Edinburgh Research Explorer

Spectral X-ray CT for fast NDT using discrete tomography

Link:
[Link to publication record in Edinburgh Research Explorer](#)

Document Version:
Peer reviewed version

Spectral X-ray CT for fast NDT using discrete tomography

D. Kamilis[*], S. Lee[†], J. Desjardins[†] and N. Polydorides[*,‡]

[*]School of Engineering, The University of Edinburgh, UK.
[†]Photonics & Quantum Sciences Core Technology Centre, Harris Corporation, FL.
[‡]The Alan Turing Institute, London, UK.

Abstract

We present progress in fast, high-resolution imaging, material classification, and fault detection using hyperspectral X-ray measurements. Classical X-ray CT approaches rely on data from many projection angles, resulting in long acquisition and reconstruction times. Additionally, conventional CT cannot distinguish between materials with similar densities. However, in additive manufacturing, the majority of materials used are known a priori. This knowledge allows to vastly reduce the data collected and increase the accuracy of fault detection. In this context, we propose an imaging method for non-destructive testing of materials based on the combination of spectral X-ray CT and discrete tomography. We explore the use of spectral X-ray attenuation models and measurements to recover the characteristic functions of materials in heterogeneous media with piece-wise uniform composition. We show by means of numerical simulation that using spectral measurements from a small number of angles, our approach can alleviate the typical deterioration of spatial resolution and the appearance of streaking artifacts.

1 Introduction

Motivated by recent advances in spectral X-ray detectors [1] and algorithms in discrete tomography [2], we show that upon formulating an inverse problem for energy-resolved attenuation data, quantitative imaging is feasible from a limited data set, comprising only projections from a small number of angles. Effectively, we anticipate that the proposed methodology can expedite the data acquisition process in non-destructive testing of such media, and increase screening throughput with minimal compromise on spatial resolution. Our approach is fundamentally based on poly-energetic attenuation models that capture beam-hardening phenomena and a new formulation of the inverse unmixing problem that leads to a number of binary image reconstruction problems for the support functions of the medium's constituent materials. Essentially this process renders the resulting imaging process suitable to the framework of discrete tomography, that can achieve full-angle CT levels of spatial resolution from a very small number of acquisition angles.

2 Spectral attenuation model

We consider the application of non-destructive testing of three-dimensional domains Ω consisting of a relatively small number of N disjoint regions such that

$$\Omega = \bigcup_{k=1}^{N} \Omega_k, \quad \Omega_k \cap \Omega_l = \emptyset \quad \text{if} \quad k \neq l, \tag{1}$$

using non-monochromatic diagnostic X-ray sources and energy-discriminating detectors. Without any loss of generality, the measurements are assumed to be captured on flat panel pixelized detectors with $N_u \times N_v$ photon-counting cells, each of which can resolve the arriving photons into N_b energy resolving bins according to some user-defined energy-sensitivity profile. Realistically, the above experimental quantities satisfy

$$N_u, N_v \gg N_b \geq N > 1. \tag{2}$$

1

In this context, let a beam trajectory L depart from the source at some direction angle θ and end on an arbitrary detector cell. Define a vector of N parameters $\boldsymbol{\beta} \in \mathbb{R}^N$ with elements

$$\beta_k = \int_L \mathrm{d}x\, \chi_k(x), \quad k = 1, \ldots, N, \tag{3}$$

where χ_k is the characteristic function of the k-th material in the sample defined as

$$\chi_k(x) = \begin{cases} 1 & \text{if } x \in \Omega_k, \\ 0 & \text{elsewhere.} \end{cases} \tag{4}$$

Then the noiseless model for the expected spectral measurement given the sample of N materials is

$$G(\boldsymbol{\beta})_b = \int_0^\infty \mathrm{d}e\, S(e) D_b(e) \exp\left\{ -\sum_{k=1}^N \mu_k(e)\beta_k \right\} \tag{5}$$

where $G : \boldsymbol{\beta} \mapsto \mathbf{y}$ is the nonlinear parameters-to-data map that describes the poly-energetic spectral attenuation [3]. In the above, $S(e)$ is the source photon flux density spectrum, $D_b(e)$ is the spectral sensitivity function of the bth bin of the detector, ideally taken uniform between some bounds, and $\mu_k(e)$ is the linear attenuation coefficient spectrum corresponding to the kth material in the sample. Normalising the source flux to a probability distribution function as

$$S(e) = p_S(e) y_0, \quad \text{with} \quad \int_0^\infty \mathrm{d}e\, p_S(e) = 1, \tag{6}$$

and introducing the mass attenuation spectrum of the kth material as

$$m_k(e) = \mu_k(e)\rho_k^{-1},$$

for a density $\rho_k > 0$, the mean spectral measurement model (5) becomes

$$G(\boldsymbol{\beta})_b = y_0 \int_0^\infty \mathrm{d}e\, p_S(e) D_b(e) \exp\left\{ -\sum_{k=1}^N \rho_k\, m_k(e)\beta_k \right\}, \tag{7}$$

for which y_0, $p_S(e)$ (resp. $S(e)$) and $D_b(e)$ can be obtained from the specifications of the source and the detector, while $m(e)$ for a selection of materials can be traced from the NIST databases or otherwise estimated from spectroscopic measurements [4]. Incorporating noise effects, the photon count measurement at the bth bin of the cell is known to be a Poisson random variable

$$y_b \sim \mathcal{P}(\bar{y}_b), \tag{8}$$

for a mean value $\bar{y}_b$ given by the conditional expectation over $\boldsymbol{\beta}$

$$\bar{y}_b \doteq \mathbb{E}[y_b|\boldsymbol{\beta}] = G(\boldsymbol{\beta})_b + \bar{\eta}_b(\boldsymbol{\beta}) \geq 0. \tag{9}$$

where $\bar{\eta}_b > 0$ denotes the mean of a random variable $\eta_b(\boldsymbol{\beta})$ incorporating the contributions of the coherent and incoherent scattering effects in the measurements. From the measurement likelihood (8), and the mean observation model (7) the probability of the measurement given the integrals-of-the-characteristics parameters $\boldsymbol{\beta}$ is

$$p(y_b|\boldsymbol{\beta}) = \frac{1}{y_b!} \bar{y}_b^{y_b} \exp\{-\bar{y}_b\}, \tag{10}$$

and as the bin photon counts $\mathbf{y} = \{y_1, \ldots, y_{N_b}\}$ are statistically independent the probability of the spectral measurements vector $\mathbf{y}$ given the parameters $\boldsymbol{\beta}$ is

$$p(\mathbf{y}|\boldsymbol{\beta}) = \prod_{b=1}^{N_b} p(y_b|\boldsymbol{\beta}). \tag{11}$$

We end this section by mentioning that from the spectral data $\mathbf{y}$ one can readily compute the 'conventional' energy-integrating scalar attenuation measurement as $\mathbf{1}^T\mathbf{y}$, while in replacing the source with a mono-energetic source the spectral integral in (7) collapses and the model reduces to the scalar expression of Beer-Lambert's law [5].

2.1 Discrete formulation

To compute the measurement expectation we must discretize the model (7). Using a three-dimensional grid of N_g voxels, we can express the discretized parameters (with a slight abuse of notation) as $\beta_k = \sum_{i=1}^{N_g} |L|_i \chi_k^i$, where $|L|_i$ denotes the length of the beam trajectory L crossing voxel i and the binary characteristic function $\chi_k^i = 1$ if the kth material occupies voxel i. Similarly, using a discretization of the energy domain into N_e intervals, we can express energy-dependent quantities in terms of N_e coefficients which we write as vectors $\boldsymbol{m}_k, \boldsymbol{p}_S, \boldsymbol{D}_b$ for the mass attenuation coefficient, the normalized source spectrum and the detector spectral sensitivity respectively. Collecting also the $\boldsymbol{m}_k$ into a $N_e \times N$ matrix M, we can write

$$G(\boldsymbol{\beta})_b = y_0 \, \boldsymbol{p}_S^T \left(\boldsymbol{D}_b \odot \exp\left\{ -M \operatorname{diag}(\boldsymbol{\rho}) \, \boldsymbol{\beta} \right\} \right), \tag{12}$$

for $b = 1, \ldots, N_b$, where $\odot$ denotes element-wise multiplication of vectors. A quantity that will be needed for the optimization problem is the derivative $\partial_{\boldsymbol{\beta}} G(\boldsymbol{\beta})_b$ which is given by

$$\partial_{\boldsymbol{\beta}} G(\boldsymbol{\beta})_b = -y_0 \, \boldsymbol{p}_S^T \left(\boldsymbol{D}_b \odot \exp\left\{ -M \operatorname{diag}(\boldsymbol{\rho}) \, \boldsymbol{\beta} \right\} M \operatorname{diag}(\boldsymbol{\rho}) \right), \tag{13}$$

and collecting the $\boldsymbol{D}_b$ into an $N_e \times N_b$ matrix D, we can write the Jacobian matrix $J(\boldsymbol{\beta}) = \partial_{\boldsymbol{\beta}} \boldsymbol{G}(\boldsymbol{\beta})$ as

$$J(\boldsymbol{\beta}) = -D^T \operatorname{diag}\left(y_0 \boldsymbol{p}_S \odot \exp\left\{ -M \operatorname{diag}(\boldsymbol{\rho}) \boldsymbol{\beta} \right\} \right) M \operatorname{diag}(\boldsymbol{\rho}). \tag{14}$$

3 Imaging methodology

Our hypothesis is that estimating $\boldsymbol{\beta}$ from the noisy data $\mathbf{y}$ decouples the inverse problem of imaging the attenuation spectra of the materials within the domain into N discrete inverse problems for the binary-valued characteristic functions $\chi_1(x), \ldots \chi_N(x)$. If necessary, the resulting images can then be scaled by the densities of the respective materials and merged into a single spectral image, often referred to as coloured X-ray image.

3.1 Recovery of $\boldsymbol{\beta}$

To begin with, an approach to estimate $\boldsymbol{\beta}$ is to compute the maximum likelihood estimator by working out the negative log-likelihood of the parameters

$$H(\boldsymbol{\beta}) = -\log p(\mathbf{y}|\boldsymbol{\beta}) = \mathbf{1}^T \bar{\mathbf{y}} - \mathbf{y}^T \log \bar{\mathbf{y}}. \tag{15}$$

The minimization of H is equivalent to the minimization of the (discrete) Kullback-Leibler (KL) divergence, which can be considered as a data misfit measure when regressing Poisson data, given by

$$D_{\mathrm{KL}}(\mathbf{y}\|\bar{\mathbf{y}}) = \mathbf{y}^T \log\left(\frac{\mathbf{y}}{\bar{\mathbf{y}}} \right) + \mathbf{1}^T (\bar{\mathbf{y}} - \mathbf{y}) \tag{16}$$

with the conventions $0 \log(0) = 0$ and $-\log(x) = \infty$, $x \leq 0$. Incorporating the positivity constraint on the parameters $\boldsymbol{\beta}$ and the additional consistency constraint

$$\sum_k \chi_k(x) = 1, \; \forall x \in \Omega \Rightarrow \sum_k \beta_k = |L|, \quad \text{for each } L, \tag{17}$$

where $|L|$ denotes the length of the beam trajectory L, we formulate the penalized minimization problem as

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in \Delta} D_{\mathrm{KL}}(\mathbf{y}\|\bar{\mathbf{y}}(\boldsymbol{\beta})), \quad \text{or equivalently} \quad \hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} D_{\mathrm{KL}}(\mathbf{y}\|\bar{\mathbf{y}}(\boldsymbol{\beta})) + \imath_\Delta(\boldsymbol{\beta}) \tag{18}$$

where $\imath_\Delta$ is the indicator function on the simplex Δ defined as

$$\Delta \doteq \left\{ \boldsymbol{\beta} \in \mathbb{R}^N | \sum_{k=1}^N \beta_k = |L| \, \text{and} \, \boldsymbol{\beta} \geq \mathbf{0} \right\}. \tag{19}$$

3

Due to the non-linearity of the measurements model

$$\bar{\mathbf{y}}(\boldsymbol{\beta}) = \boldsymbol{G}(\boldsymbol{\beta}) + \bar{\boldsymbol{\eta}}(\boldsymbol{\beta}), \tag{20}$$

D_{KL} is generally not guaranteed to be convex (unless specific conditions such as $\bar{\mathbf{y}} \geq \mathbf{y}$ are satisfied, see appendix A) and may have local minima, rendering the optimization a difficult task. In our extensive numerical tests, we have found that a non-linear variant of the Alternating Direction Method of Multipliers (ADMM) [6] performs well for this problem. The ADMM with non-linear operator constraints is a convex optimization method for non-smooth functionals, which offers flexibility in the use of a variety of smooth and non-smooth penalization/regularization terms. The formulation adapted to our problem is described in Algorithm 1 and requires as input, apart from the functions $\imath_\Delta, D_{\mathrm{KL}}, \bar{\mathbf{y}}$, two parameters δ and $0 < c < 1$ controlling the step size, which after tests we set to values $\delta = 0.001$ and $c = 0.95$. Since the algorithm is based on the linearization of the forward operator, we calculate the derivative as $\partial_{\boldsymbol{\beta}}\bar{\mathbf{y}}(\boldsymbol{\beta}) \approx J(\boldsymbol{\beta})$ where we have ignored contributions from scattering. Additionally, we require the proximal operator for the indicator function $\imath_\Delta$ which takes the form of solving a projection problem onto the simplex and can be calculated fast by e.g. the method in [7]. On the other hand, the proximal operator of D_{KL} with parameter σ is given in closed form by

$$\mathrm{prox}_{\sigma D_{\mathrm{KL}}}(\boldsymbol{x}) = \boldsymbol{x} - \frac{\sigma\mathbf{1} + \boldsymbol{x} - \sqrt{(\boldsymbol{x}/\sigma - 1)^2 + 4\sigma\boldsymbol{y}}}{2}, \tag{21}$$

After initialization of $\boldsymbol{\beta}$ to some value that satisfies constraint eq. (17), the algorithm proceeds by updating the primal variable $\boldsymbol{\beta}$ and the dual variable $\boldsymbol{\mu}$. Finally, the algorithm as requires the estimation of the spectral norm $\|J\boldsymbol{\beta}\|$ in each iteration, which quickly becomes computationally expensive. To alleviate this, we resort to a randomized SVD method to approximate the largest singular value. This is calculated for each of the first 20 iterations and then only every 30-40 iterations as the spectral norm doesn't change significantly in between.

Algorithm 1 Algorithm based on preconditioned ADMM with nonlinear operator constraint [6]

Input: $\imath_\Delta, D_{\mathrm{KL}}, \bar{\mathbf{y}}, \delta, c$
Initialization: $\boldsymbol{\beta}^0, \boldsymbol{z}^0, \boldsymbol{\mu}^0, \overline{\boldsymbol{\mu}}^0 = \boldsymbol{\mu}^0, k = 0, \sigma = \delta/2$
while convergence criterion is not met **do**

$\quad A^k = J(\boldsymbol{\beta}^k)$
$\quad$Set $\tau^k = c/\delta\|A^k\|^2$
$\quad \boldsymbol{\beta}^{k+1} = \mathrm{prox}_{\tau^k \imath_\Delta}\left(\boldsymbol{\beta}^k - \tau^k (A^k)^T \overline{\boldsymbol{\mu}}^k\right)$
$\quad$Set $\boldsymbol{w} = \bar{\mathbf{y}}(\boldsymbol{\beta}^{k+1})$
$\quad \boldsymbol{z}^{k+1} = \mathrm{prox}_{\sigma D_{\mathrm{KL}}}\left(\boldsymbol{z}^k + \sigma(\boldsymbol{\mu}^k + \delta(\boldsymbol{w} - \boldsymbol{z}^k))\right)$
$\quad \boldsymbol{\mu}^{k+1} = \boldsymbol{\mu}^k + \delta(\boldsymbol{w} - \boldsymbol{z}^{k+1})$
$\quad \overline{\boldsymbol{\mu}}^{k+1} = 2\boldsymbol{\mu}^{k+1} - \boldsymbol{\mu}^k$
$\quad k = k + 1$

end while
return $\boldsymbol{\beta}^k$

3.2 Recovery of characteristic functions of materials

At the next step, from the estimated meta-data $\boldsymbol{\beta}$, we reconstruct the binary images χ_k for each material $k = 1, \ldots, N$, using the discrete tomography DART algorithm [2] in combination with unsupervised segmentation. In the original publication of the DART algorithm, the authors use a simple global thresholding scheme to segment the reconstruction. This can be take the form for example of Otsu's thresholding method. However, more sophisticated segmentation methods can improve the segmentation accuracy and in turn the reconstruction result. We have opted to use the method of Morphological Active Contours without Edges (MorphACWE) which belongs to the family of Morphological Snakes methods for image segmentation [8]. The original Active Contours model is based on matching a deformable model (splines) to an image by minimising an energy functional. Within this general context, the ACWE model is based on a level-set formulation where the curve is evolved by propagating an interface represented by the zero level-set of a

smooth function, using a time-dependent partial differential equation (PDE). However, such a formulation suffers from high computational cost and numerical stability issues. The MorphACWE method aims to resolve these issues by replacing terms in the contour evolution PDE with morphological operators such as dilation and erosion that have equivalent infinitesimal behaviour and whose effect can be computed stably fast. We outline one iteration of the algorithm in table 1

Initialisation: compute an initial image $\hat{\chi}_k^0$ by running a conventional continuous tomography algorithm, e.g. conjugate gradient method for normal equations.

Segmentation: use MorphACWE to produce a segmented binary image $\hat{\chi}_k^1$ from $\hat{\chi}_k^0$.

Clustering: cluster the pixels of $\hat{\chi}_k^1$ in two groups: those for which we have high confidence (fixed) and the rest (free). This can be achieved by setting the free pixels as the boundary pixels in $\hat{\chi}_k^1$, identified using the gradient magnitude. The image gradient can be estimated by applying the Sobel operator (convolving in each direction with the Sobel 3×3 kernels). The boundary pixels are then augmented with some random pixels chosen with a fixed probability $0 < p \leq 1$. The fixed pixels are defined as the remaining pixels in $\hat{\chi}_k^1$.

Re-calibration: keeping the fixed pixels at their existing segmented value, we compute their contribution to the data $\boldsymbol{\beta}_k$ and then solve an imaging problem re-estimating only the free pixels from the residual data (lots of data, few pixel unknowns) which gives $\hat{\chi}_k^2$.

Smoothing: smooth the reconstructed segmented image by applying a Gaussian filter to obtain a continuous image $\hat{\chi}_k^3$ and loop back to the segmentation stage.

Table 1: DART algorithm with MorphACWE unsupervised segmentation.

4   Numerical examples
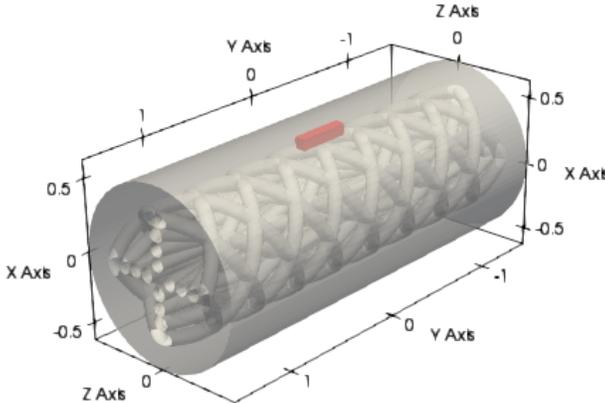
Figure 1: Target: silica (negative) preform with lead glass inhomogeneity (red).

In this section we present some numerical results demonstrating our proposed method, based on a 3D parallel-beam geometry, though the algorithm applies unchanged to cone beam sources as well. In this case the imaging target is a silica (SiO_2, density=2.32 g/cm^3) fiber (negative) preform with a $0.08 \times 0.4 \times 0.08$ cm^3 cuboid, lead glass (density=6.22 g/cm^3) inhomogeneity inserted as shown in fig. 1. The positive preform and the surrounding space are both air (density=1.205 mg/cm^3) for a total of $N = 3$ materials with attenuation coefficients shown in fig. 2 (data taken from the NIST database [4]). The materials are embedded into a $[-0.625, 0.625] \times [-1.5, 1.5] \times [-0.625, 0.625]$ cm^3 simulation space that is discretized into $N_g = 512^3$ voxels, while the flat 2D $[-1.65, 1.65] \times [-0.825, 0.825]$ cm^2 detector is discretized into $N_u \times N_v = 128 \times 64$ pixels and can discriminate photon energies within 8 bins distributed evenly in log-space as shown in fig. 2. We make measurements at $N_\theta = 20$ different angles evenly distributed from $0°$ to $90°$. The source is a 140kVp x-ray
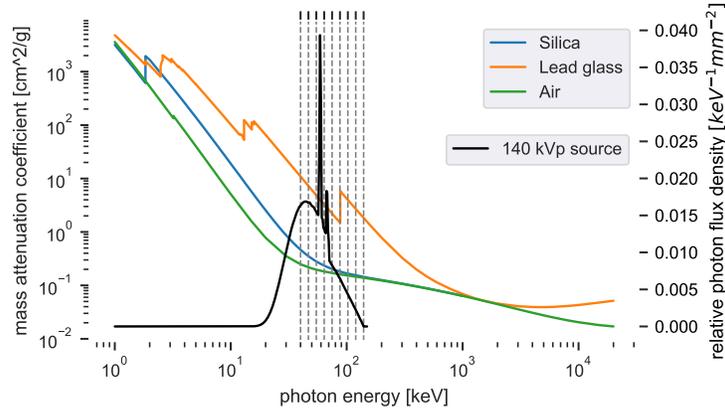
Figure 2: Mass attenuation coefficients for silica (SiO$_2$), lead glass and air from NIST data [4]. The source spectrum is for a 140kVp source. The vertical dashed lines depict the intervals for the 8 energy bins used.

source with spectrum generated with Spektr [9] as shown in fig. 2. For the numerical example we ignore scattering effects η which is a valid approximation in the energies $40 - 140$ assumed. The implementation of the model is done in Python using the well known open-source libraries NumPy, SciPy as well as the specialized libraries ASTRA toolbox [10, 11] for 3D tomography, ITK toolkit [12] for image processing and the ODL library [13] for efficient model specification and optimization.

4.1 Reconstruction of β

We set the initial intensity/dose to $y_0 = 10^8$ and simulate a synthetic dataset of dimension $8 \times 20 \times 128 \times 64$ before introducing Poisson noise. We run the optimization algorithm as described in section 3.1 starting from a single material occupying the whole volume. We set the update step size in ADMM at $\delta = 0.001$ and perform 5000 iterations, with the final results depicted in fig. 3. Note that we only report the results for one angle as those from the remaining 19 ones are very similar. The reconstruction is excellent as reported with metrics the Structural Similarity Index (SSIM) of the image and the relative error defined as $\|\beta_k - \hat{\beta}_k\| / \|\beta_k\|$ (where quantities are now taken as 128×64 long vectors for beam trajectories L at one angle). Note that the optimization method is sensitive to the choice of parameters for this problem, requiring a small step size and therefore many iterations.

4.2 Reconstruction of material images

From the reconstructed 'unmixing data' $\hat{\beta}_k$, we reconstruct the spatial support images (characteristic functions) of the three materials using the algorithm in 3.2. We perform 50 iterations of the algorithm with parameters the probability $p = 0.15$, with Gaussian smoothing of standard deviation $\sigma = 1$, and with MorphACWE set to 5 iterations for segmentation. The results ($x = 0$ slice and $y = 0$ slice) are shown in fig. 4. Note that for the reconstruction of the air image, there is no significant improvement with iterations of the algorithm so we only report the initial segmented image. The results show that the main features of the images are reconstructed to a satisfactory degree although there are some artifacts and distortions due to the limited-angle data. It's important to note that the lead glass inhomogeneity is clearly identified which may be the final task objective in an NDT setting.

5 Conclusions

Spectral X-ray CT measurements include information that can compensate for the lack of full angle data, and can thus lead to a shorter acquisition times. To exploit this advantage we have developed a discrete tomography algorithm that combines model-based inversion and segmentation within the reconstruction process to produce images of spatial resolution equivalent to that of full angle tomography. Numerical tests

(a) β_k for silica.

(b) β_k for lead glass.

(c) β_k for air.

(d) $\hat{\beta}_k$ for silica. SSIM: 0.9997, rel. error: 8.82×10^{-2}

(e) $\hat{\beta}_k$ for silica. SSIM: 1.0000, rel. error: 7.89×10^{-2}

(f) $\hat{\beta}_k$ for silica. SSIM: 0.9978, rel. error: 1.42×10^{-2}
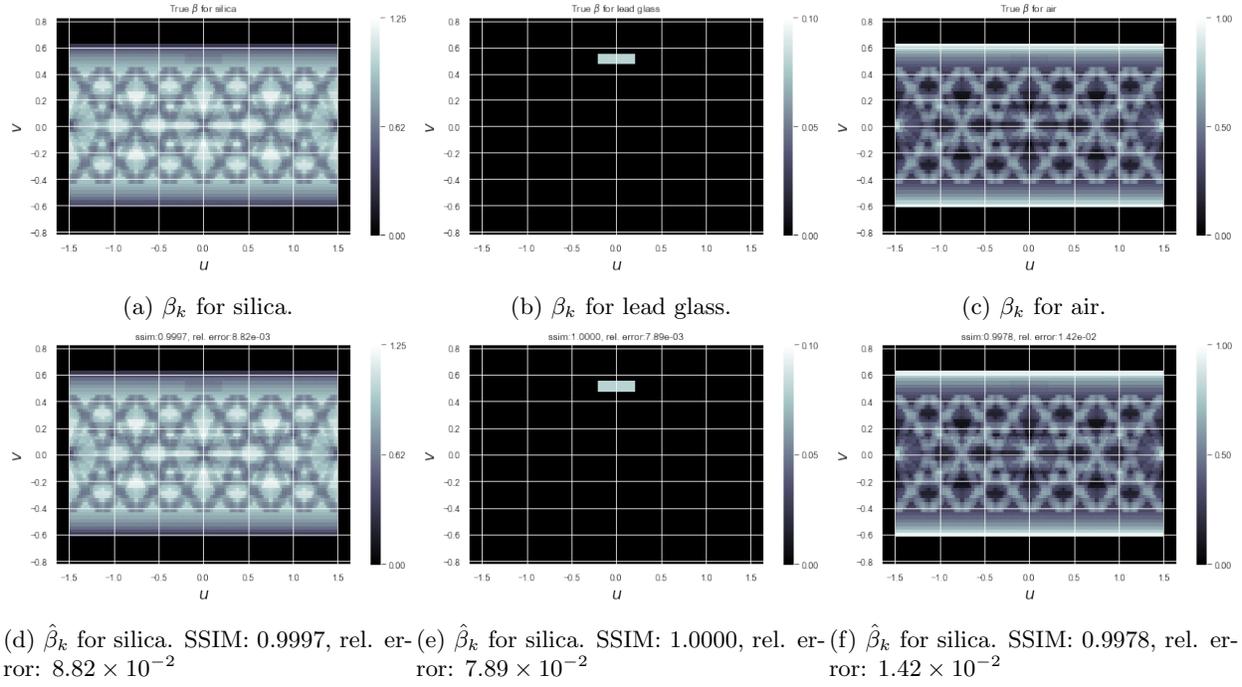
Figure 3: Reconstruction of $\boldsymbol{\beta}$. Top: true parameters β_k for each material. Bottom: reconstructed parameters $\hat{\beta}_k$ for corresponding material with SSIM and relative error metrics.

on objects with discrete material composition, like those encountered in additive manufacturing, indicate good performance and robustness to noise.

Acknowledgements
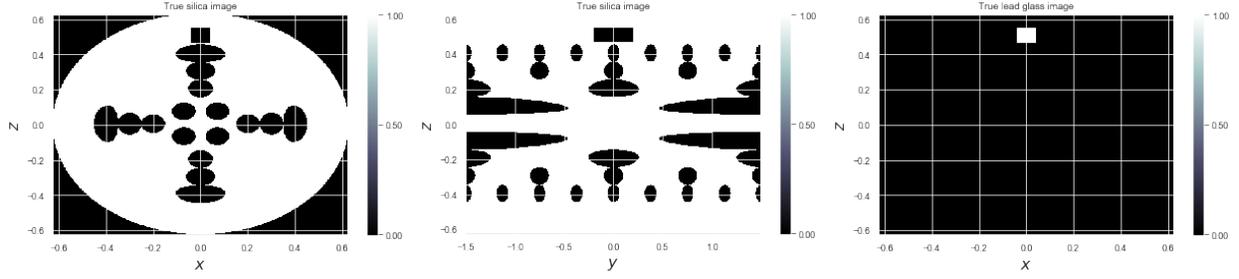
(a) $y = 0$ slice for true silica image.

(b) $x = 0$ slice for true silica image.

(c) $y = 0$ slice for true lead glass image.
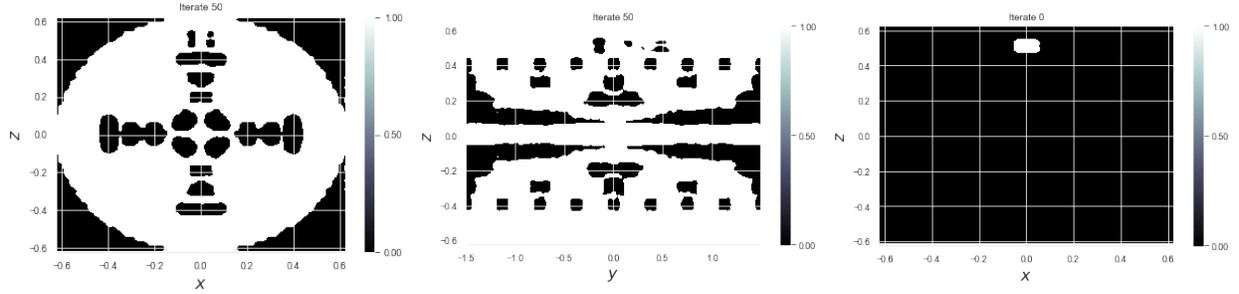
(d) $y = 0$ slice for reconstructed silica image (50 iter.).

(e) $x = 0$ slice for reconstructed silica image (50 iter.).

(f) $y = 0$ slice for reconstructed lead glass image (0 iter.).
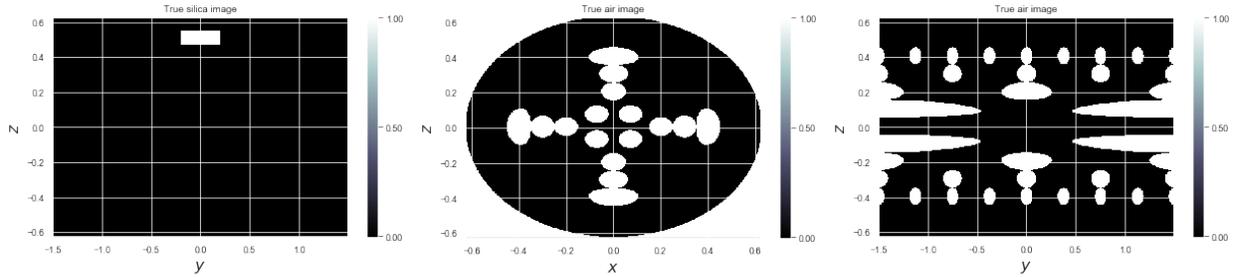
(g) $x = 0$ slice for true lead glass image.

(h) $y = 0$ slice for true air image.

(i) $x = 0$ slice for true air image.
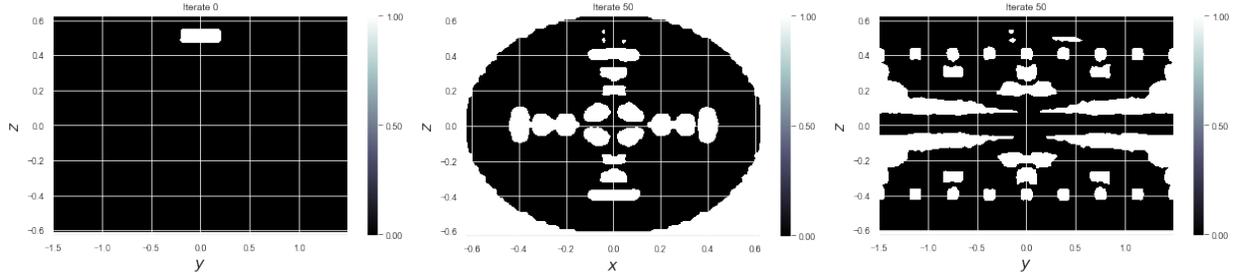
(j) $x = 0$ slice for reconstructed lead glass image (0 iter.).

(k) $y = 0$ slice for reconstructed air image (50 iter.).

(l) $x = 0$ slice for reconstructed air image (50 iter.).

Figure 4: True $\hat{\chi}_k$ and reconstructed $\hat{\chi}_k$ material images ($y = 0$ and $x = 0$ slices) for silica, lead glass and air.

A Conditions for the convexity of $L(\boldsymbol{\beta})$

Define $\boldsymbol{F}(\boldsymbol{\beta}) \doteq y_0 \, \boldsymbol{p}_S \odot \exp\left\{-M \operatorname{diag}(\boldsymbol{\rho})\, \boldsymbol{\beta}\right\}$. Then the partial derivatives of $L(\boldsymbol{\beta})$ are given by

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_k} = \left(\mathbf{1} - \frac{\mathbf{y}}{\bar{\mathbf{y}}}\right)^T \frac{\partial \bar{\mathbf{y}}}{\partial \beta_k},$$

where $\frac{\partial \bar{\mathbf{y}}}{\partial \beta_k} = \left[\frac{\partial \bar{y}_1}{\partial \beta_k}, \ldots, \frac{\partial \bar{y}_{N_b}}{\partial \beta_k}\right]$ and

$$\frac{\partial \bar{y}_b}{\partial \beta_k} = -\boldsymbol{D}_b^T(\boldsymbol{m}_k \rho_k \odot \boldsymbol{F}(\boldsymbol{\beta})).$$

Therefore, defining the $N_b \times N_e$ matrix $D = [\boldsymbol{D}_1, \ldots, \boldsymbol{D}_{N_b}]^T$, we can write for the Jacobian matrix

$$\nabla L(\boldsymbol{\beta}) = -\left(\mathbf{1} - \frac{\mathbf{y}}{\bar{\mathbf{y}}}\right)^T D \operatorname{diag}(\boldsymbol{F}(\boldsymbol{\beta})) M \operatorname{diag}(\boldsymbol{\rho}). \tag{22}$$

The elements of the Hessian matrix are given by

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_l \partial \beta_k} = \left(\mathbf{1} - \frac{\mathbf{y}}{\bar{\mathbf{y}}}\right)^T \frac{\partial^2 \bar{\mathbf{y}}}{\partial \beta_l \partial \beta_k} + \left(\frac{\mathbf{y}}{\bar{\mathbf{y}}^{\odot 2}}\right)^T \left(\frac{\partial \bar{\mathbf{y}}}{\partial \beta_k} \odot \frac{\partial \bar{\mathbf{y}}}{\partial \beta_l}\right),$$

where $\frac{\partial^2 \bar{\mathbf{y}}}{\partial \beta_l \partial \beta_k} = \left[\frac{\partial^2 \bar{y}_1}{\partial \beta_l \partial \beta_k}, \ldots, \frac{\partial^2 \bar{y}_{N_b}}{\partial \beta_l \partial \beta_k}\right]$ and

$$\frac{\partial^2 \bar{y}_b}{\partial \beta_l \partial \beta_k} = \boldsymbol{D}_b^T(\boldsymbol{m}_k \rho_k \odot \boldsymbol{m}_l \rho_l \odot \boldsymbol{F}(\boldsymbol{\beta}))$$

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_l \partial \beta_k} = \left(\mathbf{1} - \frac{\mathbf{y}}{\bar{\mathbf{y}}}\right)^T D(\boldsymbol{m}_k \rho_k \odot \boldsymbol{m}_l \rho_l \odot \boldsymbol{F}(\boldsymbol{\beta})) \tag{23}$$
$$+ \left(\frac{\mathbf{y}}{\bar{\mathbf{y}}^{\odot 2}}\right)^T (D(\boldsymbol{m}_k \rho_k \odot \boldsymbol{F}(\boldsymbol{\beta}))) \odot (D(\boldsymbol{m}_l \rho_l \odot \boldsymbol{F}(\boldsymbol{\beta})))$$

The Hessian can thus be written as

$$H = \operatorname{diag}(\boldsymbol{\rho})(H1 + H2)\operatorname{diag}(\boldsymbol{\rho})$$

where

$$H_1 = M^T \operatorname{diag}\left\{\left(\mathbf{1} - \frac{\mathbf{y}}{\bar{\mathbf{y}}}\right)^T D \operatorname{diag}(\boldsymbol{F}(\boldsymbol{\beta}))\right\} M$$

and

$$H_2 = M^T (D \operatorname{diag}(\boldsymbol{F}(\boldsymbol{\beta})))^T \operatorname{diag}\left(\frac{\mathbf{y}}{\bar{\mathbf{y}}^{\odot 2}}\right) D \operatorname{diag}(\boldsymbol{F}(\boldsymbol{\beta})) M$$

Assuming $\mathbf{y} > 0$, $\bar{\mathbf{y}} > 0$ and $D, \boldsymbol{F}(\boldsymbol{\beta})$ are positive, it is evident that H_2 is positive definite since $\operatorname{diag}\left(\frac{\mathbf{y}}{\bar{\mathbf{y}}^{\odot 2}}\right)$ is positive definite and $\operatorname{rank}[D \operatorname{diag}(\boldsymbol{F}(\boldsymbol{\beta})) M] = N$. We can therefore simultaneously diagonalize H_1 and H_2 [14, Theorem 7.6.4], to write $H_2 = SIS^T$ and $H_1 = S\Lambda S^T$ for some non-singular S and a diagonal Λ with the eigenvalues of $H_2^{-1}H_1$ in its diagonal. Then, if $\bar{\mathbf{y}} \geq \mathbf{y}$, H_1 is positive semidefinite and thus Λ is non-negative diagonal, which leads to a positive definite H. If the assumption $\bar{\mathbf{y}} \geq \mathbf{y}$ is not true, then some of the entries of Λ will be negative and we require that $\Lambda_{ii} \geq -1$ for $i = 1, \ldots, N$. In conclusion, if the above assumptions are valid $L(\boldsymbol{\beta})$ is strictly convex in the open domain $\boldsymbol{\beta} > 0$ since the Hessian is positive-definite. By continuity, it is also convex in the domain $\boldsymbol{\beta} \geq 0$.

References

[1] P. Seller et al. "Pixellated Cd(Zn)Te high-energy X-ray instrument". In: *Journal of Instrumentation* (2011). ISSN: 1748-0221. DOI: `10.1088/1748-0221/6/12/C12009`.

[2] K J Batenburg and J Sijbers. "DART: A Practical Reconstruction Algorithm for Discrete Tomography". In: *IEEE Transactions on Image Processing* 20.9 (Sept. 2011), pp. 2542–2553. DOI: `10.1109/tip.2011.2131661`.

[3] Robert E. Alvarez and Albert Macovski. "Engergy-selective reconstructions in X-ray computerized tomography". In: *Physics in medicine and biology* 21.5 (1976), pp. 733–744. DOI: `10.1088/0031-9155/21/5/002`.

[4] Stephen Seltzer. *Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients, NIST Standard Reference Database 126*. eng. 1995. DOI: `10.18434/t4d01f`.

[5] Thorsten Buzug. *Computed tomography: From photon statistics to modern cone-beam CT*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 1–521. ISBN: 9783540394075. DOI: `10.1007/978-3-540-39408-2`. arXiv: `arXiv:1011.1669v3`.

[6] Martin Benning et al. "Preconditioned ADMM with nonlinear operator constraint". In: *arXiv e-prints*, arXiv:1511.00425 (Nov. 2015), arXiv:1511.00425. arXiv: `1511.00425 [math.NA]`.

[7] Laurent Condat. "Fast projection onto the simplex and the l1 ball". In: *Mathematical Programming* 158 (2016), pp. 575–585. ISSN: 0025-5610. DOI: `10.1007/s10107-015-0946-6`.

[8] Pablo Marquez-Neila, Luis Baumela, and Luis Alvarez. "A Morphological Approach to Curvature-Based Evolution of Curves and Surfaces". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36 (2013), pp. 2–17. ISSN: 0162-8828. DOI: `10.1109/tpami.2013.106`.

[9] *SPEKTR v3.0*. https://istar.jhu.edu/downloads/.

[10] Wim van Aarle et al. "The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography". In: *Ultramicroscopy* 157 (Oct. 2015), pp. 35–47. ISSN: 0304-3991. DOI: `10.1016/j.ultramic.2015.05.002`.

[11] Wim van Aarle et al. "Fast and flexible X-ray tomography using the ASTRA toolbox". In: *Optics Express* 24.22 (Oct. 2016), p. 25129. ISSN: 1094-4087. DOI: `10.1364/OE.24.025129`.

[12] Hans J. Johnson et al. *The ITK Software Guide*. Third. *In press*. Kitware, Inc. 2013.

[13] *Operator Discretization Library*. https://odlgroup.github.io/odl/index.html.

[14] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. 2013. ISBN: 978-0-521-54823-6.