7] E. J. Candès, J. Romberg, T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489-509, Feb. 2006.

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

[9] Z. Tian, X. Jia, K. Yuan, T. Pan, and S. B. Jiang, "Low-dose CT reconstruction via edge-preserving total variation regularization," *Physics in Medicine and Biology*, vol. 56, no. 18, pp. 5949–5967, 2011.

[10] B. Goris, W. van den Broek, K. J. Batenburg, H. Heidari Mezerji, S. Bals, "Electron tomography based on a total variation minimization reconstruction technique," *Ultramicroscopy*, vol. 113, pp. 120-130, Feb. 2012.

[11] D. Laloum, T. Printemps, D.F. Sanchez, F. Lorut, G. Audoit, P. Bleuet, "Reduction of the scanning time by total variation minimization reconstruction for X-ray tomography in a SEM," *Journal of Microscopy*, vol. 256 (2), pp. 90-99, Nov. 2014.

[12] G. T. Herman, A. Kuba, Advances in Discrete Tomography and Its Applications, Birkhäuser Basel, 2007.

[13] K. J. Batenburg, J. Sijbers, DART: a practical reconstruction algorithm for discrete tomography, *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2542-2553, Sept. 2011.

[14] W. van Aarle, K. J. Batenburg, G. van Gompel, E. van de Casteele, J. Sijbers, "Super-Resolution for Computed Tomography Based on Discrete Tomography," *IEEE Trans. Image Process.*, vol. 23, pp. 1181-1193, Jan. 2014.

[15] S. Van Aert, K. J. Batenburg, M. D. Rossell, R. Erni, and G. Van Tendeloo, "Three-dimensional atomic imaging of crystalline nanoparticles", *Nature*, 470, 374-377 (2011).

[16] K. J. Batenburg, S. Bals, J. Sijbers, C. Kuebel, P. A. Midgley, J. C. Hernandez, U. Kaiser, E. R. Encina, E. A. Coronado, and G. Van Tendeloo, "3D imaging of nanomaterials by discrete tomography," *Ultramicroscopy*, 109(6), 730-740 (2009).

[17] W. van Aarle, K. J. Batenburg, J. Sijbers, "Automatic parameter estimation for the discrete algebraic reconstruction technique (DART)," *IEEE Trans. Image Process.*, vol. 21, pp. 4608-4621, 2012.

[18] F. Bleichrodt, F. Tabak, K. J. Batenburg, "SDART: An algorithm for discrete tomography from noisy projections," *Computer Vision and Image Understanding*, vol. 129, pp. 63-74, Dec. 2014.

[19] P. J. Huber, "Robust regression: Asymptotics, conjectures, and Monte Carlo", *Ann. Statist.*, vol. 1, pp.799–821, 1973.

[20] E. A. Rashed, and H. Kudo, "Statistical image reconstruction from limited projection data with intensity priors," *Phys. Med. Biol.*, vol. 57, pp. 2039-2061, 2012.

[21] T. Schule, C. Schnorr, S. Weber and J. Hornegger, "Discrete tomography by convex-concave regularizationand D.C programming," *Discr. Appl. Math.*, vol. 151, pp.229 -243, 2005.

[22] E. Y. Sidky, R. Chartrand, and X. Pan, "Image reconstruction from few views by non-convex optimization," in *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS '07)*, pp. 3526–3530, Honolulu, Hawaii, USA, November 2007.

[23] T. Blumensath, and R. Boardman, "Non-convexly constrained image reconstruction from nonlinear tomographic X-ray measurements," *Phil. Trans. R. Soc. A*, 373(2043), 20140393.

[24] J. C. R. Giraldo, J. D. Trzasko, S. Leng, C. H. McCollough, and A. Manduca, "Non-convex prior image constrained compressed sensing," In *SPIE Medical Imaging*, pp. 76222C-76222C, March 2010.

[25] B. Goris, T. Roelandts, K. J. Batenburg, H. H. Mezerji, and S. Bals, "Advanced reconstruction algorithms for electron microscopy: From comparison to combination," *Ultramicroscopy*, vol. 127, pp.40-47, April 2013.

[26] W. van Aarle, W. J. Palenstijn, et al., "The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography", *Ultramicroscopy*, vol. 157, pp.35-47, October 2015.

[27] A. Chambolle, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, pp. 120-145, 2011.