A Convex Reconstruction Model for X-ray Tomographic Imaging with Uncertain Flat-fields

Hari Om Aggrawal, Martin S. Andersen, Sean Rose, and Emil Y. Sidky

Abstract—**Classical methods for X-ray computed tomography are based on the assumption that the X-ray source intensity is known, but in practice, the intensity is measured and hence uncertain. Under normal operating conditions, when the exposure time is sufficiently high, this kind of uncertainty typically has a negligible effect on the reconstruction quality. However, in time- or dose-limited applications such as dynamic CT, this uncertainty may cause severe and systematic artifacts known as ring artifacts. By carefully modeling the measurement process and by taking uncertainties into account, we derive a new convex model that leads to improved reconstructions despite poor quality measurements. We demonstrate the effectiveness of the methodology based on simulated and real data sets.**

Index Terms—**X-ray computed tomography, ring artifacts, low intensity, reconstruction methods.**

I. INTRODUCTION

X-RAY computed tomography (CT) is a non-invasive method that is used to image the internal structure of objects without cutting or breaking them. An X-ray source illuminates an object from different directions while detectors capture the attenuated X-rays. As the X-rays propagate through the object along straight lines, they are attenuated exponentially with a rate of decay that depends on the material. This relationship is explained by the Lambert–Beer law which forms the basis of major X-ray CT reconstruction models and methods; see *e.g.* [1]. Reconstruction methods estimate the spacial attenuation of the object of interest based on a number of X-ray images, given the measurement geometry, the source intensity, and possibly some assumptions on the statistical nature of the measurement process.

In practice, the source intensity is never known exactly, but it is estimated by acquiring a number of X-ray images without an object in the scanner. Such measurements are also known as air scans [2], flat-fields, or white-fields [3]. The elementwise mean of these measurements provides an estimate of the flat-field intensity and may be used for computing reconstructions. However, in practice the measurements are noisy, and hence the flat-field intensity estimate is a random variable whose variance is proportional to the ratio of the flat-field intensity and number of flat-field samples [4]. Consequently, the signal-to-noise ratio (SNR) of the flat-field intensity estimate is proportional to the square root of the product of the flat-field intensity and the number of samples. Therefore, if the flat-field

Hari Om Aggrawal and Martin S. Andersen are with the Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Lyngby, Denmark (e-mail: {haom, mskan}@dtu.dk).

Sean Rose and Emil Y. Sidky are with the Department of Radiology, University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637, USA (email: seanrose949@gmail.com, sidky@uchicago.edu).

intensity is low or if the number of flat-field measurements is small, the flat-field estimation error may be significant and lead to reconstruction artifacts and errors. Since the flat-field estimate is used to normalize measurements from all projection directions, the estimation errors result in systematic reconstruction errors. These are known as ring artifacts [5] since they appear as concentric circles superimposed on the reconstruction, and they are a common problem that can mask important features in the reconstructed image [6], [7]. Ring artifacts may not only occur because of flat-field estimation errors; miscalibrated or dead detector elements and non-uniform sensitivities may also systematically corrupt the measurements and lead to ring artifacts in the reconstruction [3].

An experimental study [8] has pointed out that the ring artifacts are more severe when the X-ray source intensity is low, and hence a reconstruction from low-intensity measurements may be very sensitive to the assumptions upon which the reconstruction method is based. The problem may arise when the acquisition time is limited, *e.g.*, in dynamic or time-resolved tomography, or if the application imposes strict dose limitations. Thus, tomographic reconstruction based on low-intensity measurements is a challenging problem, in part because of the low SNR.

One approach to combating ring artifacts is to move the detector array between projections [9]. This has an averaging effect on the systematic error due to flat-field estimation errors and often results in noticeable improvements, but it does not address or model the underlying cause. Moreover, it requires special hardware for the acquisition, and it is not suited for applications such as dynamic CT where fast acquisition times are important. Alternative software-based methods to mitigate ring artifacts also exist. Roughly speaking, these methods can be put into three categories: sinogram preprocessing methods [10]–[14], combined ring reduction and reconstruction methods [15], [16], and post-processing methods that reduce or remove rings from a reconstruction [3], [17], [18]. The preprocessing methods detect and remove/reduce stripes in the sinogram which, in turn, reduces the ring artifacts in the image domain. These algorithms are typically based on Fourier domain filtering [11], wavelet domain filtering [12], or a normalization of measurements by estimating the sensitivity of each detector pixel [14]. The post-processing methods transform the reconstructed image from Cartesian to polar coordinates [3] and remove stripes using, *e.g.*, a median filter [17], a wavelet filter, or a variational model for destriping [18].

In two recently proposed methods [15], [16], ring artifact correction is included as an intrinsic part of the reconstruction process. Motivated by the cause of ring artifacts, which

appear as stripes in the sinogram domain, the sinogram is split into the sum of the true sinogram and a component which represents the systematic stripe errors. Although the combined ring-reduction and reconstruction methods do take the systematic nature of the flat-field estimation errors in the sinogram domain into account, they do not explicitly model the source of the errors nor their statistical properties.

Existing methods for mitigating ring artifacts have been shown to work reasonably well when applied to measurement data with high or acceptable SNRs. However, we are not aware of any studies that investigate ring artifact correction for low SNR measurements and where the intensity of X-ray beam is assumed to be uncertain. To this end, we derive a new reconstruction model that is based on a rigorous statistical description of our model assumptions. Unlike existing correction methods that, roughly speaking, are based on the geometric nature of ring artifacts in either the sinogram or the reconstruction, our approach is based on a model of a fundamental cause of these artifacts. The resulting reconstruction method jointly estimates the flat-field and the attenuation image, and we show that the estimation problem can be solved efficiently by solving a convex optimization problem. We also derive a quadratic approximation model which is similar to an existing weighted least-squares reconstruction model.

Outline: Section II introduces our model assumptions and reviews some existing approaches to CT reconstruction based on low SNR measurements. We illustrate the sensitivity of these existing methods to flat-field intensity estimation errors. Section III proposes a new reconstruction model and discusses different parameter selection strategies. We describe our numerical implementation in Section IV, and we validate the proposed model based on simulated data as well as real tomographic measurements in Section V. Section VI concludes the paper.

Notation: The set $\mathbb{R}^n$ denotes the n-dimensional real space, $\mathbb{R}^n_+$ is the nonnegative orthant of $\mathbb{R}^n$, and $\mathbb{R}^{m \times n}$ is the set of $m \times n$ real-valued matrices. Upper case letters denote matrices, lower case letters denote vectors or scalars, and boldface letters denote random variables. Given a vector $x \in \mathbb{R}^n$, the matrix $\mathrm{diag}(x)$ is the $n \times n$ diagonal matrix with the elements of x on the diagonal. Similarly, given a set of r square matrices $S_1, \ldots, S_r$, the matrix $\mathrm{blkdiag}(S_1, \ldots, S_r)$ denotes the block-diagonal matrix with diagonal blocks $S_1, \ldots, S_r$. The vector e_i denotes the ith column of an identity matrix, and $\mathbf{1}$ denotes a vector of ones. Given a vector $x \in \mathbb{R}^n$, the notation $\log(x)$ and $\exp(x)$ is interpreted as elementwise logarithm and exponentiation. $A \otimes B$ denotes the Kronecker product of $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times q}$, $\|A\|_F$ denotes the Frobenius norm of A, and $|A| \in \mathbb{R}^{m \times n}$ is the element-wise absolute value of A. The vector $y = \mathrm{vec}(Y)$ denotes the vector obtained by stacking the columns of the matrix Y. Given a discrete random variable $\mathbf{y}$, the probability of $\mathbf{y} = y$ is $\mathrm{P}(\mathbf{y} = y)$, or using shorthand notation, $\mathrm{P}(y)$. Similarly, given a continuous random variable $\mathbf{z}$, $\mathrm{P}(z)$ is shorthand for the probability density associated with $\mathbf{z}$, evaluated at z, and finally, $\mathsf{E}[\mathbf{z}]$ denotes the expectation of $\mathbf{z}$.

II. CONVENTIONAL RECONSTRUCTION APPROACH

A. *System and Measurement Model*

The Lambert–Beer law describes how an X-ray beam is attenuated as it travels through an object that is characterized by a spatial attenuation function $\mu(x)$. Specifically, the incident intensity of an X-ray beam on a detector is given by

$$I \approx I_0 \exp\left(-\int_l \mu(x)\,dx\right) \tag{1}$$

where I_0 is the intensity of the X-ray source, and l denotes the line segment between the source and a detector. This description does not take the detector efficiency and the statistical nature of the photon arrival process into account. For photon-counting detectors, it is common to assume that the photon arrival process is a Poisson process, and each measurement is assumed to be a sample from a Poisson distribution whose mean is prescribed by the Lambert–Beer law. Here we will consider a two-dimensional geometry where p projections are acquired using a one-dimensional detector array with r detector elements. We will use the notation y_{ij} to denote the measurement obtained with detector element i and projection j, and we will assume that the ith detector element has efficiency $\eta_i \in (0, 1]$ such that the effective intensity is $v_i = \eta_i I_0$. Thus, with the assumption that the arrival process is Poisson process, y_{ij} is a realization of a random variable $\mathbf{y}_{ij}$ which, conditioned on μ and v_i, is a Poisson random variable whose mean is prescribed by the Lambert–Beer law, *i.e.*,

$$\mathbf{y}_{ij} \mid \mu, v_i \sim \mathrm{Poisson}\left(v_i \exp\left(-\int_{l_{ij}} \mu(x)\,dx\right)\right) \tag{2}$$

where l_{ij} notes the line segment between the ith detector element and the source for projection angle j. For ease of notation, we define a matrix random variable $\mathbf{Y}$ of size $r \times p$ with elements $\mathbf{y}_{ij}$, and similarly, the $r \times p$ matrix Y denotes a realization of $\mathbf{Y}$ and $y = \mathrm{vec}(Y)$.

The attenuation function $\mu(x)$ may be discretized by using a parameterization

$$\mu(x) = \sum_{k=1}^{n} u_k \mu_k(x) \tag{3}$$

where $\mu_k(x)$ is one of n basis functions (*e.g.*, a pixel or voxel basis), and $u \in \mathbb{R}^n$ is a vector of unknowns (*e.g.*, pixel or voxel values). With this parameterization, the line integrals in (2) can be expressed as

$$\int_{l_{ij}} \mu(x)\,dx = e_i^T A_j u$$

where the elements of the matrix $A_j \in \mathbb{R}^{r \times n}$ are given by

$$(A_j)_{ik} = \int_{l_{ij}} \mu_k(x)\,dx,$$

and hence the columns of $\mathbf{Y}$ satisfy

$$\mathsf{E}[\mathbf{y}_j | u, v] = \mathrm{diag}(v) \exp(-A_j u), \ j = 1, \ldots, p$$

where $v = (v_1, \ldots, v_r)$.

In practice, the vector v is unknown and must be measured. As mentioned in the introduction, the measurements of v are

often referred to as flat-field measurements and are simply measurements obtained without any object in the CT scanner. We will assume that s flat-field measurements are acquired for each detector element based on the flat-field measuring model

$$\mathbf{f}_{ij} \mid v_i \sim \mathrm{Poisson}\,(v_i) \tag{4}$$

for $i = 1, \ldots, r$ and $j = 1, \ldots, s$, and $\mathbf{F}$ denotes a $r \times s$ matrix random variable with elements $\mathbf{f}_{ij}$. As for the measurements Y, the matrix $F \in \mathbb{R}^{r \times s}$ denotes a realization of $\mathbf{F}$.

B. Maximum Likelihood Estimation

Given the flat-field measurements F, a maximum likelihood (ML) estimate of v is given by

$$\hat{v}_{\mathrm{f}} = \underset{v}{\mathrm{argmin}} \left\{ -\log \mathrm{P}(F \mid v) \right\} \tag{5}$$

$$= \underset{v}{\mathrm{argmin}} \left\{ s\mathbf{1}^T v - \mathbf{1}^T F^T \log(v) \right\} = \frac{1}{s} F \mathbf{1},$$

i.e., $\hat{v}_{\mathrm{f}}$ is simply the arithmetic average of the s flat-field measurements. This estimate can be used to compute an *approximate* ML estimate of the vector u which is given by

$$\hat{u}_{\mathrm{y}} = \underset{u}{\mathrm{argmin}} \left\{ -\log \mathrm{P}(Y \mid u, \hat{v}_{\mathrm{f}}) \right\} \tag{6}$$

$$= \underset{u}{\mathrm{argmin}} \left\{ (\mathbf{1} \otimes \hat{v}_{\mathrm{f}})^T \exp(-Au) + y^T Au \right\}$$

where $A \in \mathbb{R}^{rp \times n}$ is defined as $A = [A_1^T \ \cdots \ A_p^T]^T$. The estimation problem (6) is a convex optimization problem, and it is essentially an approximate ML estimation problem since with our model assumptions, the true likelihood $\mathrm{P}(Y \mid u, v)$ is a function of both u and v. We will return to this issue in the next section.

If y is positive, a quadratic approximation of (6) can be obtained by means of a second-order Taylor expansion of the likelihood function [19], and this yields the following weighted least-squares objective function

$$\frac{1}{2} \|\mathrm{diag}(y)^{1/2}(Au - b)\|_2^2 \tag{7}$$

where $b = \mathbf{1} \otimes \log(\hat{v}_{\mathrm{f}}) - \log(y)$. Notice that if A has full rank and $rp \le n$, both (6) and the quadratic approximation (7) reduce to the problem of solving the consistent system of equations $Au = b$, but the two problems are generally different when the system of equations $Au = b$ is inconsistent. The noise properties of reconstructions based on the weighted least-squares objective (7) have been studied in [20].

C. The Effect of Flat-field Estimation Errors

The flat-field estimate $\hat{v}_{\mathrm{f}}$ in (5) satisfies $\mathsf{E}[\hat{v}_{\mathrm{f}}] = v$, and hence it is an unbiased estimate. However, $\hat{v}_{\mathrm{f}}$ is itself a random variable with covariance $(1/s)\mathrm{diag}(v)$, and the flat-field estimation error may lead to artifacts in the reconstruction. To study how flat-field estimation errors influence the reconstruction, we now consider a simplified model based on Gaussian approximations. Specifically, we assume that $(\hat{v}_{\mathrm{f}})_i | v_i \sim \mathcal{N}(v_i, s^{-1} v_i)$ and $\mathbf{y}_{ij} | v_i, u \sim \mathcal{N}(v_i \exp(-e_i^T A_j u), v_i \exp(-e_i^T A_j u))$. With these assumptions, $\mathbf{b}_{ij} = \log((\hat{v}_{\mathrm{f}})_i) - \log(\mathbf{y}_{ij})$ can be approximated by

linearizing each of the log terms around the mean of their arguments, *i.e.*,

$$\mathbf{b}_{ij} \approx \log(v_i) + \frac{(\hat{v}_{\mathrm{f}})_i - v_i}{v_i} - \log(\mathsf{E}[\mathbf{y}_{ij}]) - \frac{\mathbf{y}_{ij} - \mathsf{E}[\mathbf{y}_{ij}]}{\mathsf{E}[\mathbf{y}_{ij}]}$$

$$= e_i^T A_j u + \mathbf{z}_i + \mathbf{w}_{ij}$$

for $i = 1, \ldots, r$ and $j = 1, \ldots, p$, and where

$$\mathbf{z}_i = ((\hat{v}_{\mathrm{f}})_i - v_i)/v_i, \ \mathbf{z}_i \sim \mathcal{N}(0, (s v_i)^{-1})$$

and

$$\mathbf{w}_{ij} = (\mathbf{y}_{ij} - \mathsf{E}[\mathbf{y}_{ij}])/\mathsf{E}[\mathbf{y}_{ij}], \ \mathbf{w}_{ij} \sim \mathcal{N}(0, v_i^{-1} \exp(e_i^T A_j u)).$$

The terms $\mathbf{z}_i$ arise because of the flat-field estimation errors, and the terms $\mathbf{w}_{ij}$ represent the effect of measurement noise. If we define $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_r)$ and $\mathbf{w} = \mathrm{vec}(\mathbf{W})$ where $\mathbf{W}$ is the $r \times p$ matrix with elements $\mathbf{w}_{ij}$, then

$$\mathbf{b} \approx Au + \mathbf{1} \otimes \mathbf{z} + \mathbf{w}. \tag{8}$$

Not surprisingly, this shows that flat-field estimation errors affect all projections, and hence give rise to structured errors.

The linear approximation reaffirms that the variance of the flat-field errors is inversely proportional to the flat-field intensity and the number of flat-field measurements s. Thus, if s is sufficiently large, the flat-field estimation errors play a negligible role. However, a twofold reduction of the flat-field error-to-noise ratio

$$\sqrt{\frac{\mathsf{E}[\mathbf{z}_i^2]}{\mathsf{E}[\mathbf{w}_{ij}^2]}} = \frac{1}{\sqrt{s \exp(-e_i^T A_j u)}}$$

requires a fourfold increase in the number of flat-field samples, and hence it may require many samples to obtain a sufficiently small flat-field error-to-noise ratio.

We now demonstrate the effect of flat-field estimation errors by considering the behavior of reconstructions based on (6). We will use a constant flat-field $v = \omega \mathbf{1}$ for $\omega > 0$ to generate a set of measurements according to the model (2) with $r = 200$ detector elements and $p = 720$ parallel beam projections covering a full rotation. For the reconstruction we use the flat-field ML estimate $\hat{v}_{\mathrm{f}}$, as defined in (5), where only one flat-field sample ($s = 1$) is acquired for each detector element based on (4).

Our object u, shown in Fig. 1a, consists of three squares of different sizes where the attenuation of the innermost square is 0.5 cm^{-1}, the enclosing square has attenuation 0.25 cm^{-1}, and the outermost square has no attenuation. The domain size is 1 cm, and the reconstruction grid is 128×128 pixels. Fig. 1 shows three reconstructions based on (6) with different values of the parameter ω. The effect of the flat-field error appears as a ring in the reconstructions, and it is clear that the severity of both noise and the ring in the reconstruction decreases as the flat-field intensity is increased. In the next section, we propose and investigate a new reconstruction model that takes a statistical model of the flat-field into account.

The effect of a flat-field estimation error on the reconstruction may also be analyzed by means of an analytic reconstruction of the sinogram $h_\theta(t) = \delta(t - t_0)$ where $t_0 \ne 0$ is a given constant. This corresponds to a "line" in

(a) Phantom

(b) $\omega = 10^3$ photons

(c) $\omega = 10^4$ photons
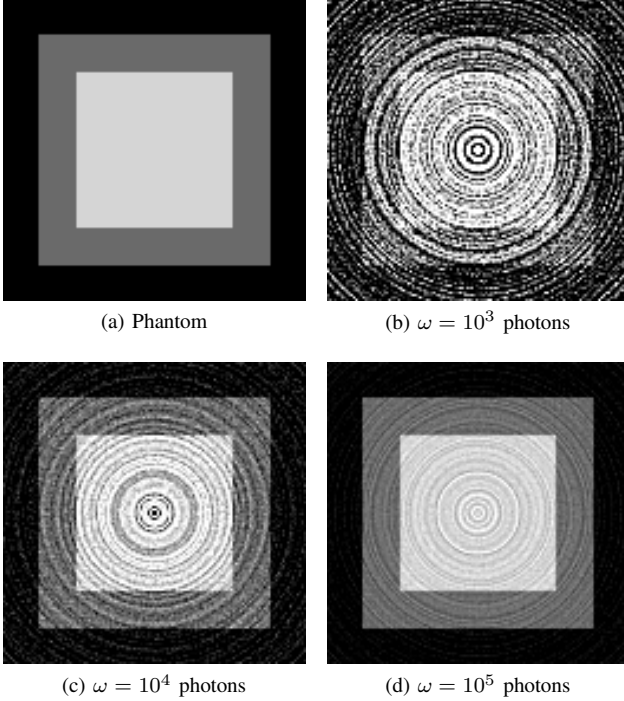
(d) $\omega = 10^5$ photons

Fig. 1. Phantom (a) and reconstructions (b), (c), and (d), based on (6) with flat-field estimation errors. The display range for each of the images is $[0, 0.6]$.
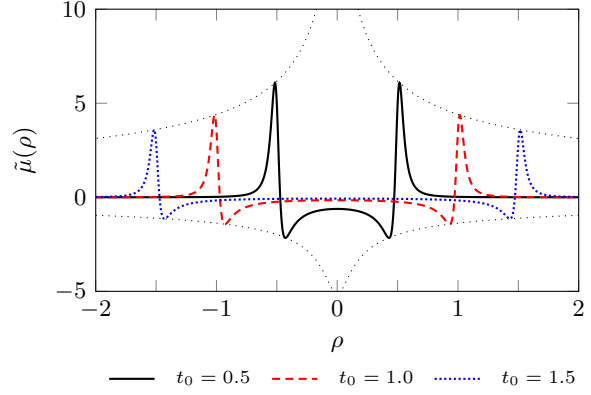

Fig. 2. Examples of radial profile of reconstruction of $h_\theta(t)$ for three different values of t_0 (0.5, 1.0, and 1.5) and $\epsilon = 0.05$.

the sinogram. The function $h_\theta(t)$ is a radial function (*i.e.*, it does not depend on θ), but it is not the Radon transform of a function since $h_\theta(t) \neq h_{\theta+\pi}(-t)$. As a consequence, the Fourier slice theorem does not hold. However, we may still compute a reconstruction using filtered backprojection. The reconstruction $\mu(x)$ is itself a radial function, and if we let $x = \rho n_\phi$ where $n_\phi = (\cos\phi, \sin\phi)$ such that $|\rho|$ is the distance to the origin, we obtain the expression [21]

$$\mu(\rho n_\phi) = \frac{1}{2} \int_{-\pi}^{\pi} \int_{-\infty}^{\infty} H_\theta(\zeta)|\zeta|e^{-2\pi\epsilon|\zeta|}e^{i2\pi\zeta\rho n_\phi^T n_\theta}\, d\zeta\, d\theta$$

$$= \pi \int_{-\infty}^{\infty} H_\theta(\zeta)|\zeta|e^{-2\pi\epsilon|\zeta|}J_0(2\pi\zeta\rho)\, d\zeta$$

$$= \pi \int_{0}^{\infty} [H_\theta(\zeta) + H_\theta(-\zeta)]\, e^{-2\pi\epsilon\zeta}\zeta J_0(2\pi\zeta\rho)\, d\zeta$$

where J_0 denotes the zeroth-order Bessel function of the first kind, $H_\theta(\zeta) = e^{-i2\pi\zeta t_0}$ is the Fourier transform of $h_\theta(t)$, and $|\zeta|e^{-2\pi\epsilon|\zeta|}$ is an apodizing filter with parameter $\epsilon > 0$. Using the Hankel transform pair (20) in [22, p. 9], we obtain the closed-form expression

$$\tilde{\mu}(\rho) = \frac{1}{4\pi} \left(\frac{\sigma}{(\sigma^2 + \rho^2)^{3/2}} + \frac{\bar{\sigma}}{(\bar{\sigma}^2 + \rho^2)^{3/2}} \right) \qquad (9)$$

where $\sigma = \epsilon + it_0$ and $\tilde{\mu}(\rho) = \mu(\rho n_\phi)$. Fig. 2 shows three examples of what this function may look like. It is clear from the figure that a systematic error in the sinogram in the form of a "line" will appear as spikes in the radial reconstruction. In particular, the reconstruction will have two "rings" of opposite sign near $\rho = t_0$, corresponding to the positive and negative peaks in the profile $\tilde{\mu}(\rho)$. The extrema of $\tilde{\mu}(\rho)$ (*i.e.*, the spike magnitudes) depend on both t_0 and ϵ. The dotted curves in the

figure provide an envelope of the extrema for $\epsilon = 0.05$, and it shows that the magnitude of a spike is large when $|t_0|$ is small and vice versa. Our analysis of the extrema of $\tilde{\mu}(\rho)$, which is included in Appendix A, shows that they are approximately inversely proportional to $\sqrt{\epsilon^3 |t_0|}$ when $|t_0| \gg \epsilon$. Moreover, $\tilde{\mu}(\rho)$ may have a significant offset near $\rho = 0$, as is the case for the example with $t_0 = 0.5$ in Fig. 2.

D. Including Prior Information

If the prior probability density $\mathrm{P}(u)$ is assumed to be known, a so-called maximum a posteriori (MAP) estimate can be expressed as

$$\hat{u}_{\mathrm{map}} = \underset{u}{\mathrm{argmin}}\left\{ -\log\mathrm{P}(u \mid y, v) \right\} \qquad (10)$$

where, according to Bayes' rule, the posterior probability density $\mathrm{P}(u|y,v)$ satisfies

$$\mathrm{P}(u \mid y, v) \propto \mathrm{P}(y \mid u, v)\mathrm{P}(u). \qquad (11)$$

Again, since v is generally unknown, an approximate MAP (AMAP) estimate can be obtained by maximizing an approximation of the posterior distribution, *i.e.*,

$$\hat{u}_{\mathrm{amap}} = \underset{u}{\mathrm{argmin}}\left\{ -\log\mathrm{P}(u \mid y, \hat{v}_{\mathrm{f}}) \right\}. \qquad (12)$$

We will restrict our attention to priors of the form

$$\mathrm{P}(u \mid \gamma) \propto e^{-\gamma\phi(u)} \qquad (13)$$

where $\phi(u)$ is a convex function and $\gamma > 0$ is a hyperparameter. With this prior, the AMAP estimation problem can be expressed as

$$\hat{u}_{\mathrm{amap}} = \underset{u}{\mathrm{argmin}}\left\{ (\mathbf{1} \otimes \hat{v}_{\mathrm{f}})^T \exp(-Au) + y^T Au + \gamma\phi(u) \right\} \qquad (14)$$

which is a convex optimization problem. Alternatively, using the quadratic approximation (7) in place of the log-likelihood function, we obtain the regularized weighted least-squares problem

$$\hat{u}_{\mathrm{wls}} = \underset{u}{\mathrm{argmin}}\left\{ \frac{1}{2}\|\mathrm{diag}(y)^{1/2}(Au - b)\|_2^2 + \gamma\phi(u) \right\} \qquad (15)$$

as an approximation to the AMAP estimation problem.

III. Joint Reconstruction Approach

We now turn to the main contribution of this paper, namely a model for jointly estimating the flat-field v as well as the absorption image u. Recall from the example in section II-C that the approximate ML model (6) may lead to ring artifacts. As will be evident from our numerical experiments in section V, the approximate MAP model (14) suffers the same drawback. To mitigate this, we consider joint MAP estimation of u and v. This approach is motivated by the fact that the measurements Y contain information about both u and v. Indeed, given u, an ML estimate of v can be computed as

$$\hat{v}_{\text{y}}(u) = \underset{v}{\arg\min} \left\{ -\log \text{P}(Y \mid u, v) \right\} \qquad (16)$$

$$= \text{diag} \left(\sum_{j=1}^{p} \exp(-A_j u) \right)^{-1} Y\mathbf{1}. \qquad (17)$$

A. MAP Estimation Problem

With the model assumptions described in II-A and given a flat-field prior $\text{P}(v|\alpha, \beta)$, the joint posterior distribution of the unknown parameters u and v can be expressed as

$$\text{P}(u, v \mid Y, F) \propto \text{P}(Y, F \mid u, v)\text{P}(u \mid \gamma)\text{P}(v \mid \alpha, \beta)$$

where $\text{P}(Y, F|u, v) = \text{P}(Y|u, v)\text{P}(F|v)$, and $\alpha \in \mathbb{R}^r$ and $\beta \in \mathbb{R}^r$ are hyperparameters associated with the flat-field prior. Here we will assume that v_i and v_j, $i \neq j$ are independent, and the flat-field prior is $v_i|\alpha_i, \beta_i \sim \text{Gamma}(\alpha_i, \beta_i)$ for $i = 1, \ldots, r$, i.e.,

$$\text{P}(v_i \mid \alpha_i, \beta_i) = \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} v_i^{\alpha_i - 1} \exp(-\beta_i v_i).$$

The Gamma prior is chosen because of computational convenience; it is the so-called conjugate prior for the Poisson likelihood function, and as a consequence, the posterior distribution of v given u is itself a Gamma distribution. For the Gamma distribution, the hyperparameter α_i is commonly referred to as the shape, and β_i is referred to as the rate. The corresponding MAP estimation problem can be expressed as

$$(\hat{u}, \hat{v}) = \underset{(u,v)}{\arg\min} \left\{ -\log \text{P}(u, v \mid Y, F) \right\} \qquad (18)$$

$$= \underset{(u,v)}{\arg\min} \left\{ J(u, v) + \gamma\phi(u) \right\}$$

where

$$J(u, v) = v^T d(u) + y^T Au - c^T \log(v) \qquad (19)$$

and

$$c = F\mathbf{1} + Y\mathbf{1} + \alpha - \mathbf{1}, \ d(u) = s\mathbf{1} + \sum_{j=1}^{p} \exp(-A_j u) + \beta. \qquad (20)$$

The function $J(u, v)$ is convex in u given v and vice versa, but it is not jointly convex in u and v. However, by setting the gradient of $J(u, v)$ with respect to v equal to zero, we obtain the first-order optimality condition $\hat{v}(u) = \text{diag}(d(u))^{-1}c$.

This allows us to eliminate v from the estimation problem (18), i.e.,

$$J(u, \hat{v}(u)) \propto y^T Au + c^T \log(d(u)),$$

which is a convex function of u. Thus, the problem (18) is equivalent to the following convex reconstruction model

$$\hat{u} = \underset{u}{\arg\min} \left\{ y^T Au + c^T \log(d(u)) + \gamma\phi(u) \right\} \qquad (21)$$

with the flat-field estimate $\hat{v}$ given by

$$\hat{v} = \text{diag}(d(\hat{u}))^{-1}c. \qquad (22)$$

We note that $\hat{v}$ has an interesting interpretation: each element of $\hat{v}$ can be expressed as a convex combination of three independent estimates, i.e.,

$$\hat{v} = \text{diag}(\theta_1)\hat{v}_{\text{f}} + \text{diag}(\theta_2)\hat{v}_{\text{y}}(\hat{u}) + \text{diag}(\theta_3)\hat{v}_{\text{pr}}(\alpha, \beta) \qquad (23)$$

where $\theta_1, \theta_2, \theta_3 \in \mathbb{R}_+^r$, $\theta_1 + \theta_2 + \theta_3 = \mathbf{1}$, are parameters that depend on both data and $\hat{u}$, α, and β. The ML estimate $\hat{v}_{\text{f}}$, defined in (5), is based on the flat-field measurements F, the estimate $\hat{v}_{\text{y}}(\hat{u})$ is based on the measurements Y and defined in (16), and the estimate $\hat{v}_{\text{pr}}(\alpha, \beta) = \text{diag}(\beta)^{-1}(\alpha - \mathbf{1})$ is based on the flat-field prior; see Appendix B for further details on this interpretation.

B. Choosing The Hyperparameters

The estimation problem (21) depends on the flat-field hyperparameters α and β. We now discuss different ways to choose these hyperparameters.

1) Uniform Positive Prior: The simplest prior is perhaps the uniform positive (UP) prior which is obtained by setting $\alpha_i = 1$ and $\beta_i = 0$ for $i = 1, \ldots, r$. In the present case, this corresponds to simply omitting the prior $\text{P}(v|\alpha, \beta)$ from the model, and hence the flat-field estimates $\hat{v}(u)$ become convex combinations of only two estimates instead of three. This is an improper prior since it does not integrate to one.

2) Jeffreys Prior: The Jeffreys prior (JP) for the Poisson distribution is $p(v_i|\alpha_i, \beta_i) \propto 1/\sqrt{v_i}$ which is obtained by letting $\alpha_i = 0.5$ and $\beta_i = 0$. This is also an improper prior.

3) Type-II ML Estimation: The flat-field measurements can be used to estimate the hyperparameters by maximizing the marginal probability of $f_{i1}, \ldots, f_{is}$ given the hyperparameters α_i and β_i, i.e.,

$$(\hat{\alpha}_i, \hat{\beta}_i) = \underset{(\alpha_i, \beta_i)}{\arg\min} \left\{ -\log \text{P}(f_{i1}, \ldots, f_{is} \mid \alpha_i, \beta_i) \right\}. \qquad (24)$$

This is known as type-II ML estimation or empirical Bayes estimation [23]. As shown in Appendix C, this approach leads to the AMAP model, i.e., a zero-variance prior with mean $\hat{v}_{\text{f}}$.

4) Flat-field Emphasizing Prior: Recall that the flat-field estimate $\hat{v}(u)$ can be expressed as convex combinations of three estimates. Specifically,

$$\hat{v}_i(u) = \frac{s}{d_i(u)}(\hat{v}_{\text{f}})_i + \frac{\tau_i(u)}{d_i(u)}(\hat{v}_{\text{y}})_i + \frac{\beta_i}{d_i(u)}\frac{\alpha_i - 1}{\beta_i} \qquad (25)$$

where $\tau_i(u) = \sum_{j=1}^{p} \exp(-e_i^T A_j u)$. If we set the mode of the Gamma prior (i.e., $(\alpha_i - 1)/\beta_i$) equal to the flat-field
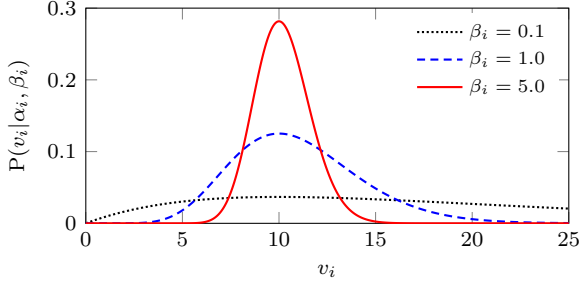
Fig. 3. Gamma distributions with hyperparameters β_i and $\alpha_i = 1 + (\hat{v}_{\mathrm{f}})_i\beta_i$ for $(\hat{v}_{\mathrm{f}})_i = 10$ and $\beta_i \in \{0.1, 1.0, 5.0\}$.

ML estimate $(\hat{v}_{\mathrm{f}})_i$ by letting $\alpha_i = 1 + \beta_i(\hat{v}_{\mathrm{f}})_i$, we obtain the estimate

$$\hat{v}_i(u) = \frac{s + \beta_i}{d_i(u)}(\hat{v}_{\mathrm{f}})_i + \frac{\tau_i(u)}{d_i(u)}(\hat{v}_{\mathrm{y}})_i \qquad (26)$$

which is a convex combination of two estimates. It is easy to verify that $\hat{v}_i(u) \to (\hat{v}_{\mathrm{f}})_i$ as $\beta \to \infty$, and with $\beta_i = 0$, the estimate $\hat{v}_i(u)$ is equivalent to the estimate obtained with the UP prior. Thus, choosing $\beta_i > 0$ and $\alpha_i = 1 + \beta_i(\hat{v}_{\mathrm{f}})_i$ allows us to emphasize the flat-field ML estimate $(\hat{v}_{\mathrm{f}})_i$. This is consistent with the fact that the parameter β_i is the rate parameter associated with the Gamma distribution: the larger the rate, the more concentrated the distibution is around its mode. This is illustrated in Fig. 3. We call this corresponding prior the flat-field emphasizing (FE) prior.

C. Quadratic Approximation

A quadratic approximation of the first two terms in (21) can be derived by means of a second-order Taylor expansion with respect to Au. Substituting y for $(I \otimes \mathrm{diag}(\hat{v}(u))) \exp(-Au)$, we obtain the following approximate MAP estimation problem

$$\hat{u}_{\mathrm{swls}} = \underset{u}{\mathrm{argmin}}\left\{\frac{1}{2}\|Au - b\|_{\widehat{\Sigma}_{\mathbf{b}}^{-1}}^2 + \gamma\phi(u)\right\} \qquad (27)$$

where the covariance matrix $\widehat{\Sigma}_{\mathbf{b}}$ is defined as

$$\widehat{\Sigma}_{\mathbf{b}} = (\mathbf{1}\mathbf{1}^T) \otimes \mathrm{diag}(s\hat{v}_{\mathrm{f}} + \alpha - 1)^{-1} + \mathrm{diag}(y)^{-1}. \qquad (28)$$

This is also the covariance matrix associated with $\mathbf{b}$ in the linear approximation (8). Note that the weighted least-squares data fidelity term takes the systematic errors induced by flat-field estimation errors into account without explicitly modeling the flat-field, and hence we label this a regularized stripe-weighted least-squares (SWLS) problem. The model depends on the hyperparameter vector α, which appears in the covariance matrix, but β does not appear in the model.

IV. IMPLEMENTATION

The MAP estimation problems (14) and (21) as well as the WLS (15) and SWLS (27) quadratic approximations are all convex problems that can be solved with a wide range of numerical optimization methods. Here we will focus on simple first-order methods which are suitable for large-scale problems.

A. Attenuation Priors

Before we describe our implementation of the different reconstruction methods, we briefly discuss two attenuation priors of the form (13), namely the nonnegativity prior (corresponding to nonnegativity constraints $u_i \geq 0$), and a combination of the nonnegativity prior and total variation (TV) regularization [24]. Both of these priors can be combined with the existing AMAP model (14), the proposed model (21), the WLS model (15) and the SWLS model (27).

1) Nonnegativity: The nonnegativity constraints can be expressed as $\phi(u) = I_+(u)$ where $I_+(u)$ denotes the indicator function of the nonnegative orthant, *i.e.*, $I_+(u) = 0$ if and only if u is a nonnegative vector, and otherwise $I_+(u) = \infty$.

2) Nonnegativity and TV: The combination of nonnegativity constraints and TV may be expressed as

$$\phi(u) = I_+(u) + \mathrm{TV}_\delta(u)$$

where $\mathrm{TV}_\delta(u) = \sum_{i=1}^n \xi_\delta(\|D_i u\|_2)$ is a differentiable TV-approximation, ξ_δ denotes the Huber-norm

$$\xi_\delta(t) = \begin{cases} (t)^2/(2\delta) & |t| \leq \delta \\ |t| - \delta/2 & \text{otherwise} \end{cases}$$

with parameter δ, and $D_i u$ is a finite-difference approximation of the gradient at pixel i. We will use a pixel basis corresponding to an $M \times N$ grid (*i.e.*, $n = MN$). Specifically, we define

$$D_i = \begin{bmatrix} e_i^T(I_N \otimes \bar{D}_M) \\ e_i^T(\bar{D}_N \otimes I_M) \end{bmatrix}$$

where I_M and I_N are identity matrices, and $\bar{D}_M$ and $\bar{D}_N$ are square difference matrices of order M and N, respectively, and of the form

$$\begin{bmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \\ & & & 0 \end{bmatrix}$$

where the last row is zero, corresponding to Neumann boundary conditions.

The function $\mathrm{TV}_\delta(u)$ has a Lipschitz continuous gradient with constant $L_{\mathrm{tv}}(\delta) = \|D\|_2^2/\delta$ where $D = \begin{bmatrix} D_1^T & \cdots D_n^T \end{bmatrix}^T$.

B. Reconstruction Models

We now consider five different reconstruction models of the form

$$\text{minimize} \quad J_i(u) + \gamma\phi(u), \quad i = 1, \ldots, 5, \qquad (29)$$

where $J_i(u)$ is based on either (10), (14), (21), (15) or (27).

1) Baseline and AMAP Estimation: The reconstruction model (10) requires the true flat-field v which is not available in practice. However, the model may be used to compute a baseline reconstruction in simulation studies. The baseline reconstruction problem corresponds to $J_1(u) = J(u, v)$ where the true flat-field v is assumed to be known. If we replace v by $\hat{v}_{\mathrm{f}}$, we obtain the AMAP model (14) with objective $J_2(u) = J(u, \hat{v}_{\mathrm{f}})$.

To solve the reconstruction problem (29) using a first-order method, we need the gradient of $J(u,v)$ with respect to u, *i.e.*,

$$\nabla_u J(u,v) = A^T(y - \hat{y}(u,v)) \tag{30}$$

where $\hat{y}(u,v) = (I \otimes \mathrm{diag}(v))\exp(-Au)$. It is easy to verify that the gradient $\nabla_u J(u,v)$ is Lipschitz continuous on the nonnegative orthant since the norm of the Hessian

$$\nabla_u^2 J(u,v) = A^T \mathrm{diag}(\hat{y}(u,v)) A$$

is bounded for $u \geq 0$ and with v fixed. We will use the Lipschitz constants $L_1 = \max_i\{v_i\}\|A\|_2^2$ and $L_2 = \max_i\{(\hat{v}_\mathrm{f})_i\}\|A\|_2^2$.

2) Joint MAP Estimation: The MAP estimation problem (21) is a special case of (29) if we let $J_3(u) = J(u, \hat{v}(u))$. The gradient of $J_3(u)$ is

$$\nabla J_3(u) = A^T y + D_d(u)^T \hat{v}(u) \tag{31}$$
$$= A^T(y - \hat{y}(u, \hat{v}(u)))$$

where $D_d(u) = -\sum_{j=1}^p \mathrm{diag}(\exp(-A_j u))A_j$ denotes the Jacobian matrix of $d(u)$. Comparing with (30), we see that the only difference is that the residual $y - \hat{y}(u, \hat{v}(u))$ is based on the flat-field estimate $\hat{v}(u)$ instead of the true flat-field v or the ML estimate $\hat{v}_\mathrm{f}$.

To derive the Hessian of $J_3(u)$, note that

$$c^T \log(d(u)) = \sum_{i=1}^r c_i \log(d_i(u))$$

where $d_i(u) = s + \sum_{j=1}^p \exp(-e_i^T A_j u) + \beta_i$. This implies that the Hessian can be expressed as

$$\sum_{i=1}^r c_i\left(\frac{\nabla^2 d_i(u)}{d_i(u)} - \frac{\nabla d_i(u)\nabla d_i(u)^T}{d_i(u)^2}\right).$$

Now let $\Pi_i = I \otimes e_i^T$ such that $\Pi_i y = Y^T e_i$ corresponds to the ith row of Y, and define a permutation matrix $\Pi = [\Pi_1^T \cdots \Pi_r^T]^T$. This allows us to express the Hessian $\nabla^2 J_3(u)$ as

$$\nabla^2 J_3(u) = A^T \Pi^T \mathrm{blkdiag}(B_1(u), \ldots, B_r(u))\Pi A \tag{32}$$

where $B_i(u) = \mathrm{diag}(\Pi_i \hat{y}) - \frac{1}{c_i}\Pi_i \hat{y}\hat{y}^T \Pi_i^T$, and where $\hat{y}$ is used as shorthand for $\hat{y}(u, \hat{v}(u))$. (We remark that $\hat{v}(u)$ depends on both α and β, and consequently, so does the Hessian $\nabla^2 J_3(u)$.) It follows that

$$\|\nabla^2 J_3(u)\|_2 \leq \|A^T \mathrm{diag}(y)A\|_2$$

which implies that $\nabla J_3(u)$ is Lipschitz continuous with constant $L_3 = \|A^T \mathrm{diag}(y)A\|_2$.

3) WLS Estimation: The quadratic approximation (15) corresponds to (29) with $J_4(u) = \frac{1}{2}\|Au - b\|_{\widehat{\Sigma}_\mathbf{b}^{-1}}^2$ and $\widehat{\Sigma}_\mathbf{b} = \mathrm{diag}(y)^{-1}$. The gradient of $J_4(u)$ is

$$\nabla J_4(u) = A^T \Sigma_\mathbf{b}^{-1}(Au - b)$$

which is Lipschitz continuous with constant $\|A^T \widehat{\Sigma}_\mathbf{b}^{-1} A\|_2$.

4) Regularized SWLS: The quadratic approximation (27) corresponds to (29) with $J_5(u) = \frac{1}{2}\|Au - b\|_{\widehat{\Sigma}_\mathbf{b}^{-1}}^2$ and

$$\widehat{\Sigma}_\mathbf{b} = \Pi^T\left[\mathrm{diag}(\Pi y)^{-1} + \mathrm{diag}(s\hat{v}_\mathrm{f} + \alpha - \mathbf{1})^{-1} \otimes (\mathbf{1}\mathbf{1}^T)\right]\Pi.$$

Thus, $\widehat{\Sigma}_\mathbf{b}$ is a symmetric permutation of a block-diagonal matrix with diagonal-plus-rank-one blocks, and hence matrix-vector products with $\widehat{\Sigma}_\mathbf{b}^{-1}$ can be efficiently evaluated using the Woodbury identity, *i.e.*, $\widehat{\Sigma}_\mathbf{b}^{-1} = \Pi^T\mathrm{blkdiag}(S_1, \ldots, S_r)\Pi$ where

$$S_i = \mathrm{diag}(\Pi_i y) - \frac{1}{s(\hat{v}_\mathrm{f})_i + e_i^T Y\mathbf{1} + \alpha_i - 1}\Pi_i yy^T \Pi_i^T. \tag{33}$$

This allows us to evaluate the gradient as

$$\nabla J_5(u) = A^T \Sigma_\mathbf{b}^{-1}(Au - b)$$

which is Lipschitz continuous with constant $\|A^T\widehat{\Sigma}_\mathbf{b}^{-1}A\|_2$.

It is instructive to compare the SWLS model to the WLS model considered in [16]. This model implicitly includes the flat-fields using the following objective function

$$J_6(u,z) = \frac{1}{2}\|\mathrm{diag}(y)^{1/2}(Au - b + \mathbf{1} \otimes z)\|_2^2 + \frac{\lambda}{2}\|z\|_2^2 \tag{34}$$

where $z \in \mathbb{R}^r$ is an auxiliary variable that can be thought of as the relative flat-field error (cf. the analysis in Section II-C). Taking the gradient with respect to z and setting it equal to zero yields $z = \mathrm{diag}(Y\mathbf{1} + \lambda\mathbf{1})^{-1}(\mathbf{1}^T \otimes I)\mathrm{diag}(y)(b - Au)$, and using this expression in (34) yields

$$J_6(u) = \frac{1}{2}\|Au - b\|_{\widehat{\Sigma}^{-1}}^2 \tag{35}$$

where $\widehat{\Sigma}^{-1} = \Pi^T\mathrm{blkdiag}(\bar{S}_1, \ldots, \bar{S}_r)\Pi$ and

$$\bar{S}_i = \mathrm{diag}(\Pi_i y) - \frac{1}{e_i^T Y\mathbf{1} + \lambda}\Pi_i yy^T \Pi_i^T. \tag{36}$$

The blocks $\bar{S}_i$ clearly resemble the blocks S_i from the SWLS model in (33): the only difference is the scalar weight in front of the rank-1 term in each of the r blocks. In particular, notice that the weights in the SWLS model include information derived from all measurements as well as the flat-field prior. Moreover, the parameter λ in (36) plays a similar role as the flat-field hyperparameters α in (33), but the SWLS model is more general and flexible because it allows the use of a different hyperparameter α_i for each of the r blocks.

C. Algorithm

The functions $J_1(u), \ldots, J_5(u)$ are all differentiable with Lipschitz continuous gradients on the nonnegative orthant, and hence we can apply a proximal gradient method which is suitable for minimizing problems of the form

$$\text{minimize} \quad g(u) + h(u).$$

Here $g : \mathbb{R}^n \to \mathbb{R}$ is convex with a Lipschitz continuous gradient with Lipschitz constant L, $h : \mathbb{R}^n \to \mathbb{R}$ is convex, and the prox-operator

$$\mathrm{prox}_{th}(\bar{u}) = \underset{u}{\mathrm{argmin}}\left\{th(u) + \frac{1}{2}\|u - \bar{u}\|_2^2\right\}$$

is assumed to be cheap to evaluate. We will define $g(u) = J_i(u) + \gamma TV_\delta(u)$ and $h(u) = I_+(u)$, and hence the Lipschitz constant is given by $L = L_i + \gamma L_{tv}(\delta)$. Given a starting point $u^{(0)}$ and a fixed number of iterations K, the algorithm can be summarized as

$$u^{(k)} = \text{prox}_{th}(u^{(k-1)} - t\nabla g(u^{(k-1)})), \ k = 1, 2, \ldots, K$$

where $t \in (0, 2/L)$ is the step size and $\text{prox}_{th}(\bar{u}) = \max(0, \bar{u})$ is the projection onto the nonnegative orthant. With this step size, the method is a descent method. The Lipschitz constant L can be estimated without an explicit representation of A or D by means of the power iteration algorithm. Our MATLAB implementation of the method is available for download at https://github.com/hariagr/R2CT.

V. NUMERICAL EXPERIMENTS

A. Simulation Study

To evaluate the proposed reconstruction models, we conducted a series of experiments in MATLAB based on simulated data. In these experiments, we used a parallel beam geometry with $p = 720$ equidistant projection angles covering half a rotation, and a 2 cm wide photon counting detector array with $r = 512$ detector elements. To model a non-uniform detection efficiency, the elements of the flat-field vector v were drawn from a Poisson distribution with mean I_0. We used $s = 5$ measurements of the flat-field which were generated according to (4), and the measurements Y were generated according to (2) using a $2N \times 2N$ pixel discretization of a 2D phantom defined on a 4 cm^2 square. To avoid inverse crimes, we computed our reconstructions on an $N \times N$ ($N = 512$) pixel grid with a circular mask. The value of the TV-smoothing parameter δ was 0.01 cm^{-1} in all experiments with the TV-prior. We used as step size $t = 1.8/L$, and we used the ASTRA Toolbox [25] (version 1.7.1beta) to compute filtered backprojection (FBP) reconstructions and to implicitly compute products with A and A^T on a GPU. We generated the phantoms using the AIR Tools package [26] (version 1.3), and we used the method outlined in Section IV-C to numerically solve the reconstruction problems. As a remark, we note that the ASTRA GPU code for backprojection (i.e., multiplication by A^T) is not an exact adjoint of the forward operator (multiplication by A), and this may introduce small errors in the gradient computations. However, it is significantly faster than matched implementations, and we did not see any noticeable differences in reconstruction quality when using the exact adjoint.

As initial guess we used a vector of zeros, and we used a fixed number of iterations as stopping criteria (500 iterations for reconstructions without the TV-prior and 1,500 iterations for reconstructions with the TV-prior). We determine the parameter γ for the TV-prior based on the subjective visualization. As flat-field prior $P(v|\alpha, \beta)$ we used $\alpha_i = 1 + \beta_i(\hat{v}_f)_i$ and $\beta_i \geq 0$ (corresponding to the UP flat-field prior if $\beta_i = 0$ and the FE prior if $\beta_i > 0$), and for the attenuation prior $P(u|\gamma)$ we used either nonnegativity or total variation combined with nonnegativity. Note that SWLS only depends on α, but since

we also use $\alpha_i = 1 + \beta_i(\hat{v}_f)_i$ for SWLS, we report the value of β in the experiments.

To quantitatively compare the quality of reconstructions, we report the relative attenuation error (RAE)

$$e_u^{\text{rel}}(\hat{u}) = 100 \cdot \frac{\|\hat{u} - u\|_2}{\|u\|_2},$$

the relative flat-field error (RFE)

$$e_v^{\text{rel}}(\hat{v}) = 100 \cdot \frac{\|\hat{v} - v\|_2}{\|v\|_2},$$

the structural similarity (SSIM) index[1] [27], and a "ring ratio" (RR), defined as

$$\|\psi_v(\hat{v}(\hat{u}))\|_F / \|\psi_v(\hat{v}_f)\|_F$$

with $\psi_v(\hat{v})$ defined as

$$\psi_v(\hat{v}) = \text{FBP}(\text{diag}(v)^{-1}(\hat{v} - v)\mathbf{1}^T) \tag{37}$$

and where FBP denotes the filtered backprojection reconstruction method. In other words, $\psi_v(\hat{v})$ is the FBP reconstruction of the sinogram stripes due to flat-field estimation errors, and hence the norm $\|\psi_v(\hat{v})\|_F$ quantifies how severely the flat-field estimation errors affect the reconstruction. Thus, the RR can be viewed as an indication of the expected ring artifact reduction if we were to use the flat-field estimate $\hat{v}(\hat{u})$ instead of the ML estimate $\hat{v}_f$ (smaller is better) to compute a reconstruction. Recall that all but the JMAP reconstruction model are based on the ML estimate $\hat{v}_f$, so for the other models, the RFE and the RR simply reflect what we obtain if we were to use the reconstruction $\hat{u}$ to compute a new flat-field estimate $\hat{v}(\hat{u})$, using (22). We used $\alpha = 1$ and $\beta = 0$ to compute $\hat{v}(\hat{u})$ for all but the JMAP and SWLS reconstruction models.

1) Low Intensity: In our first experiment, we used a phantom based on the "grains" phantom from AIR Tools, shown in the upper left corner of Fig. 4. We applied a circular mask of radius 0.8 cm to obtain a phantom that is fully contained by the reconstruction grid. We used $I_0 = 500$ in this experiment, corresponding to approximately 500 photons per detector element per projection. As a result, the SNR is relatively low. Estimates based on low SNR measurements generally have a high variance, and hence a good model and strong priors are of paramount importance. The reconstructions shown in Fig. 4 demonstrate this. The baseline reconstructions were computed using the true flat-field, and hence they are "inverse crime" reconstructions that serve only as a baseline for comparison. The two baseline MAP reconstructions (with and without the TV prior) are based on the model (10). Using the flat-field estimate $\hat{v}_f$ instead of the true flat-field, we obtained the FBP and AMAP reconstructions. It is clear from these reconstructions that the flat-field estimation errors introduce severe ring artifacts, even in the presence of a strong prior such as the TV-prior. The ring artifacts are especially severe near the center of the image (cf. Section II-C).

---

[1]We used the MATLAB `ssim` function with the radius parameter equal to 0.2 for reconstructions without the TV-prior and equal to 2.0 for reconstructions with the TV-prior.
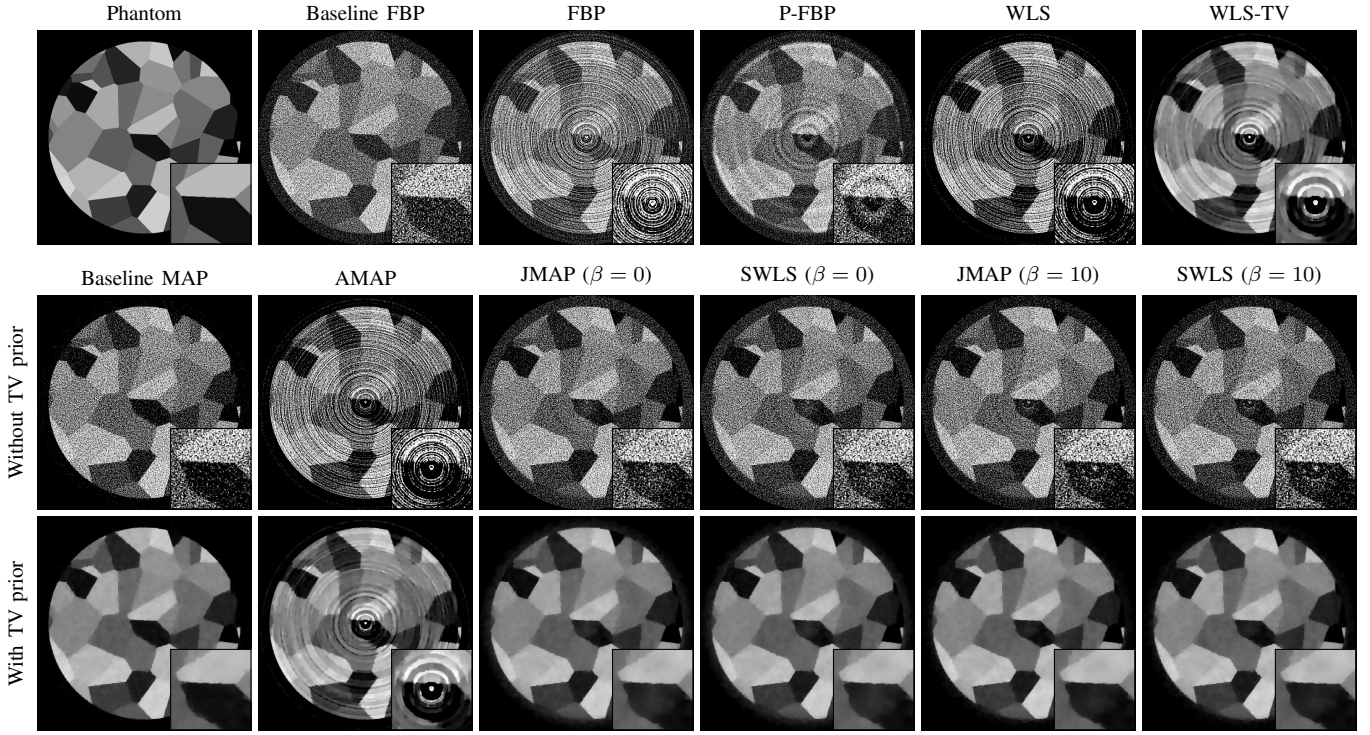
Fig. 4. Phantom and reconstructions based on simulated low-intensity measurements. The display range for the images is 0 to 1.2 cm^{-1}. The reconstructions with the TV-prior were computed with $\gamma = 3$. The insets are blow-ups of the reconstructions at the isocenter. The number of iterations was 500 for reconstructions without TV prior and 1,500 for reconstructions with TV prior.

Model	Without TV (full domain, 2×2 cm)				Without TV (disc, radius 0.8 cm)				TV $\gamma = 3$ (disc, radius 0.8 cm)			
	RAE	SSIM	RFE	RR	RAE	SSIM	RFE	RR	RAE	SSIM	RFE	RR
Baseline FBP	71.9	0.62	0.2	0.03	65.7	0.77	0.2	0.03	-	-	-	-
FBP	101.2	0.55	3.5	0.66	94.7	0.70	3.6	0.66	-	-	-	-
P-FBP	73.1	0.62	1.8	0.20	66.7	0.77	1.6	0.19	-	-	-	-
Baseline MAP	58.7	0.79	0.3	0.04	58.4	0.80	0.2	0.04	6.1	0.93	0.3	0.06
AMAP	77.3	0.72	2.8	0.50	76.9	0.74	2.9	0.52	15.2	0.71	1.7	0.19
WLS	76.9	0.72	2.8	0.50	76.6	0.74	2.9	0.52	15.2	0.71	1.7	0.19
JMAP ($\beta = 0$)	63.8	0.72	5.4	0.25	58.1	0.80	2.7	0.12	8.2	0.92	0.9	0.09
SWLS ($\beta = 0$)	63.9	0.72	5.5	0.26	58.0	0.80	2.7	0.12	8.3	0.92	1.0	0.10
JMAP ($\beta = 10$)	61.8	0.74	3.1	0.20	58.4	0.80	1.5	0.15	7.6	0.92	0.7	0.09
SWLS ($\beta = 10$)	61.8	0.74	3.2	0.20	58.3	0.80	1.4	0.15	7.7	0.92	0.8	0.09
JMAP ($\beta = 50$)	62.0	0.75	2.3	0.30	60.3	0.79	2.0	0.31	7.6	0.91	1.2	0.17
SWLS ($\beta = 50$)	61.9	0.75	2.3	0.30	60.2	0.79	2.0	0.31	7.7	0.91	1.3	0.17

TABLE I
ERROR MEASURES FOR RECONSTRUCTIONS BASED ON SIMULATED LOW-INTENSITY MEASUREMENTS.

The preprocessed FBP (P-FBP) reconstruction is the result of applying the combined wavelet and FFT filtering preprocessing method[2] by Münch et al. [12] to the sinogram, followed by FBP. This removes stripes from the sinogram, and although there are still some noticeable ring artifacts in the reconstruction, the preprocessing step clearly reduces the severity of the artifacts. However, the preprocessing step involves several parameters that must be carefully tuned, and it does not directly allow us to use the AMAP or MAP-based reconstruction models for reconstruction.

The proposed models are quite effective at reducing ring artifacts, as can be seen from the JMAP reconstructions as well as the SWLS reconstruction. Notice that both the SWLS

[2]We used a damping factor of 0.9 and a Daubechies 5 wavelet with a three-level decomposition.

($\beta = 0$) reconstruction and the JMAP ($\beta = 0$) reconstruction without the TV prior do not involve any parameters.

For the experiments without the TV-prior, Table I shows the error measures based on both the full reconstruction domain and based on a disc of radius 0.8 cm (corresponding to the support of the phantom). The latter approach ignores noise and ring artifacts outside the phantom, and hence this gives a more practical picture of the performance. For the reconstructions with a TV-prior, we report our results based on a disc of radius 0.8 cm. Notice that in all cases, we obtain the best reconstruction (in terms of both RAE and SSIM) using either the JMAP reconstruction model or the SWLS model. Moreover, these reconstructions have RAEs that are similar to those of the baseline MAP reconstructions. We also see that RRs and the RAEs for the JMAP reconstructions appear to be
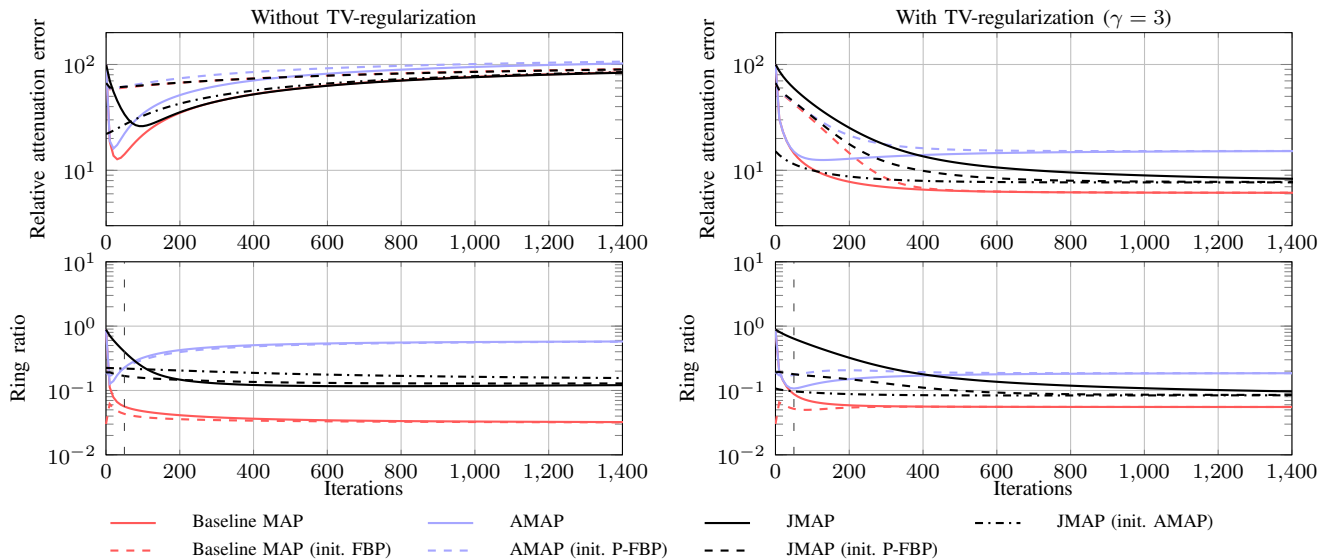
Fig. 5. Results of semi-convergence and initialization study. Reconstructions are computed with a UP prior $\beta = 0$.

correlated, but interestingly, the RFEs do not seem to agree with the RAEs.

Despite the fact that the P-FBP reconstruction is worse than the JMAP reconstructions, it is interesting to note that it may be used to compute an improved flat-field estimate. In our experiment, the ML estimate $\hat{v}_f$ had a relative error of 4.8%, but the flat-field estimate computed based on the P-FBP reconstruction had a relative error of only around 1.8%. However, using the TV-prior, the JMAP and SWLS model still produced the best flat-field estimate of all the models.

Finally, we remark that the AMAP and WLS reconstructions may be improved slightly by increasing the parameter γ. Using $\gamma = 10$, we obtained AMAP and WLS reconstructions with a relative error of around 10%, and although these reconstructions did not have noticeable ring artifacts, they contained an increased amount of undesirable TV-artifacts. On the other hand, the JMAP and SWLS reconstructions obtained with $\gamma = 3$ only have a limited amount of ring artifacts and TV artifacts, and hence we conclude that the proposed model allows us to reduce ring artifacts using a smaller regularization parameter γ than with the AMAP or WLS models, thus limiting unnecessary TV-induced artifacts.

2) Semi-convergence and Initialization: We now investigate the role of regularization and its influence on the reconstruction. Recall that X-ray tomographic imaging is an ill-posed problem where a small amount of noise in the measurements may results in a large change in the reconstruction if it is not regularized by a suitable prior. Thus, without regularization, intermediate iterates sometimes provide better reconstructions than iterates close to convergence. This behavior is known as semi-convergence and depends on the reconstruction method as well as initialization. Semi-convergence behavior often indicates that the reconstruction is under-regularized, and hence a solution to our convex reconstruction model may be a poor reconstruction. In practice it is difficult to rely on semi-convergence as the true solution is unknown.

We use the same experimental setup as in the previous experiment. Fig. 5 shows RAE and RR as a function of the number of iterations, with and without the TV-prior (*i.e.*, regularization). The semi-convergence behavior is evident without the TV-prior, and not surprisingly, the baseline reconstruction obtains the lowest RAE at the semi-convergence point after approximately 50 iterations. After the semi-convergence point, noise start to dominate the reconstruction and the RAE starts to increase monotonically. Comparing the AMAP and JMAP models, we see that the AMAP model has a lower RAE at the semi-convergence point, but it converges to a higher RAE. Taking the definition of the AMAP and JMAP estimators into account, we can conclude that the JMAP model still converges to a better reconstruction than the AMAP model. Fig. 5 also shows the RR error measure, and while the AMAP model exhibits semi-convergence both with respect to the RAE and the RR, the JMAP model appears to monotonically reduce the RR despite semi-convergence with respect to the RAE.

The dashed curves in Fig. 5 show the results of the same experiment, but using the P-FBP reconstruction of u as initialization (the baseline MAP was initialized with the baseline FBP reconstruction). The FBP reconstruction has a smaller RAE than the zero-initialization, but FBP reconstructions may be quite noisy when the SNR is low. Consequently, this initialization may not lead to faster convergence without regularization, as can be seen in Fig 5. The figure also shows that the AMAP reconstruction method still exhibits a mild degree of semi-convergence when using the TV-prior, but the baseline method and the JMAP method appear to reduce the RAE and the RR monotonically. Moreover, it is clear that the FBP-initialization helps when combined with the TV-prior. Finally, using the 50th AMAP iterate as initialization for JMAP (corresponding to the semi-convergence point for the RR), we obtained a significant improvement in the number of iterations when compared to initialization with zeros.
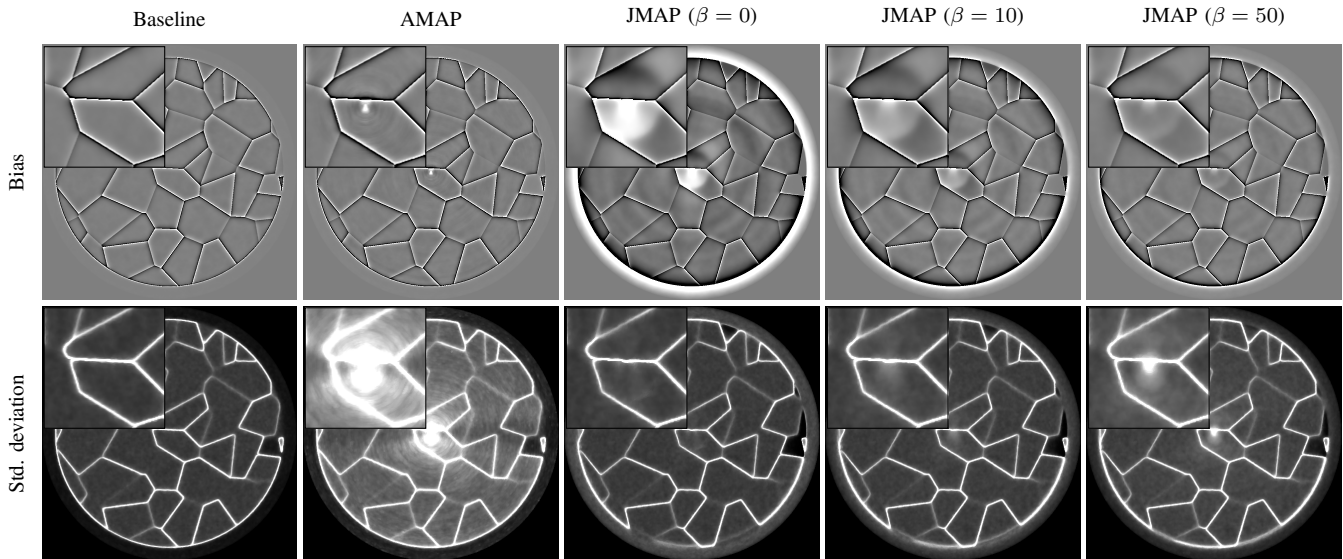
Fig. 6. Pixelwise bias and standard deviation based on 200 realizations of all measurements. The display range for the bias images is -0.1 to 0.1 cm^{-1}, and the display range for the standard deviation images is 0 to 0.06 cm^{-1}. The reconstructions are computed with TV-prior with $\gamma = 3$. The insets are blow-ups of the reconstructions at the isocenter.

3) Noise Analysis: To investigate the noise properties of the proposed reconstruction model, we generated 200 realizations of all measurements based on the grains phantom (see Fig. 4) and with $I_0 = 500$. We then computed pixelwise bias (the difference between the mean of the reconstructions and the phantom) and standard deviation for reconstructions based on the baseline MAP, the AMAP, and the JMAP reconstruction models. All reconstructions were computed with the TV-prior ($\gamma = 3$) and 1,500 iterations. The results are shown in Fig. 6. Generally speaking, the AMAP model is less biased than the JMAP model. For small values of β, the JMAP bias is somewhat large in comparison to the AMAP bias, especially near the boundary of the object and at the isocenter. However, the JMAP bias decreases when the parameter β is increased, but at the cost of increasing the standard deviation. This is consistent with the fundamental trade-off between bias and variance in statistical learning. More importantly, the standard deviation is significantly lower for the JMAP model in comparison to the AMAP model, and it is even comparable to that of the baseline MAP model when β is small. Notice that in all instances, the standard deviation is particularly large near the interfaces of the grains where the intensity jumps.

Recall from the previous experiment that the flat-field estimate may converge very slowly. As a consequence, the bias component that is induced by flat-field estimation errors decreases slowly as we increase the number of iterations. The results therefore depend on the stopping criteria (*i.e.*, the number of iterations). Finally we note that the noise results for the SWLS model were very similar to those of the the JMAP model, and hence we have chosen to omit the SWLS results for the sake of brevity.

4) Flat-field Regularization: Our next experiment demonstrates a potential shortcoming of the proposed model when using the UP flat-field prior for reconstruction. We used the Shepp–Logan phantom for the experiment, but unlike in

the previous experiments, we generated the measurements by evaluating the line integrals analytically. The intensity parameter was $I_0 = 10^5$. The reconstruction based on (14), the leftmost reconstruction in Fig. 7, has some low-level ring artifacts. JMAP with the FE prior and $\beta = 0$ leads to the reconstruction in the middle of Fig. 7. Somewhat surprisingly, while the low-level rings are mostly gone, the reconstruction has a few wide and very noticeable rings. These rings arise because of the structure of the flat-field estimation errors which can be seen by looking at the reconstruction $\psi_v(\hat{v})$, defined in (37) and shown in Fig. 7. Several high-intensity rings appear clearly, and these can be linked to large flat-field estimation errors associated with a small number of detector elements. In particular, the detector elements corresponding to rays that intersect the outer ellipsoidal shell of the Shepp–Logan phantom tangentially give rise to large estimation errors. We remark that we have observed experimentally that these artifacts seem to be exacerbated by the fact that the two outer Shepp–Logan ellipses are centered at the isocenter.

Now recall that the flat-field estimate $\hat{v}(u)$ can be expressed as (23), *i.e.*, a convex combination of independent estimates. Thus, the weights θ indicate the emphasis of the different flat-field estimates. The plots in Fig. 7 show these weights for two different priors parameterized by β. We see that when $\beta = 0$ (corresponding to the UP flat-field prior), the flat-field estimate is based almost entirely on $\hat{v}_y$, and the estimates $\hat{v}_f$ and $\hat{v}_{pr}$ both receive negligible (but nonzero) weights. Inspecting the corresponding flat-field estimate (the bottom plot in Fig. 7) reveals that for $\beta = 0$, the JMAP estimate is worse than the ML estimate $\hat{v}_f$. This indicates over-fitting. To mitigate this, we can emphasize the flat-field ML estimate $\hat{v}_f$ by using the FE prior (*i.e.*, $\alpha = 1 + \beta(\hat{v}_f)$), as described in III-B. Doing so effectively removes the major rings that were present with the FE prior with $\beta = 0$, as shown in the rightmost reconstruction in Fig. 7. Moreover, the rightmost plot in the figure confirms
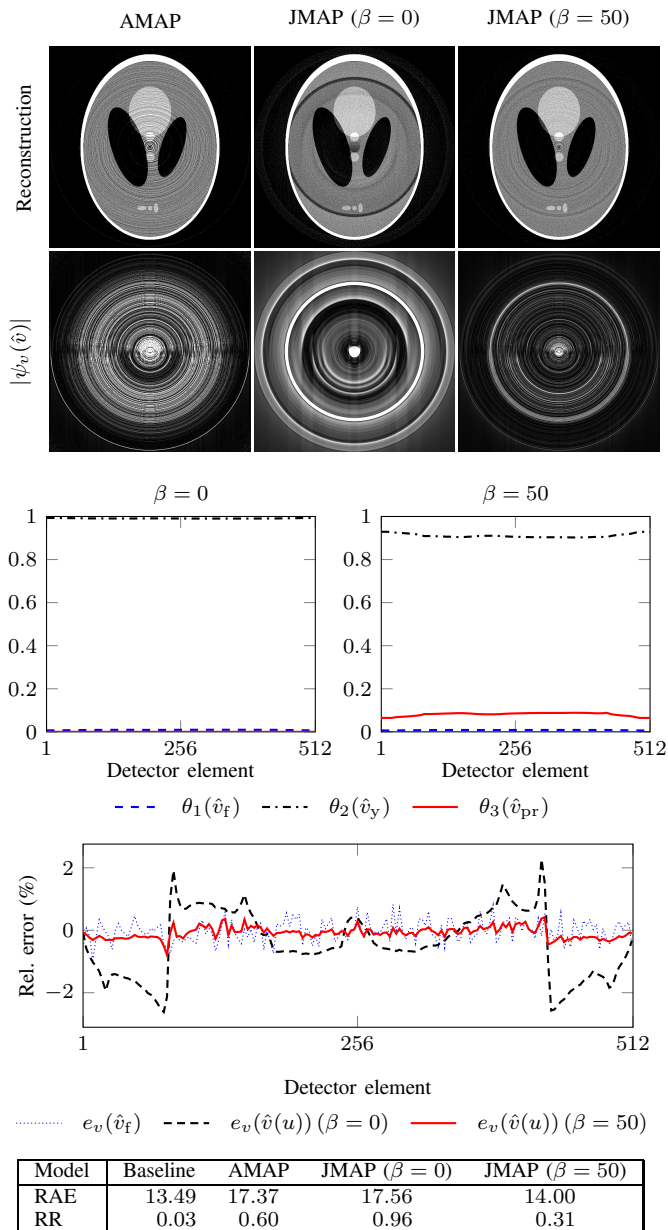
Fig. 7. Reconstructions of the Shepp–Logan phantom after 1,000 iterations, without the TV-prior on u. The display range for the reconstruction images is 0 to 0.4 cm^{-1}, and 0 to 0.04 for the ring images $\psi_v(\hat{v})$. The first two plots show the values of $\theta_1, \theta_2, \theta_3$, as defined in (23), for $\beta = 0$ and $\beta = 50$. The third plot shows the element-wise relative error with respect to true flat-field v, defined as $e_v(\hat{v}) = 100 \cdot \mathrm{diag}(v)^{-1}(\hat{v} - v)$, for the ML flat-field estimate $\hat{v}_{\mathrm{f}}$ and two JMAP flat-field estimates. The table lists the RAE and RR error measures.

that the resulting flat-field estimate depends less on $\hat{v}_{\mathrm{y}}$ than with the FE prior with $\beta = 0$. The FBP reconstructions of the flat-field error, shown below the reconstructions in Fig. 7, clearly show a reduction in ring artifacts compared to the basic AMAP and JMAP reconstructions.

B. Real Data Study

We now evaluate the performance of the proposed model based on real measurement data provided by the Advanced Photon Source (APS) facility operated by Argonne National Laboratory (USA). The data set provides tomographic measurements of a sample of glass beads with some dried potassium from $p = 900$ projection angles between $0°$ and $180°$ in a parallel beam geometry and with a 600×960 pixel detector array. In this experiment, we will consider only a 2D reconstruction of the center slice (slice 300) so we take $r = 960$. The energy of the X-ray source was 33.27 keV, and the photon flux per pixel in each projection was approximately 1200 photons/s. With an exposure time of only 6 ms, that amounts to pixelwise photon counts in the range 0-20 per projection. Out of a total of 20 flat-field measurements collected before and after the experiment, 8 appear to be corrupted, so we used $s = 12$ flat-field measurements for our reconstructions. Moreover, we used a square grid with side length 0.3053 cm and 768×768 pixels for the reconstructions. Our reconstructions are shown in the Fig. 8.

Without the TV-prior on the attenuation image, the reconstructions are quite noisy because of the low SNR. The FBP reconstruction and the AMAP reconstruction both have ring artifacts which heavily distort the reconstruction. The P-FBP reconstruction does not have noticeable ring artifacts, but the reconstruction is quite noisy. Thus, to reduce noise, we smoothed the FBP and P-FBP reconstructions using a Gaussian filter with standard deviation 1.0, and although this help, the resulting images are still somewhat noisy compared to the other reconstructions. The JMAP reconstruction with the UP prior ($\beta = 0$) has no noticeable ring artifacts, but it has a significant amount of noise. This is especially noticeable near the circular boundary of the object, and it may be because of flat-field estimation errors. Indeed, using the FE prior with $\beta = 200$ yields a reconstruction that is somewhat improved near the outer circles. Notice that the JMAP reconstructions do not have such a "hole" in the middle like the FBP, P-FBP, and AMAP reconstruction. Finally, including the TV-prior on u results in the AMAP-TV and JMAP-TV reconstructions. These results verify the applicability of proposed model for tomographic reconstruction based on low-intensity measurements.

VI. CONCLUSION

In X-ray computed tomography, the X-ray source intensity is typically estimated based on a number of flat-field measurements. This estimation introduces unavoidable errors in popular reconstruction models such as AMAP, WLS, and FBP, and these errors lead to systematic reconstruction errors in the form of ring artifacts. By investigating the filtered backprojection of a line in the sinogram, we have demonstrated that such systematic errors introduce structural changes in the reconstruction in the form of a ring. Based on the statistics of X-ray measurements, our analysis shows an inverse relationship between severity of ring artifacts and the source intensity. Therefore, these systematic errors can have a significant impact on the reconstruction quality of dose-constrained and time-constrained problems. To mitigate this problem, we have introduced a convex reconstruction model (JMAP) that jointly estimates the attenuation image and the flat-field. We have also introduced a quadratic approximation of the JMAP model, the
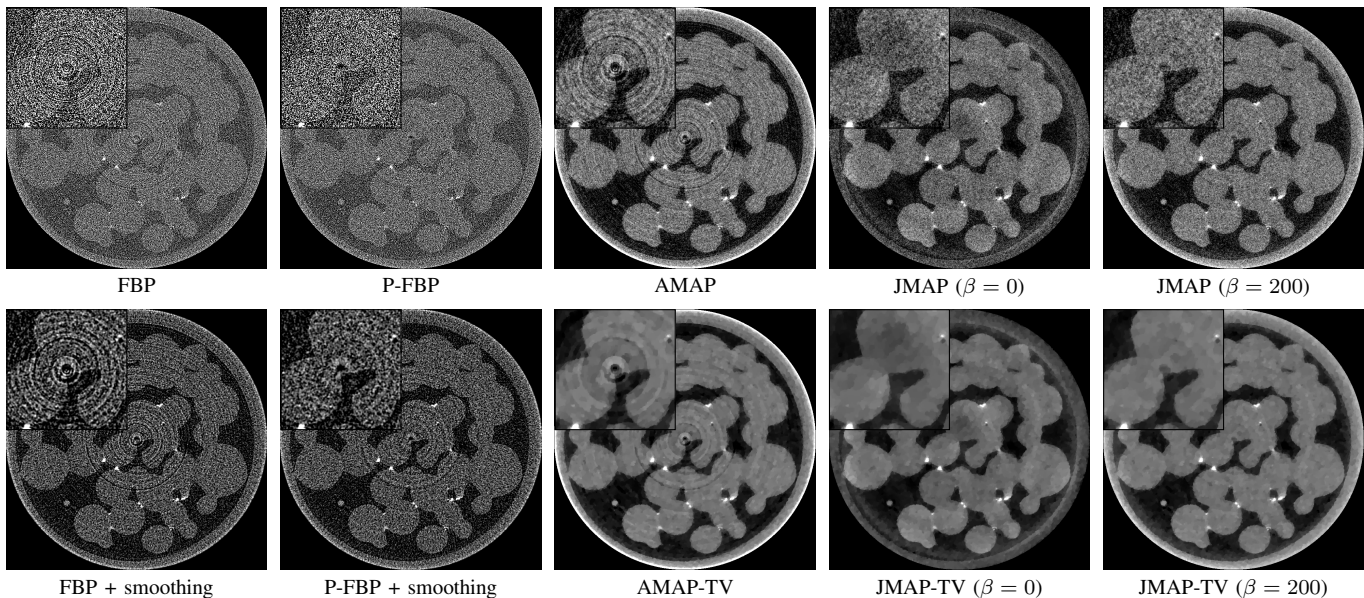
Fig. 8. Reconstructions of real tomographic measurements. The display range for the images is 0 to 10 cm^{-1}. The reconstructions using the TV-prior were obtained with $\gamma = 0.01$. The number of iterations were 50 for reconstructions without TV prior and 1,000 with TV prior. The insets are blow-ups of the reconstructions at the isocenter.

stripe-weighted least-squares (SWLS) model, which provides insight about the model and its similarities with existing models.

To assess the reduction of ring artifacts in the reconstructions, we have proposed a "ring ratio" error measure which quantifies the flat-field error in the image domain. Our experimental results indicate that the model effectively mitigates ring artifacts even for low SNR data, not only with simulated data but also with real data sets. In some cases, the proposed method may itself introduce artifacts when not appropriately regularized. These artifact essentially arise because of overfitting, and we have shown that they can be mitigated or supressed by means of a suitable regularizing flat-field prior. Moreover, we have shown experimentally that the JMAP and the SWLS models have similar performance in terms of noise and reconstruction quality.

Finally, we mention that the proposed methodology can readily be extended to estimate a time-varying flat-field which may be useful in applications where the flat-field does not remain stable while acquiring the tomographic measurements and/or when the scanner acquires projection images and flat-field images in an interleaved temporal order.

APPENDIX A
EXTREMA OF THE RADIAL PROFILE

The extrema of the radial profile $\tilde{\mu}(\rho)$, defined in (9), depend on the parameters t_0 and $\epsilon > 0$. To see this, we derive the critical points of $\tilde{\mu}(\rho)$. Setting the derivative equal to zero yields the equation

$$\tilde{\mu}'(\rho) = -3\rho\left(\sigma(\sigma^2 + \rho^2)^{-5/2} + \bar{\sigma}(\bar{\sigma}^2 + \rho^2)^{-5/2}\right) = 0$$

where $\sigma = \epsilon + it_0$. It follows that the critical points are $\rho = 0$ and any solution to the equation

$$\sigma(\sigma^2 + \rho^2)^{-5/2} + \bar{\sigma}(\bar{\sigma}^2 + \rho^2)^{-5/2} = 0,$$

or equivalently, $\rho = 0$ and solutions to the equation

$$\frac{\sigma}{\bar{\sigma}} = -\left(\frac{\sigma^2 + \rho^2}{\bar{\sigma}^2 + \rho^2}\right)^{5/2}.$$

Taking the complex logarithm of both sides of the equation yields the equation $2\angle\sigma + 2k\pi = \pi + 5\angle(\sigma^2 + \rho^2), k \in \mathbb{Z}$, and hence

$$\angle(\sigma^2 + \rho^2) = \frac{2}{5}\angle\sigma + \frac{2k-1}{5}\pi, \quad k \in \mathbb{Z}. \tag{38}$$

This implies that the tangent of $\angle(\sigma^2 + \rho^2)$ is equal to

$$\frac{2\epsilon t_0}{\rho^2 + \epsilon^2 - t_0^2} = \tan\left(\frac{2}{5}\angle\sigma + \frac{2k-1}{5}\pi\right), \quad k \in \mathbb{Z}, \tag{39}$$

or equivalently, if we define $c_k^{-1} = \tan\left(\frac{2}{5}\angle\sigma + \frac{2k-1}{5}\pi\right)$ and solve for ρ^2, we get $\rho^2 = 2\epsilon t_0 c_k + t_0^2 - \epsilon^2, k \in \mathbb{Z}$. Thus, in addition to $\rho = 0$, the real roots of the right-hand side of this equation are the critical points of $\tilde{\mu}(\rho)$, and hence we may limit our attention to $k \in \mathbb{Z}$ for which $2\epsilon t_0 c_k + t_0^2 - \epsilon^2 \geq 0$.

In order to find the extrema of $\tilde{\mu}(\rho)$, we now rewrite (9) as

$$\tilde{\mu}(\rho) = \frac{1}{4\pi}\frac{|\sigma|}{|\sigma^2 + \rho^2|^{3/2}}\cos(\angle\sigma - \angle(\sigma^2 + \rho^2)).$$

At a nonzero critical point $\rho_k \neq 0$, the angle $\angle(\sigma^2 + \rho_k^2)$ is given by (38), and it follows from (39) that

$$|\sigma^2 + \rho_k^2| = 2\epsilon|t_0|\left(c_k^2 + 1\right)^{1/2}.$$

This allows us to express the extrema associated with ρ_k as

$$\tilde{\mu}(\rho_k) = \frac{(\epsilon^2 + t_0^2)^{1/2}}{4\pi(1 + c_k^2)^{3/4}(2\epsilon|t_0|)^{3/2}}\cos\left(\frac{2}{5}\angle\sigma + \frac{2k-1}{5}\pi\right),$$

and it immediately follows that for $|t_0| \gg \epsilon$, the extrema are approximately inversely proportional to $\sqrt{\epsilon^3|t_0|}$.

APPENDIX B
INTERPRETATION OF FLAT-FIELD ESTIMATE

The ith element of flat-field estimate $\hat{v}$, defined in (22), is given by

$$\hat{v}_i(u) = \frac{\mathbf{1}^T f_i + \mathbf{1}^T y_i + \alpha_i - 1}{d_i(u)} \tag{40}$$

where $f_i \in \mathbb{R}^s$, $y_i \in \mathbb{R}^p$, $d_i(u) = s + \tau_i(u) + \beta_i$, and $\tau_i(u) = \sum_{j=1}^{p} \exp(-e_i^T A_j u)$. This expression can be reformulated as

$$\hat{v}_i(u) = \frac{s}{d_i(u)} \frac{\mathbf{1}^T f_i}{s} + \frac{\tau_i(u)}{d_i(u)} \frac{\mathbf{1}^T y_i}{\tau_i(u)} + \frac{\beta_i}{d_i(u)} \frac{\alpha_i - 1}{\beta_i}$$
$$= \frac{s}{d_i(u)}(\hat{v}_{\mathrm{f}})_i + \frac{\tau_i(u)}{d_i(u)}(\hat{v}_{\mathrm{y}})_i + \frac{\beta_i}{d_i(u)}\hat{v}_{\mathrm{pr}}(\alpha_i, \beta_i) \tag{41}$$

where the ML estimate $\hat{v}_{\mathrm{f}}$ is defined in (5), the estimate $\hat{v}_{\mathrm{y}}(\hat{u})$ is defined in (16), and

$$\hat{v}_{\mathrm{pr}}(\alpha, \beta) = \mathrm{diag}(\beta)^{-1}(\alpha - \mathbf{1})$$

is the mean of the Gamma prior. It follows from the definition (20), i.e., $d_i(u) = s + \tau_i(u) + \beta_i$, that

$$\frac{s}{d_i(u)} + \frac{\tau_i(u)}{d_i(u)} + \frac{\beta_i}{d_i(u)} = 1$$

and hence $\hat{v}_i(u)$ is a convex combination of three estimates. Thus, the full flat-field vector $\hat{v}(u)$ can be expressed as

$$\hat{v}(u) = \mathrm{diag}(\theta_1)\hat{v}_{\mathrm{f}} + \mathrm{diag}(\theta_2)\hat{v}_{\mathrm{y}}(\hat{u}) + \mathrm{diag}(\theta_3)\hat{v}_{\mathrm{pr}}(\alpha, \beta)$$

where $\theta_1 = \mathrm{diag}(d(u))^{-1}s\mathbf{1}$, $\theta_2 = \mathrm{diag}(d(u))^{-1}\tau(u)$, and $\theta_3 = \mathrm{diag}(d(u))^{-1}\beta$ with $\theta_1 + \theta_2 + \theta_3 = \mathbf{1}$.

APPENDIX C
TYPE-II ML ESTIMATION OF HYPERPARAMETERS

The marginal probability of $f_{i1}, \ldots, f_{is}$ given the hyperparameters α_i and β_i can be computed analytically and is given by

$$\mathrm{P}(f_{i1}, \ldots, f_{is} \mid \alpha_i, \beta_i)$$
$$= \int_0^\infty \mathrm{P}(f_{i1}, \ldots, f_{is} \mid v_i)\mathrm{P}(v_i \mid \alpha_i, \beta_i)dv_i$$
$$= \frac{\Gamma(k_i + \alpha_i)}{(\prod_{k=1}^s f_{ik}!)\,\Gamma(\alpha_i)\,s^{k_i}} \left(\frac{\beta_i}{s + \beta_i}\right)^{\alpha_i} \left(\frac{s}{s + \beta_i}\right)^{k_i} \tag{42}$$

where $k_i = \sum_{k=1}^s f_{ik}$. Here the identity $\int_0^\infty x^b e^{-ax}\,dx = \frac{\Gamma(b+1)}{a^{b+1}}$ was used to derive this expression. This probability distribution resembles the negative binomial distribution, and it follows from the first-order optimality conditions associated with (24) that $\beta_i = s\alpha_i/k_i$, or equivalently, $\alpha_i/\beta_i = k_i/s$. This implies that the mean of the Gamma prior is equal to the flat-field ML estimate $(\hat{v}_{\mathrm{f}})_i$. Substituting the expression for β_i in (24), we obtain the one-dimensional problem $\mathrm{argmin}_{\alpha_i} \kappa_i(\alpha_i)$ where

$$\kappa_i(\alpha_i) = -\log \frac{\Gamma(k_i + \alpha_i)}{\Gamma(\alpha_i)} - \alpha_i \log \frac{\alpha_i}{\alpha_i + k_i} - k_i \log \frac{k_i}{\alpha_i + k_i}.$$

The derivative of $\kappa_i(\alpha_i)$ is

$$\kappa_i'(\alpha_i) = -\left[F(k_i + \alpha_i) - F(\alpha_i) - \log\left(1 + \frac{k_i}{\alpha_i}\right)\right]$$
$$= -\sum_{l=0}^{k_i-1} \frac{1}{\alpha_i + l} + \log\left(1 + \frac{k_i}{\alpha_i}\right),$$

where $F(x)$ denotes the digamma function. Similarly, the second derivative is given by

$$\kappa_i''(\alpha_i) = \sum_{l=0}^{k_i-1} \frac{1}{(\alpha_i + l)^2} - \frac{k_i}{\alpha_i(\alpha_i + k_i)} \tag{43}$$

where the summation satisfies the inequality

$$\sum_{l=0}^{k_i-1} \frac{1}{(\alpha_i + l)^2} = \sum_{n=\alpha_i}^{\alpha_i + k_i - 1} \frac{1}{n^2} > \int_{\alpha_i}^{\alpha_i + k_i} \frac{1}{x^2}\,dx$$
$$= \frac{k}{\alpha_i(\alpha_i + k_i)} \tag{44}$$

for $\alpha_i > 0$. This shows that $\kappa_i''(\alpha_i) > 0$ for $\alpha > 0$, and hence κ_i is convex on the positive real line. Moreover, since $\kappa'(\alpha_i)$ tends to zero as as α_i tends to infinity, $\kappa'(\alpha_i)$ can not have a positive zero. Consequently, the resulting flat-field Gamma prior has zero variance (i.e., α_i/β_i^2 tends to zeros for $\alpha_i \to \infty$ since $\beta_i = s\alpha_i/k_i$) and its mean is equal to the empirical flat-field estimate, i.e., $\alpha_i/\beta_i = (\hat{v}_{\mathrm{f}})_i$.

REFERENCES

[1] T. M. Buzug, *Computed Tomography - From photon statistics to modern cone-beam CT*, 1st ed. Springer-Verlag Berlin Heidelberg, 2008.
[2] B. R. Whiting, P. Massoumzadeh, O. A. Earl, J. A. O'Sullivan, D. L. Snyder, and J. F. Williamson, "Properties of preprocessed sinogram data in X-ray computed tomography." *Med. Phys.*, vol. 33, no. 9, pp. 3290–3303, 2006.
[3] J. Sijbers and A. Postnov, "Reduction of ring artefacts in high resolution micro-CT reconstructions." *Phys. Med. Biol.*, vol. 49, no. 14, pp. N247–53, jul 2004.
[4] M. H. DeGroot and M. J. Schervish, *Probability and Statistics*, 4th ed. Pearson, 2012.
[5] G. Kowalski, "The influence of fixed errors of a detector array on the reconstruction of objects from their projections," *IEEE Trans. Nucl. Sci.*, vol. 24, no. 5, pp. 2006–2016, jun 1977.

[6] C. Thomas, M. Heuschmid, D. Schilling, D. Ketelsen, I. Tsiflikas, A. Stenzl, C. D. Claussen, and H.-P. Schlemmer, "Urinary calculi composed of uric acid, cystine, and mineral salts: differentiation with dual-energy CT at a radiation dose comparable to that of intravenous pyelography." *Radiology*, vol. 257, no. 2, pp. 402–9, nov 2010.

[7] P. Dahlman, A. J. Van Der Molen, M. Magnusson, and A. Magnusson, "How much dose can be saved in three-phase CT urography? A combination of normal-dose corticomedullary phase with low-dose unenhanced and excretory phases." *AJR. Am. J. Roentgenol.*, vol. 199, no. 4, pp. 852–60, oct 2012.

[8] R. Fahrig, R. Dixon, T. Payne, R. L. Morin, A. Ganguly, and N. Strobel, "Dose and image quality for a cone-beam C-arm CT system," *Med. Phys.*, vol. 33, no. 12, p. 4541, dec 2006.

[9] S. J. Doran, K. K. Koerkamp, M. A. Bero, P. Jenneson, E. J. Morton, and W. B. Gilboy, "A CCD-based optical CT scanner for high-resolution 3D imaging of radiation dose distributions: equipment specifications, optical simulations and preliminary results," *Phys. Med. Biol.*, vol. 46, no. 12, p. 3191, 2001.

[10] G. Kowalski, "Suppression of ring artefacts in CT fan-beam scanners," *IEEE Trans. Nucl. Sci.*, vol. 25, no. 5, pp. 1111–1116, oct 1978.

[11] C. Raven, "Numerical removal of ring artifacts in microtomography," *Rev. Sci. Instrum.*, vol. 69, no. 8, p. 2978, aug 1998.

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

[13] S. Rashid, S. Lee, and M. Hasan, "An improved method for the removal of ring artifacts in high resolution CT imaging," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, p. 93, 2012.

[14] T. Kim, J. Baek, and D. Hwang, "Ring artifact correction using detector line-ratios in computed tomography," *Opt. Express*, vol. 22, no. 11, pp. 13 380–92, jun 2014.

[15] P. Paleo and A. Mirone, "Ring artifacts correction in compressed sensing tomographic reconstruction." *J. Synchrotron Radiat.*, vol. 22, no. 5, pp. 1268–78, sep 2015.

[16] K. A. Mohan, S. V. Venkatakrishnan, J. W. Gibbs, E. B. Gulsoy, X. Xiao, M. De Graef, P. W. Voorhees, and C. A. Bouman, "TIMBIR: A method for time-space reconstruction from interlaced views," *IEEE Trans. Comput. Imag.*, vol. 1, no. 2, pp. 96–111, jun 2015.

[17] D. Prell, Y. Kyriakou, and W. A. Kalender, "Comparison of ring artifact correction methods for flat-detector CT." *Phys. Med. Biol.*, vol. 54, no. 12, pp. 3881–95, jun 2009.

[18] L. Yan, T. Wu, S. Zhong, and Q. Zhang, "A variation-based ring artifact correction method with sparse constraint for flat-detector CT." *Phys. Med. Biol.*, vol. 61, no. 3, pp. 1278–92, feb 2016.

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

[20] S. Rose, M. S. Andersen, E. Y. Sidky, and X. Pan, "Noise properties of CT images reconstructed by use of constrained total-variation, data-discrepancy minimization," *Med. Phys.*, vol. 42, no. 5, pp. 2690–8, may 2015.

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

[22] H. Bateman, *Table of Integral Transforms, Vol. 2*. McGraw-Hill, 1954.

[23] J. Berger, *Statistical Decision Theory and Bayesian Analysis*. Springer New York, 1985.

[24] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D Nonlinear Phenom.*, vol. 60, no. 1-4, pp. 259–268, nov 1992.

[25] W. van Aarle, W. J. Palenstijn, J. D. Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, "The ASTRA toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, oct 2015.

[26] P. C. Hansen and M. Saxild-Hansen, "AIR Tools—a MATLAB package of algebraic iterative reconstruction methods," *J. Comput. Appl. Math.*, vol. 236, no. 8, pp. 2167–2178, feb 2012.

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

Hari Om Aggrawal received his M.Tech. in Electrical Engineering from the Indian Institute of Technology Kanpur, India, in 2011, and he is presently pursuing a PhD degree at the Technical University of Denmark in the Section for Scientific Computing at the Department of Applied Mathematics and Computer Science. His interests include image reconstruction models, methods for tomographic imaging, and image registration.

Martin S. Andersen received his M.S. in Electrical Engineering from Aalborg University, Denmark, in 2006 and his PhD in Electrical Engineering from the University of California, Los Angeles, in 2011. After receiveing his degree, he was a postdoc in the Division of Automatic Control at Linköping University, Sweden, and at the Technical University of Denmark. He is currently an Associate Professor at the Technical University of Denmark in the Section for Scientific Computing at the Department of Applied Mathematics and Computer Science. His research interests include optimization, numerical methods, signal and image processing, and systems and control.

Sean D. Rose was born in Cincinnati, OH. He received a B.S. degree in physics and biochemistry from The Ohio State University in 2013. He began his graduate studies in Medical Physics at the University of Chicago in the summer of 2013 in the lab of Dr. Xiaochuan Pan where he continues to work on the development and optimization of iterative methods for tomographic image reconstruction. His research interests are in optimization based image reconstruction for tomographic imaging modalities, objective assessment of image quality and its application to parameter selection, and large-scale optimization.

Emil Y. Sidky has held academic positions in Physics at the University of Copenhagen and Kansas State University. He then switched to Medical Physics joining the University of Chicago in 2001, where he is now a Research Associate Professor. His current interests are image reconstruction in X-ray computed tomography and digital breast tomosynthesis, large-scale optimization, and objective assessment of image quality.