Journal of Mathematical Imaging and Vision (2021) 63:133–143
https://doi.org/10.1007/s10851-020-00972-7

Computed Tomography Reconstruction with Uncertain View Angles by Iteratively Updated Model Discrepancy

Nicolai André Brogaard Riis[1] · **Yiqiu Dong**[1] · **Per Christian Hansen**[1]

Received: 15 October 2019 / Accepted: 8 June 2020 / Published online: 4 July 2020

Abstract
We propose a new model and a corresponding iterative algorithm for Computed Tomography (CT) when the view angles are uncertain. The uncertainty is described by an additive model discrepancy term which is included in the data fidelity term of a total variation regularized variational model. We approximate the model discrepancy with a Gaussian distribution. Our iterative algorithm alternates between updating the CT reconstruction and parameters of the model discrepancy. By assuming that the uncertainties in the view angles are independent we achieve a covariance matrix structure that we can take advantage of in a stochastic primal dual method to greatly reduce the computational work compared to classical primal dual methods. Using simulations with 2D problems we demonstrate that our method is able to reduce the reconstruction error and improve the visual quality, compared to methods that ignore the uncertainties in the angles.

Keywords Computed Tomography · Uncertain view angles · Model error · Model discrepancy

Mathematics Subject Classification 65F22 · 65K10

1 Introduction

In this paper we consider Computed Tomography (CT) reconstruction where the view angles are only known approximately. This uncertainty may arise from several sources, e.g., inaccuracies in the physical set-up and inexact estimates from a calibration procedure. The goal is to reconstruct a CT image in such a way that we take into account both measurement noise and the uncertainty associated with the view angles.

We formulate the *CT reconstruction problem with uncertain view angles* as estimating the unknown attenuation coefficient image, represented by the vector $\mathbf{x} \in \mathbb{R}^n$, from a measured and noisy sinogram represented by the vector $\mathbf{b} \in \mathbb{R}^m$. We use the model

$$\mathbf{b} = \mathbf{R}(\boldsymbol{\theta})\,\mathbf{x} + \mathbf{e}, \quad \boldsymbol{\theta} \sim \pi_{\text{angles}}(\cdot), \quad \mathbf{e} \sim \pi_{\text{noise}}(\cdot), \tag{1}$$

where the uncertainty in the view angles and the measured sinogram are characterized by the probability distributions π_{angles} and π_{noise}, respectively.

The CT forward model (the forward projection) is represented by the matrix $\mathbf{R}(\boldsymbol{\theta}) \in \mathbb{R}^{m\times n}$, which is the discrete approximation of the Radon transform $\mathcal{R}$ explicitly parameterized by the view angles $\boldsymbol{\theta} \in \mathbb{R}^q$. Specifically, if f is the continuous representation of the attenuation image, the discretization satisfies

$$(\mathbf{R}(\theta_i)\mathbf{x})_l \approx (\mathcal{R}f)(\theta_i, \mathrm{s}_l) = \int_{\mathbb{R}} f(\mathrm{s}_l \overrightarrow{\mathbf{v}}(\theta_i) + t\,\overrightarrow{\mathbf{v}}^{\perp}(\theta_i))\,\mathrm{d}t, \tag{2}$$

where s_l with $l = 1, \ldots, p$ is the position of the lth pixel on the detector, θ_i with $i = 1, \ldots, q$ is the view angle, and hence $m = qp$. Moreover, $\overrightarrow{\mathbf{v}}(\theta) = (\cos\theta, \sin\theta)$ and $\overrightarrow{\mathbf{v}}^{\perp}(\theta) = \overrightarrow{\mathbf{v}}(\theta + \pi/2)$. For more details on the physical and mathematical models of CT see, e.g., [5,17].

This work was supported by a Villum Investigator Grant (No. 25893) from The Villum Foundation and by the National Natural Science Foundation of China via Grant 11701388.

✉ Nicolai André Brogaard Riis
nabr@dtu.dk

Yiqiu Dong
yido@dtu.dk

Per Christian Hansen
pcha@dtu.dk

[1] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark

We emphasize that the goal in this work is to reconstruct the CT image $\mathbf{x}$ from a measured sinogram $\mathbf{b}$ according to the model (1) with uncertain view angles $\boldsymbol{\theta}$ and noise $\mathbf{e}$. We consider $\boldsymbol{\theta}$ and $\mathbf{e}$ as nuisance or uninteresting parameters, and they are only taken into account when reconstructing $\mathbf{x}$ without being explicitly estimated.

While our work focuses on 2D CT reconstruction, our methodology can also be applied to 3D CT and other applications, because there are no restrictions on the matrix in the forward model. We emphasize the computational advantage of having a block representation of the covariance matrix for the model discrepancy introduced in the next section (cf. Sect. 3).

1.1 Related Work

Variational methods have been proposed for CT reconstruction problems that, in addition to a data fidelity term which incorporates the noise model, explicitly incorporate prior information via a regularization term, see the survey [3]. Different regularization techniques can be applied, for example, Tikhonov regularization [23,29] or total variation (TV) regularization [26,27]. But these methods do not take parametric uncertainty in the model into account. Therefore, good performance of these methods is not guaranteed if the view angles are uncertain.

Reconstruction methods that aim to estimate the view angles $\boldsymbol{\theta}$ in addition to the CT image $\mathbf{x}$ using only the measured sinogram without reference objects or instruments are proposed in [2,10,18,22]. These methods can be categorized into two groups: estimating view angles directly from projection data and then estimating the CT image, and simultaneously estimating view angles and CT image. In the former case an inaccurate angle estimation could lead to an unsatisfactory reconstruction due to error propagation. Algorithms for the latter case (such as the Bayesian sampling-based method [22]) can effectively avoid error propagation, but they are limited by large computing times which makes them unfeasible for large-scale problems.

There are also a few methods for characterizing and reducing model errors in general inverse problems, see [6,11,12,15,16,21]. Most of these methods are based on Bayesian sampling-based methods which may also suffer from large computing times, except in the special case of a Gaussian prior for the solution where a closed-form expression exist.

1.2 Our Contribution

We propose a new model and algorithm for CT reconstruction when the view angles are uncertain. Our model incorporates the uncertainty of the view angles in the data fidelity term of a variational method. This provides a CT reconstruction where the uncertainty is accounted for and we avoid error propagation without the extra cost of simultaneously estimating the view angles and CT image.

Compared to our preliminary work in [25] the main contribution in this work is the formulation and utilization of block-structure of the computational problem. This allows us to utilize a stochastic primal dual hybrid gradient method [9] to significantly reduce the computational work and allows the method to solve large-scale problems. Moreover, we use a numerically stable and efficient approach to factorize and invert the covariance matrix required in our data fidelity term.

2 Iteratively Updated Model Discrepancy

In this section we summarize our previous work in [25] where we reformulate (1) by fixing the forward model with an estimate $\hat{\boldsymbol{\theta}}$ of the view angles. This is done to avoid dealing with a distribution of models $\mathbf{R}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \sim \pi_{\text{angles}}(\cdot)$ in the CT reconstruction. The angles $\hat{\boldsymbol{\theta}}$ are set to the nominal angles of the scanner, and we obtain

$$\mathbf{b} = \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x} + \boldsymbol{\eta} + \mathbf{e},\ \boldsymbol{\eta} \sim \pi_{\text{discrep}}(\cdot),\ \mathbf{e} \sim \pi_{\text{noise}}(\cdot), \tag{3}$$

where $\boldsymbol{\eta}$ acquires its push-forward distribution from

$$\boldsymbol{\eta} = \eta(\boldsymbol{\theta}, \mathbf{x}) = \mathbf{R}(\boldsymbol{\theta})\,\mathbf{x} - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x}, \tag{4}$$

and we call $\boldsymbol{\eta}$ the *model discrepancy term*. The advantage of doing so is that the uncertainty in the forward model is moved into $\boldsymbol{\eta}$ and the CT model $\mathbf{R}(\hat{\boldsymbol{\theta}})$ is fixed. If we ignore the fact that $\boldsymbol{\eta}$ depends on $\mathbf{x}$ and consider $\boldsymbol{\eta}$ as independent additive noise, then (3) becomes a standard CT reconstruction model with the fixed view angles $\hat{\boldsymbol{\theta}}$.

Representing the model uncertainties as an additive model discrepancy term was first proposed in [14], where $\boldsymbol{\eta}$ was assumed a Gaussian process and used for calibration from a physical model to a computational model. In [12,13] a similar idea was applied to Bayesian inverse problems and referred to as the approximation error approach (AEA), in which $\boldsymbol{\eta}$ denotes the model discrepancy. This approach is used successfully in a number of different imaging applications such as fine-to-coarse mesh approximation in diffuse optical tomography [1], unknown domain boundaries in electric impedance tomography [20], unknown scattering in both diffuse optical tomography and quantitative photoacoustic tomography [15,24], truncation errors in magnetic particle imaging [4] and uncertain sound speed in photoacoustic tomography [28].

Inspired by AEA, we derive the likelihood for (3) by marginalizing both $\boldsymbol{\eta}$ and $\mathbf{e}$. Let $\boldsymbol{\nu} = \boldsymbol{\eta} + \mathbf{e}$, then

$$\begin{aligned}\pi(\mathbf{b}|\mathbf{x}) &= \int_{\mathbb{R}^m} \pi(\mathbf{b}, \boldsymbol{\nu}|\mathbf{x})\mathrm{d}\boldsymbol{\nu} \\ &= \int_{\mathbb{R}^m} \pi(\mathbf{b}|\mathbf{x}, \boldsymbol{\nu})\pi(\boldsymbol{\nu}|\mathbf{x})\mathrm{d}\boldsymbol{\nu} \\ &= \pi_{\boldsymbol{\nu}|\mathbf{x}}(\mathbf{b} - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x}|\mathbf{x}). \end{aligned} \tag{5}$$

This formulation raises two main issues:

(1) The distribution $\pi_{\boldsymbol{\nu}|\mathbf{x}}$ can be rather complicated and may not have a closed-form expression, so the evaluation of the likelihood becomes difficult.
(2) The distribution $\pi_{\boldsymbol{\nu}|\mathbf{x}}$ depends on the unknown $\mathbf{x}$.

To deal with the former issue, we introduce a Gaussian approximation to $\pi_{\boldsymbol{\nu}|\mathbf{x}}$, i.e., we simplify the model. The latter issue, on the other hand, is dealt with from an algorithmic point of view by using a computational method that alternates between updating $\pi_{\boldsymbol{\nu}|\mathbf{x}}$ and $\mathbf{x}$.

2.1 Gaussian Approximation

The distribution $\pi_{\boldsymbol{\nu}|\mathbf{x}}$ may not have a closed-form expression, so we approximate it by a simple distribution. Experiments suggest that Gaussian approximations are useful in many applications related to model discrepancies [1,4,6,11–13,15,19,20,24,28]. In this work, we therefore assume $\boldsymbol{\eta}|\mathbf{x}$ follows a Gaussian distribution, i.e.,

$$\boldsymbol{\eta}|\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\eta}|\mathbf{x}}, \mathbf{C}_{\boldsymbol{\eta}|\mathbf{x}}) \tag{6}$$

with mean $\boldsymbol{\mu}_{\boldsymbol{\eta}|\mathbf{x}}$ and covariance $\mathbf{C}_{\boldsymbol{\eta}|\mathbf{x}}$, both depending on $\mathbf{x}$. Moreover, we assume

$$\mathbf{e} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{e}}, \mathbf{C}_{\mathbf{e}}), \tag{7}$$

with mean $\boldsymbol{\mu}_{\mathbf{e}}$ and covariance $\mathbf{C}_{\mathbf{e}}$, which are independent on $\mathbf{x}$. Following these two assumptions, we obtain the distribution of $\boldsymbol{\nu}|\mathbf{x}$

$$\pi_{\boldsymbol{\nu}|\mathbf{x}} = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\eta}|\mathbf{x}} + \boldsymbol{\mu}_{\mathbf{e}}, \mathbf{C}_{\boldsymbol{\eta}|\mathbf{x}} + \mathbf{C}_{\mathbf{e}}). \tag{8}$$

Then, the likelihood (5) admits a closed-form expression. By taking the negative logarithm we get

$$\begin{aligned}-\log \pi(\mathbf{b}|\mathbf{x}) &\propto \tfrac{1}{2}\|\mathbf{b} - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}}\|^2_{\mathbf{C}^{-1}_{\boldsymbol{\nu}|\mathbf{x}}} \\ &= \tfrac{1}{2}\|\mathbf{L}_{\boldsymbol{\nu}|\mathbf{x}}(\mathbf{b} - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}})\|_2^2, \end{aligned} \tag{9}$$

where $\boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}} = \boldsymbol{\mu}_{\boldsymbol{\eta}|\mathbf{x}} + \boldsymbol{\mu}_{\mathbf{e}}$, $\mathbf{C}_{\boldsymbol{\nu}|\mathbf{x}} = \mathbf{C}_{\boldsymbol{\eta}|\mathbf{x}} + \mathbf{C}_{\mathbf{e}}$, and $\mathbf{L}^T_{\boldsymbol{\nu}|\mathbf{x}}\mathbf{L}_{\boldsymbol{\nu}|\mathbf{x}} = \mathbf{C}^{-1}_{\boldsymbol{\nu}|\mathbf{x}}$ is the Cholesky factorization of the inverse of the combined covariance matrix.

2.2 Alternate Updates

Since the distribution $\pi_{\boldsymbol{\nu}|\mathbf{x}}$ depends on the unknown $\mathbf{x}$, the question is how the mean $\boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}}$ and the covariance $\mathbf{C}_{\boldsymbol{\nu}|\mathbf{x}}$ of the combined uncertainties can be determined.

Considering a given reconstruction $\hat{\mathbf{x}}$, one can generate samples of $\boldsymbol{\eta}|\mathbf{x}$ with $\mathbf{x} = \hat{\mathbf{x}}$ by drawing samples $\boldsymbol{\theta}^s$ with $s = 1, \ldots, S$ following the distribution π_{angles}, and evaluate the model discrepancy term by

$$\boldsymbol{\eta}^s_{\hat{\mathbf{x}}} = \mathbf{R}(\boldsymbol{\theta}^s)\,\hat{\mathbf{x}} - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\hat{\mathbf{x}}, \quad s = 1, \ldots, S. \tag{10}$$

Then, the sample mean and sample covariance can be calculated by

$$\boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}} = \frac{1}{S}\sum_{s=1}^{S} \boldsymbol{\eta}^s_{\hat{\mathbf{x}}} \tag{11}$$

and

$$\mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}} = \frac{1}{S-1}\sum_{s=1}^{S} (\boldsymbol{\eta}^s_{\hat{\mathbf{x}}} - \boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}})(\boldsymbol{\eta}^s_{\hat{\mathbf{x}}} - \boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}})^T. \tag{12}$$

In (9), we can then use the sample mean and the sample covariance to approximate $\boldsymbol{\mu}_{\boldsymbol{\eta}|\mathbf{x}}$ and $\mathbf{C}_{\boldsymbol{\eta}|\mathbf{x}}$, respectively.

Good estimation of the model discrepancy term strongly relies on a reconstruction that resembles the ground truth. Variational methods with TV regularization have shown good performance in large-scale CT reconstruction [27]. Adding a TV regularization term to the negative log-likelihood in (9) we obtain the variational model

$$\min_{\mathbf{x}\geq 0} \tfrac{1}{2}\|\mathbf{L}_{\boldsymbol{\nu}|\mathbf{x}}(\mathbf{b} - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}})\|_2^2 + \lambda \mathrm{TV}(\mathbf{x}), \tag{13}$$

where $\lambda > 0$ denotes the regularization parameter, and the TV term is defined as

$$\mathrm{TV}(\mathbf{x}) = \|\nabla \mathbf{x}\|_{2,1} \equiv \sum_{i=1}^{n} \|[\nabla \mathbf{x}]_i\|_2\,, \tag{14}$$

in which $[\nabla \mathbf{x}]_i$ denotes the discrete gradient of $\mathbf{x}$ at the ith pixel, computed via a forward difference scheme with reflexive boundary conditions. The non-negativity constraint represents the fact that the attenuation coefficients $\mathbf{x}$ cannot be negative. Since $\mathbf{L}_{\boldsymbol{\nu}|\mathbf{x}}$ depends on the unknown $\mathbf{x}$, the objective function in (13) is non-convex, which means that a solution relies on the initialization as well as the numerical algorithm.

According to (13), the reconstruction $\mathbf{x}$ also depends on $\boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}}$ and $\mathbf{C}_{\boldsymbol{\nu}|\mathbf{x}}$. To further strengthen the relation between $\mathbf{x}$ and $(\boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}}, \mathbf{C}_{\boldsymbol{\nu}|\mathbf{x}})$ we introduce an alternately updating scheme **MD-TV** shown in Algorithm 1.

Algorithm 1: Iteratively Updated Model Discrepancy with TV (MD-TV)

Input: $\mathbf{b}$, λ, $\hat{\boldsymbol{\theta}}$, π_{angles}, $\boldsymbol{\mu}_{\mathbf{e}}$, $\mathbf{C}_{\mathbf{e}}$. Initialize $\mathbf{L}^0_{\boldsymbol{\nu}|\mathbf{x}}$, $\boldsymbol{\mu}^0_{\boldsymbol{\nu}|\mathbf{x}}$.
Output: $\mathbf{x}$

1: **for** $k = 1, 2, \ldots, K$
2: $\quad \mathbf{x}^k = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2}\|\mathbf{L}^{k-1}_{\boldsymbol{\nu}|\mathbf{x}}(\mathbf{b} - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x} - \boldsymbol{\mu}^{k-1}_{\boldsymbol{\nu}|\mathbf{x}})\|_2^2 + \lambda \mathrm{TV}(\mathbf{x})$
3: $\quad$ **for** $s = 1, 2, \ldots, S$
4: $\quad\quad \boldsymbol{\eta}^s_{\mathbf{x}^k} = \mathbf{R}(\boldsymbol{\theta}^s)\,\mathbf{x}^k - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x}^k$ with $\boldsymbol{\theta}^s \sim \pi_{\text{angles}}(\cdot)$
5: $\quad$ **end**
6: $\quad \boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\mathbf{x}^k} = \frac{1}{S}\sum_{s=1}^{S} \boldsymbol{\eta}^s_{\mathbf{x}^k}$
7: $\quad \mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\mathbf{x}^k} = \frac{1}{S-1}\sum_{s=1}^{S}(\boldsymbol{\eta}^s_{\mathbf{x}^k} - \boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\mathbf{x}^k})(\boldsymbol{\eta}^s_{\mathbf{x}^k} - \boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\mathbf{x}^k})^T$
8: $\quad \boldsymbol{\mu}^k_{\boldsymbol{\nu}|\mathbf{x}} = \boldsymbol{\mu}_{\mathbf{e}} + \boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\mathbf{x}^k}$
9: $\quad \mathbf{L}^k_{\boldsymbol{\nu}|\mathbf{x}} = \mathrm{chol}\big((\mathbf{C}_{\mathbf{e}} + \mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\mathbf{x}^k})^{-1}\big)$
10: **end**

Here, $\mathbf{L} = \mathrm{chol}(\mathbf{C}^{-1})$ gives the Cholesky factor of the inverse covariance $\mathbf{C}^{-1}$, i.e., $\mathbf{L}^T\mathbf{L} = \mathbf{C}^{-1}$. Note that although the minimization problem for obtaining $\mathbf{x}$ is convex now, it does not guarantee that the overall algorithm converges.

3 Block-Wise Representation

In each iteration of Algorithm 1 we require S forward projections to generate samples (10) of $\boldsymbol{\eta}|\mathbf{x}$ in step 4, which can dominate the cost of each iteration for large S. Moreover, even for small S the algorithm can be of limited use in solving large-scale CT problems in its "naive" form due to the amount of work in solving the TV optimization problem in step 2 and inverting and factorizing the covariance matrix $\mathbf{C}_{\mathbf{e}} + \mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}} \in \mathbb{R}^{m\times m}$ in step 9.

In our previous work [25], the latter issue was partly addressed by a block-diagonal representation of the covariance matrix. This greatly reduces the computational burden of the inversion and factorization and it reduces the cost of multiplication with the factorized matrix when solving the TV optimization problem. In this work, we further extend these ideas by reformulating the entire variational model using the block diagonal representation. Our new formulation can thus benefit from algorithms that utilize this structure as well as parallelization in the block-wise sampling, inversion and factorization of the covariance matrix.

3.1 Block Covariance Matrix Representation

The covariance matrix $\mathbf{C}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}}$ is block-diagonal when the view angles are independent. The sampled matrix in (12) may not be block diagonal due to the finite sampling.

Our experience is that the norm of the off-diagonal blocks decreases as the number of samples S increases. Therefore, we replace the sample covariance matrix $\mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}} \in \mathbb{R}^{m\times m}$ by a block diagonal matrix $\widetilde{\mathbf{C}}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}}$ given by

$$\widetilde{\mathbf{C}}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}} = \begin{bmatrix} \mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}},1} & & \\ & \ddots & \\ & & \mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}},q} \end{bmatrix}, \tag{15}$$

where $\mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}},i} \in \mathbb{R}^{p\times p}$ are the block diagonal parts of $\mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}}$. Then, assuming the Gaussian measurement noise is independent identically distributed, i.e., $\mathbf{C}_{\mathbf{e}} = \sigma^2\mathbf{I}$, we can compute the Cholesky factor of the inverse covariance matrix $(\mathbf{C}_{\mathbf{e}} + \widetilde{\mathbf{C}}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}})^{-1}$ block-wise as follows

$$\widetilde{\mathbf{L}}^{\text{sample}}_{\boldsymbol{\nu}|\mathbf{x}=\hat{\mathbf{x}}} = \begin{bmatrix} \mathrm{chol}\big((\sigma^2\mathbf{I}_p + \widetilde{\mathbf{C}}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}},1})^{-1}\big) & & \\ & \ddots & \\ & & \mathrm{chol}\big((\sigma^2\mathbf{I}_p + \widetilde{\mathbf{C}}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}},q})^{-1}\big) \end{bmatrix}. \tag{16}$$

These aspects are illustrated in Fig. 1, which shows that the block-diagonal representation captures the structure of the actual covariance matrix with fewer samples compared to the full version.

Furthermore, to achieve a computationally efficient and numerically stable inversion and factorization, we note that we can interchange the inversion and factorization and utilize that the covariance matrix can be written as a low-rank update of a scaled identity matrix,

$$\widetilde{\mathbf{C}}^{\text{sample}}_{\boldsymbol{\nu}|\mathbf{x}=\hat{\mathbf{x}}} = \sigma^2\mathbf{I}_m + \sum_{s=1}^{S} \mathbf{w}_s\mathbf{w}_s^T, \tag{17}$$

where $\mathbf{w}_s = \frac{\boldsymbol{\eta}^s_{\hat{\mathbf{x}}} - \boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}}}{\sqrt{S-1}}$. Therefore, the computation of the inverse Cholesky factors that constitute the blocks in $\widetilde{\mathbf{L}}^{\text{sample}}_{\boldsymbol{\nu}|\mathbf{x}=\hat{\mathbf{x}}}$ merely consists of low-rank updates of $\sigma^{-2}\mathbf{I}_m$. Updating an inverse Cholesky factor is discussed in, e.g., [8]; this approach is numerically more stable than using the Sherman-Morrison formula for updating the inverse covariance matrix.

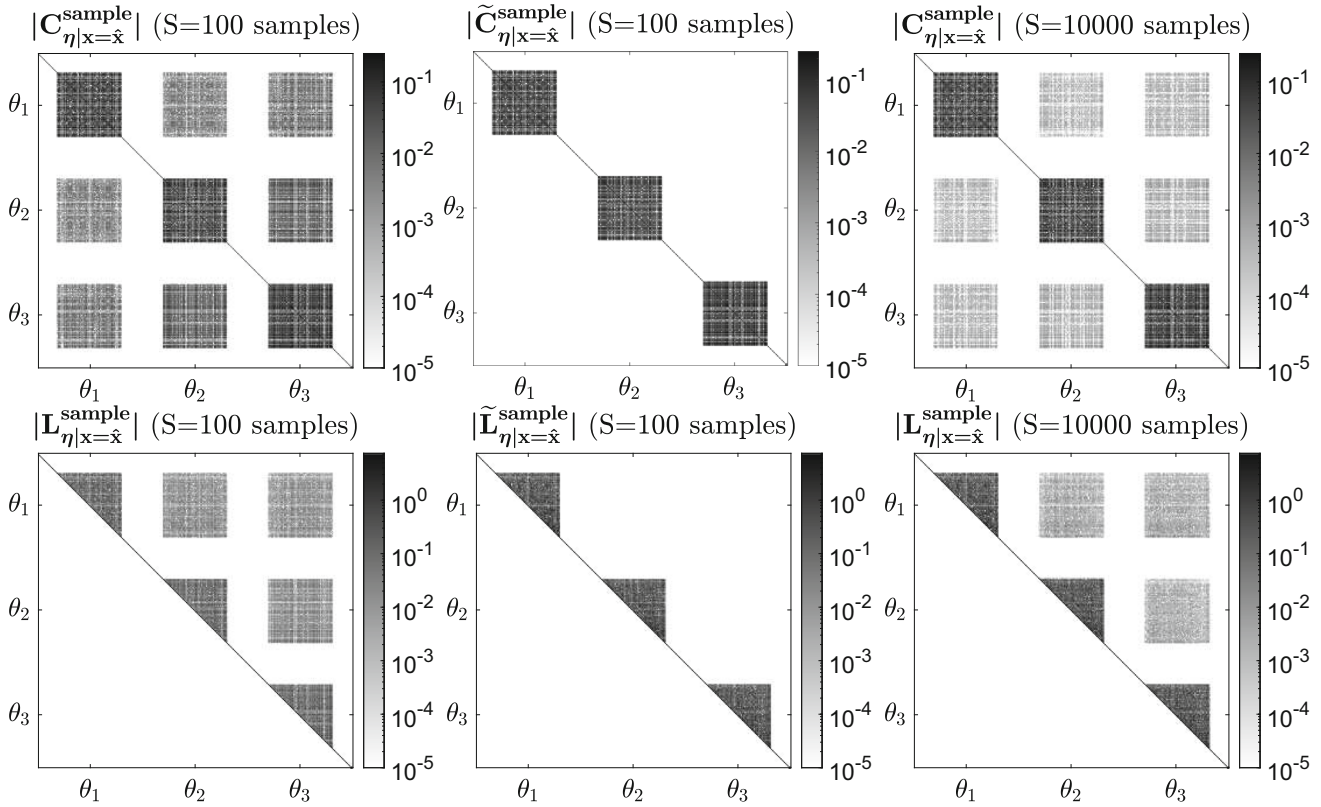


Fig. 1 These plots show the small top-left submatrix of the entire matrices. Each plot shows the absolute value of the matrix elements, and to visualize the zero-nonzero structure we use a logarithmic colormap. Left: the full sampled covariance matrix $\mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}}$ and the Cholesky factor $\mathbf{L}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}}$ of its inverse with $S = 100$ samples. Middle: the block-diagonal representation $\widetilde{\mathbf{C}}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}}$ according to (15) and the Cholesky factor $\widetilde{\mathbf{L}}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}}$ of its inverse with $S = 100$ samples. Right: full sampled covariance matrix and Cholesky factor of its inverse with $S = 10000$ samples. We see that the block-diagonal representation captures the structure of the actual covariance matrix with fewer samples compared to the full version

3.2 Block-Wise Variational Model

Given the block-wise approximation of the covariance matrix, the variational model in (13) can be formulated as

$$\min_{\mathbf{x}\geq\mathbf{0}} \frac{1}{2} \left\| \begin{bmatrix} \widetilde{\mathbf{L}}_{\boldsymbol{v}|\mathbf{x},1} & & \\ & \ddots & \\ & & \widetilde{\mathbf{L}}_{\boldsymbol{v}|\mathbf{x},q} \end{bmatrix} \left(\begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_q \end{bmatrix} - \begin{bmatrix} \mathbf{R}(\hat{\theta}_1) \\ \vdots \\ \mathbf{R}(\hat{\theta}_q) \end{bmatrix} \mathbf{x} - \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{v}|\mathbf{x},1} \\ \vdots \\ \boldsymbol{\mu}_{\boldsymbol{v}|\mathbf{x},2} \end{bmatrix} \right) \right\|_2^2 + \lambda \mathrm{TV}(\mathbf{x}), \tag{18}$$

where we use the short-hand notation $\boldsymbol{\mu}_{\boldsymbol{v}|\mathbf{x},i}$ and $\widetilde{\mathbf{L}}_{\boldsymbol{v}|\mathbf{x},i}$ for the blocks of the sample mean (11) and Cholesky factor of the inverse sample covariance matrix (16).

We can conveniently utilize the block structure to arrive at a variational model using q data-fitting terms, i.e.,

$$\min_{\mathbf{x}\geq 0} \frac{1}{2} \sum_{i=1}^{q} \| \widetilde{\mathbf{L}}_{\boldsymbol{v}|\mathbf{x},i} (\mathbf{b}_i - \mathbf{R}(\hat{\theta}_i)\, \mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{v}|\mathbf{x},i}) \|_2^2 + \lambda \mathrm{TV}(\mathbf{x}). \tag{19}$$

This formulation allows us to take the block structure into account via a stochastic optimization algorithm, namely, the stochastic primal dual hybrid gradient (SPDHG) method [9]. This greatly reduces the number of outer iterations and thus the overall computational work, compared to classical primal dual methods. The SPDHG algorithm solves the generic optimization problem

$$\min_{\mathbf{x}} \sum_{i=1}^{q+1} f_i(\mathbf{A}_i \mathbf{x}) + g(\mathbf{x}), \tag{20}$$

where for (19) we specifically have

$$f_{1,\dots,q}(\cdot) = \tfrac{1}{2}\|\cdot + \widetilde{\mathbf{L}}_{\boldsymbol{v}|\mathbf{x},i}(\mathbf{b}_i - \boldsymbol{\mu}_{\boldsymbol{v}|\mathbf{x},i})\|_2^2, \tag{21}$$

$$f_{q+1}(\cdot) = \lambda\,\|\cdot\|_{2,1}, \tag{22}$$

$$\mathbf{A}_{1,\dots,q} = -\tilde{\mathbf{L}}_{\boldsymbol{v}|\mathbf{x},i}\mathbf{R}(\hat{\theta}_i), \tag{23}$$

$$\mathbf{A}_{q+1} = \nabla, \tag{24}$$

$$g(\cdot) = \begin{cases} 0 & \text{if } x_i \geq 0, \\ \infty & \text{if } x_i < 0. \end{cases} \tag{25}$$

The procedure for solving the generic problem (20) is shown in Algorithm 2, where we specially use the proximal operators:

$$\operatorname{prox}_g^{\tau}(\mathbf{x}) = \max(\mathbf{x}, \mathbf{0}), \tag{26}$$

$$\operatorname{prox}_{f_i^*}^{\omega}(\mathbf{y}) = \frac{1}{1+\omega}\left(\mathbf{y} + \omega\widetilde{\mathbf{L}}_{\boldsymbol{v}|\mathbf{x},i}(\mathbf{b}_i - \boldsymbol{\mu}_{\boldsymbol{v}|\mathbf{x},i})\right), \quad i = 1,\dots,q, \tag{27}$$

$$\operatorname{prox}_{f_{q+1}^*}^{\omega}(\mathbf{y}) = \frac{1}{\lambda}\frac{\mathbf{y}}{\max(1, \|\mathbf{y}\|_2)}. \tag{28}$$

Algorithm 2: SPDHG

Input: Initial $\mathbf{x}$, step parameters ω_i, τ.
Output: $\mathbf{x}$

1: $\mathbf{y} = \mathbf{0}, \mathbf{z} = \bar{\mathbf{z}} = \mathbf{A}^T\mathbf{y} = \mathbf{0}$
2: **for** $k = 1, \dots$
3: $\quad \mathbf{x} = \operatorname{prox}_g^{\tau}(\mathbf{x} - \tau\bar{\mathbf{z}})$
4: $\quad$ Select $i \in 1, \dots, q$ with probability P_i
5: $\quad \mathbf{y}_i^+ = \operatorname{prox}_{f_i^*}^{\omega_i}(\mathbf{y}_i + \omega_i\mathbf{A}_i\mathbf{x})$
6: $\quad \boldsymbol{\Delta}\mathbf{z} = \mathbf{A}_i^T(\mathbf{y}_i^+ - \mathbf{y}_i)$
7: $\quad \mathbf{z} = \mathbf{z} + \boldsymbol{\Delta}\mathbf{z}, \quad \mathbf{y} = \mathbf{y}^+$
8: $\quad \bar{\mathbf{z}} = \mathbf{z} + \frac{1}{P_i}\boldsymbol{\Delta}\mathbf{z}$
9: **end**

A key choice here is the selection of the probabilities P_i in step 4. First note that if $P_i = 1$ then we obtain the standard the primal dual hybrid gradient (PDHG) aka. the Chambolle-Pock algorithm [7]. However, if $P_i < 1$ only a few blocks are updated in the dual variable before $\mathbf{x}$ is updated, which reduced the computational work in each iteration. To ensure that the algorithm puts equal weight on both regularization and the data-fitting blocks, we choose

$$P_i = \begin{cases} \frac{1}{2q} & \text{if } \ i = 1,\dots,q, \\ \frac{1}{2} & \text{if } \ i = q+1. \end{cases} \tag{29}$$

The step sizes are selected according to

$$\omega_i = \gamma\frac{\rho}{\|\mathbf{A}_i\|_2}\mathbf{I}, \quad \tau_i = \gamma^{-1}\frac{\rho P_i}{\|\mathbf{A}_i\|_2}\mathbf{I}, \quad \tau = \min\tau_i, \tag{30}$$

where $\rho < 1$ and $\gamma > 0$. In our implementation we choose $\gamma = 1$ and $\rho = 0.999$, i.e., we put equal balance between primal and dual variable and as large a step size as possible.

4 Numerical Experiments

We present simulated numerical experiments in 2D to show the performance of our method and compare it to our previous non-block version in [25]. The simulations are carried out in MATLAB and we use the ASTRA Toolbox [30] for the matrix-free forward and back projections, i.e., for multiplication with $\mathbf{R}(\boldsymbol{\theta})$ and $\mathbf{R}(\boldsymbol{\theta})^T$. The block formulation involves multiplication with the block matrices $\mathbf{R}(\theta_i) \in \mathbb{R}^{p\times n}$ and their transpose for $i = 1, \dots q$.

We assume that the view angles are independent and distributed according to $\pi_{\text{angles}} = \mathcal{N}(\boldsymbol{\theta}^{\text{equid}}, \delta^2\mathbf{I})$, where $\boldsymbol{\theta}^{\text{equid}}$ are equidistant view angles in [0° to 360°). We generate the noisy data according to

$$\mathbf{b} = \mathbf{R}(\bar{\boldsymbol{\theta}})\bar{\mathbf{x}} + \mathbf{e}, \tag{31}$$

where $\bar{\mathbf{x}}$ represents MATLAB's Shepp-Logan phantom and $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ with $\sigma = 0.005\,\|\mathbf{R}(\bar{\boldsymbol{\theta}})\bar{\mathbf{x}}\|_2/\sqrt{m}$. Moreover, $\bar{\boldsymbol{\theta}}$ is a realization of $\mathcal{N}(\boldsymbol{\theta}^{\text{equid}}, \delta^2\mathbf{I})$.

As in [25] we compare our results with a TV reconstruction that does not take the uncertainty into account, and a TV reconstruction that uses the true view angles $\bar{\boldsymbol{\theta}}$. That is, we compare

Table 1 The physical and discretization parameters in the simulated CT experiments

Parameter	Value
Scan geometry	Fan-beam
Reconstruction domain size	50 cm × 50 cm
Source to center distance	50 cm
Source to detector distance	100 cm
Detector length	130 cm
Small example	
Image pixels	$n = 45^2$
Detector pixels	$p = 90$
Number of view angles	$q = 90$
View angle standard deviation	$\delta = 1.2°$
Medium example	
Image pixels	$n = 135^2$
Detector pixels	$p = 270$
Number of projection angles	$q = 270$
View angle standard deviation	$\delta = 0.4°$

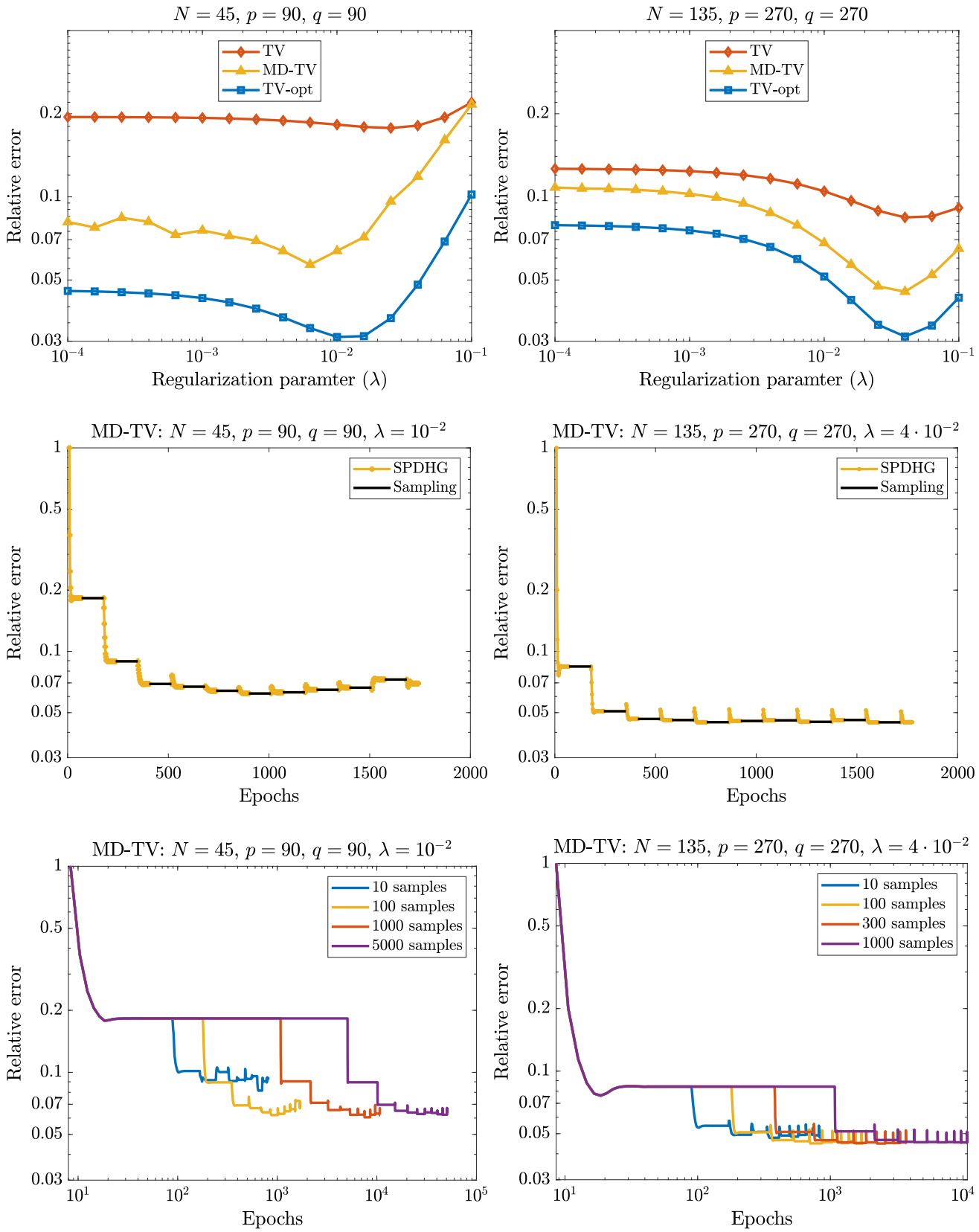


Fig. 2 We compare relative error (y-axis) with regularization parameter choice (top) for all 3 methods, show convergence history of our algorithm (middle) and compare number of samples of the model discrepancy for our method (bottom)

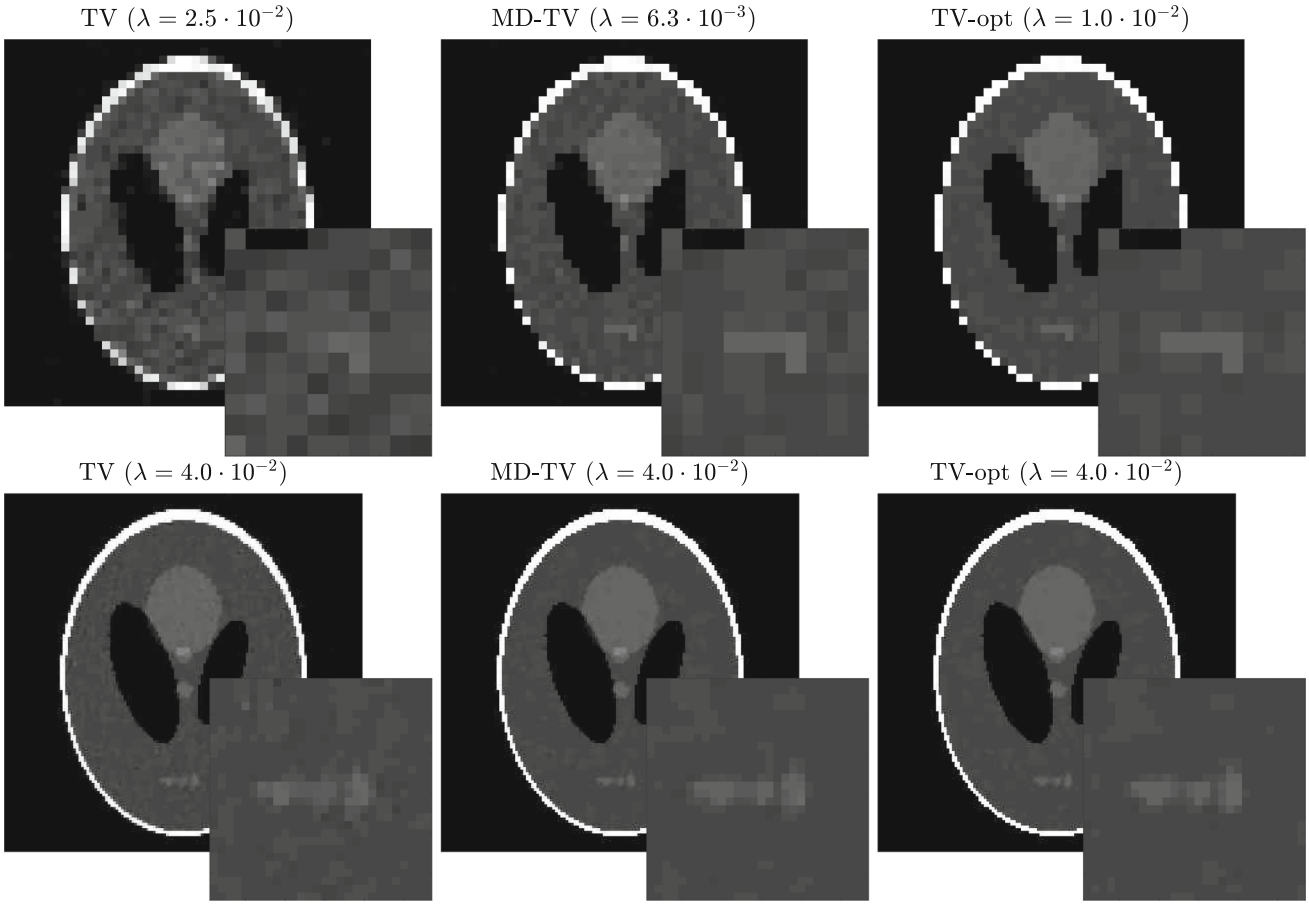


Fig. 3 We compare reconstruction quality of the solutions for all 3 methods with the optimal regularization parameter choice. All images are shown with the same greyscale. Top row $N = 45$, bottom row $N = 135$. We see that our method (MD-TV) improves reconstruction quality visually compared to a standard TV reconstruction, and is visually similar to the TV reconstruction using the true view angles

$$\mathbf{x}_{\text{TV}} = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2}\sum_{i=1}^{q} \|\mathbf{L}_{\mathbf{e},i}(\mathbf{b}_i - \mathbf{R}(\hat{\theta}_i)\,\mathbf{x} - \boldsymbol{\mu}_{\mathbf{e},i})\|_2^2 + \lambda\,\text{TV}(\mathbf{x}), \tag{32}$$

$$\mathbf{x}_{\text{MD-TV}} = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2}\sum_{i=1}^{q} \|\widetilde{\mathbf{L}}_{\boldsymbol{v}|\mathbf{x},i}(\mathbf{b}_i - \mathbf{R}(\hat{\theta}_i)\,\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{v}|\mathbf{x},i})\|_2^2 + \lambda\,\text{TV}(\mathbf{x}), \tag{33}$$

$$\mathbf{x}_{\text{TV-opt}} = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2}\sum_{i=1}^{q} \|\mathbf{L}_{\mathbf{e},i}(\mathbf{b}_i - \mathbf{R}(\bar{\theta}_i)\,\mathbf{x} - \boldsymbol{\mu}_{\mathbf{e},i})\|_2^2 + \lambda\,\text{TV}(\mathbf{x}), \tag{34}$$

where $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{\text{equid}}$ and MD-TV is our method.

In all cases, when solving for $\mathbf{x}$ we have a choice between using SPDHG (Algorithm 2) and PDHG. We use $K = 10$ outer iterations and $S = 100$ samples unless stated otherwise. According to our numerical tests, after $K = 10$ outer iterations the reconstruction error levels off, see Fig. 2 (middle plots). The stopping criterion in Algorithm 2 when using SPDHG is

$$\frac{\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2}{\|\mathbf{x}^k\|_2} \leq \frac{10^{-6}}{q}, \tag{35}$$

and when using PDHG we use the stopping criterion

$$\frac{\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2}{\|\mathbf{x}^k\|_2} \leq 10^{-6}. \tag{36}$$

We found empirically that dividing by q in (35) for SPDHG provides similar reconstructions as using (36) for PDHG.

In Table 1, we summarize the parameters for our simulated experiments. We consider the same two examples from [25]. We compare the quality of a solution $\mathbf{x}$ using the relative error $\|\mathbf{x} - \bar{\mathbf{x}}\|_2/\|\bar{\mathbf{x}}\|_2$ and visual quality.

4.1 Reconstructions

In the top row of Fig. 2 we show the relative error versus the regularization parameter λ for the two examples and

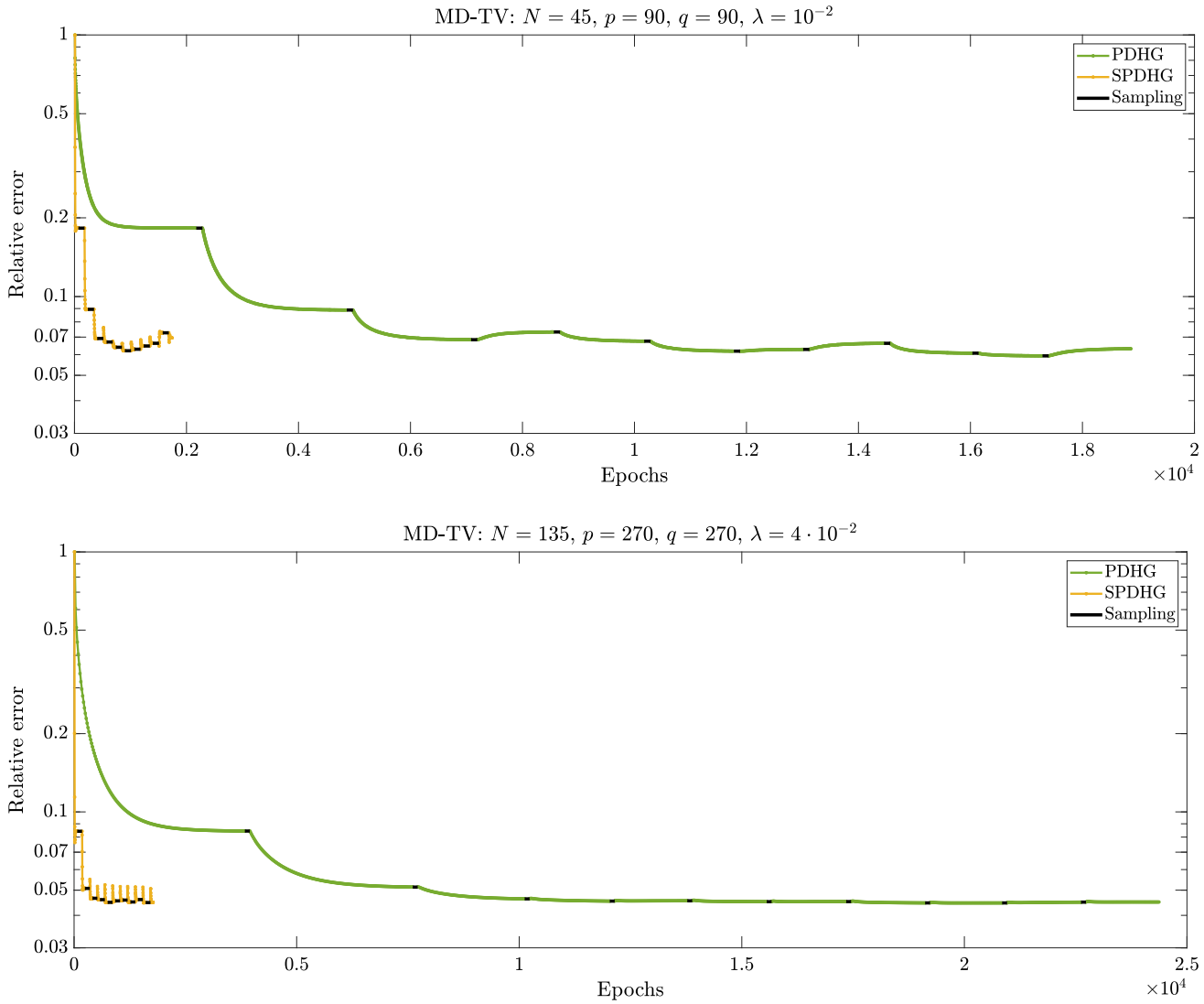


Fig. 4 We compare the convergence history of Algorithm 1 using either PDHG or SPDHG on our block-structure problem (19). Every "corner" in these plots indicates the beginning of a new outer iteration. We see that SPDHG greatly decreases the number of epochs required for this problem

all three methods. First note that for the ranges shown our method (MD-TV) always has a lower relative error compared to the TV reconstruction (TV) from (32). Furthermore, we see that the optimal choice of regularization parameter for MD-TV is similar to the one from the TV reconstruction using the true view angles (TV-opt) from (34). This indicates that the uncertainty due to the view angles is better handled by incorporating it in the data fidelity, rather than increasing the amount of TV regularization.

In the middle row of Fig. 2 we show the convergence history of MD-TV for the $K = 10$ outer iterations. The SPDHG algorithm uses one multiplication with $\mathbf{R}(\hat{\boldsymbol{\theta}})$ and its transpose every $2q$ iterations in expectation. Hence, to show convergence and compare the amount of work we use the unit of an *epoch* defined as the work involved in one multiplication with $\mathbf{R}(\hat{\boldsymbol{\theta}})$ or its transpose, since the computation is dominated by the forward and back projections.

The SPDHG algorithm computes q times as many gradients (multiplication with ∇) as the PDHG algorithm. This leads to a small increase in computational cost that is not accounted for when only comparing epochs. However, evaluation of each finite difference approximation $[\nabla \mathbf{x}]_i$ in (14) requires summing only the neighboring pixels in the CT reconstruction, compared to summing all pixels along each line for the forward and back projections. Hence—except for very underdetermined problems—computation of the gradient is significantly less expensive than computing the forward projection. In our numerical tests using MATLAB, we found that a direct computation of the gradient is 1000 times faster than computing a forward projection with the optimized CPU-ASTRA library. The convergence plot includes the

work associated with sampling the model discrepancy which requires $S + 1$ multiplications of $\mathbf{R}(\boldsymbol{\theta})$.

In the bottom row of Fig. 2 we show the convergence history of MD-TV for different number of samples S. We note that relative error of the final solution does not improve when using more than about $S = 100$ samples, so we use that for our other experiments.

Finally, in Fig. 3 we show the reconstructions obtained from the regularization parameter that yielded the lowest relative error using all 3 methods.

4.2 Convergence

In Fig. 4, we compare the convergence history of our algorithm using SPDHG which takes the block structure into account with our previous algorithm using PDHG, which does not. Every "corner" in these plots indicates the beginning of a new outer iteration, and we see a significant decrease in the number of epochs required in each outer iteration. We see for example that in the small example the SPDHG algorithm reaches convergence in the first outer iteration using 69 epochs compared to PDHG with 2174 epochs.

5 Conclusion

We proposed a new model and iterative algorithm for CT reconstruction when the view angles are uncertain. The uncertainty is described by a model discrepancy term and included in the data fidelity term of a TV regularized variational model. To establish the new model, we have overcome two key difficulties: (i) the probability distribution for the combined measurement noise and model discrepancy may not have a closed-form expression leading to difficulties for deriving a data fidelity term and (ii) the combined noise and discrepancy depend on the unknown CT reconstruction. We handle these issues by approximating the model discrepancy by a Gaussian, leading to a closed-form expression for the data-fidelity, and we alternately update the reconstruction and the parameters of the model discrepancy. 2D numerical experiments show that this approach improves the reconstruction quality in terms of relative error and visual quality compared to a standard TV reconstruction. Furthermore, our model admits a block structure, which we take advantage of to greatly reduce the overall computational work.

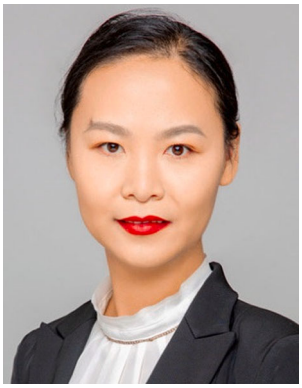
Acknowledgements We thank Aku Seppänen for a helpful discussion regarding the block formulation of our variational method. We also thank Matthias Ehrhardt for a discussion on the implementation details of SPDHG for this problem. Finally, we thank the referees for comments and suggestions that helped to improve the presentation.

References

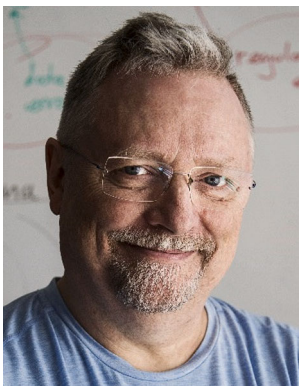
1. Arridge, S.R., Kaipio, J.P., Kolehmainen, V., Schweiger, M., Somersalo, E., Tarvainen, T., Vauhkonen, M.: Approximation errors and model reduction with an application in optical diffusion tomography. Inverse Probl. **22**(1), 175–195 (2006)
2. Basu, S., Bresler, Y.: Uniqueness of tomography with unknown view angles. IEEE Trans. Image Process. **9**(6), 1094–1106 (2000)
3. Benning, M., Burger, M.: Modern regularization methods for inverse problems. Acta Numer. **27**, 1–111 (2018)
4. Brandt, C., Seppänen, A.: Recovery from errors due to domain truncation in magnetic particle imaging: approximation error modeling approach. J. Math. Imaging Vis. **60**(8), 1196–1208 (2018)
5. Buzug, T.M.: Computed Tomography: From Photon Statistics to Modern Cone-Beam CT. Springer, Berlin (2008)
6. Calvetti, D., Dunlop, M., Somersalo, E., Stuart, A.: Iterative updating of model error for Bayesian inversion. Inverse Probl. **34**(2), 025008 (2018)
7. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**(1), 120–145 (2011)
8. Dongarra, J., Moler, C., Bunch, J., Stewart, G.: LINPACK Users' Guide. SIAM, Philadephia (1979)
9. Ehrhardt, M.J., Markiewicz, P.J., Schönlieb, C.B.: Faster PET reconstruction with non-smooth priors by randomization and preconditioning. Phys. Med. Biol. **64**(22), 225019 (2019)
10. Fang, Y., Murugappan, S., Ramani, K.: Estimating view parameters from random projections for tomography using spherical MDS. BMC Med. Imaging **10**(1), 12 (2010)
11. Hansen, T.M., Cordua, K.S., Holm Jacobsen, B., Mosegaard, K.: Accounting for imperfect forward modeling in geophysical inverse problems—exemplified for crosshole tomography. Geophysics **79**(3), H1–H21 (2014)
12. Kaipio, J., Somersalo, E.: Statistical and Computational Inverse Problems. Springer, New York (2005)
13. Kaipio, J., Somersalo, E.: Statistical inverse problems: discretization, model reduction and inverse crimes. J. Comput. Appl. Math. **198**(2), 493–504 (2007)
14. Kennedy, M.C., O'Hagan, A.: Bayesian calibration of computer models. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **63**(3), 425–464 (2001)
15. Kolehmainen, V., Tarvainen, T., Arridge, S.R., Kaipio, J.P.: Marginalization of uninteresting distributed parameters in inverse problems—application to diffuse optical tomography. Int. J. Uncertain. Quantif. **1**(1), 1–17 (2011)
16. Korolev, Y., Lellmann, J.: Image reconstruction with imperfect forward models and applications in deblurring. SIAM J. Imaging Sci. **11**(1), 197–218 (2018)
17. Natterer, F.: The Mathematics of Computerized Tomography. Wiley, Chicago (1986)
18. Niebler, S., Schömer, E., Tjaden, H., Schwanecke, U., Schulze, R.: Projection-based improvement of 3D reconstructions from motion-impaired dental cone beam CT data. Med. Phys. **46**(10), 4470–4480 (2019)
19. Nissinen, A., Heikkinen, L.M., Kaipio, J.P.: The Bayesian approximation error approach for electrical impedance tomography—experimental results. Meas. Sci. Technol. **19**(1), 015501 (2007)
20. Nissinen, A., Kolehmainen, V., Kaipio, J.P.: Reconstruction of domain boundary and conductivity in electrical impedance tomography using the approximation error approach. Int. J. Uncertain. Quantif. **1**(3), 203–222 (2011)
21. Madsen, R.B., Hansen, T.M.: Estimation and accounting for the modeling error in probabilistic linearized amplitude variation with offset inversion. Geophysics **83**(2), N15–N30 (2018)

22. Mallick, S.P., Agarwal, S., Kriegman, D.J., Belongie, S.J., Carragher, B., Potter, C.S.: Structure and view estimation for tomographic reconstruction: a Bayesian approach. In: Fitzgibbon, A., Taylor, C.J., Lecun, Y. (eds.) 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), **2**(1):2253–2260 (2006)
23. Peng, C.B., Rodi, W.L., Toksöz, M.N.: A Tikhonov regularization method for image reconstruction. In: Wei, Y., Gu, B. (eds.) Acoustical Imaging, vol. 20, pp. 153–164. Springer, Berlin (1993)
24. Pulkkinen, A., Kolehmainen, V., Kaipio, J.P., Cox, B.T., Arridge, S.R., Tarvainen, T.: Approximate marginalization of unknown scattering in quantitative photoacoustic tomography. Inverse Probl. Imaging **8**(3), 811–829 (2014)
25. Riis, N.A.B., Dong, Y.: A new iterative method for CT reconstruction with uncertain view angles. In: Lellmann, J., Burger, M., Modersitzki, J. (eds.) Scale Space and Variational Methods in Computer Vision. SSVM 2019. Lecture Notes in Computer Science, vol. 11603, pp. 156–167. Springer, Berlin (2019)
26. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1–4), 259–268 (1992)
27. Sidky, E.Y., Pan, X.: Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Phys. Med. Biol. **53**(17), 4777–4807 (2008)
28. Tick, J., Pulkkinen, A., Tarvainen, T.: Modelling of errors due to speed of sound variations in photoacoustic tomography using a Bayesian framework. Biomed. Phys. Eng. Express **6**(1), 015003 (2019)
29. Tikhonov, A.N.: Solution of incorrectly formulated problems and the regularization method. Soviet Math. Dokl. **4**:1035–1038 (1963); English translation of Dokl. Akad. Nauk. SSSR **151**:501–504 (1963)
30. van Aarle, W., Palenstijn, W.J., Cant, J., Janssens, E., Bleichrodt, F., Dabravolski, A., Sijbers, J.: Fast and flexible X-ray tomography using the ASTRA toolbox. Opt. Express **24**(22), 25129–25147 (2016)

Nicolai André Brogaard Riis is a Ph.D. researcher at the Technical University of Denmark. His research focuses on the mathematical modelling of inverse problems with uncertainty in the physical models.



Yiqiu Dong is associate professor at the Technical University of Denmark. She received the B.Sc. degree in mathematics from Yantai University, Yantai, China, in 2002 and the Ph.D. degree in mathematics from Peking University, Beijing, China, in 2007. Her research area includes mathematical modeling and algorithms in imaging sciences, inverse problems and variational methods, matrix application and computation, and optimization methods.



Per Christian Hansen is professor of scientific computing at the Technical University of Denmark. His research focuses on numerical analysis, iterative reconstruction methods, and computational methods for inverse problems. He is the author of four books and several MATLAB packages, and he is a SIAM Fellow.