Blind X-ray CT Image Reconstruction from Polychromatic Poisson Measurements

Renliang Gu and Aleksandar Dogandžić, *Senior Member, IEEE*

***Abstract*—We develop a framework for reconstructing images that are sparse in an appropriate transform domain from polychromatic computed tomography (CT) measurements under the blind scenario where the material of the inspected object and incident-energy spectrum are unknown. Assuming that the object that we wish to reconstruct consists of a single material, we obtain a parsimonious measurement-model parameterization by changing the integral variable from photon energy to mass attenuation, which allows us to combine the variations brought by the unknown incident spectrum and mass attenuation into a single unknown *mass-attenuation spectrum* function; the resulting measurement equation has the Laplace-integral form. The mass-attenuation spectrum is then expanded into basis functions using B-splines of order one. We consider a Poisson noise model and establish conditions for biconvexity of the corresponding negative log-likelihood (NLL) function with respect to the density-map and mass-attenuation spectrum parameters. We derive a block-coordinate descent algorithm for constrained minimization of a penalized NLL objective function, where penalty terms ensure nonnegativity of the mass-attenuation spline coefficients and nonnegativity and gradient-map sparsity of the density-map image, imposed using a convex total-variation (TV) norm; the resulting objective function is biconvex. This algorithm alternates between a Nesterov's proximal-gradient (NPG) step and a limited-memory Broyden-Fletcher-Goldfarb-Shanno with box constraints (L-BFGS-B) iteration for updating the image and mass-attenuation spectrum parameters, respectively. We prove the Kurdyka-Łojasiewicz property of the objective function, which is important for establishing local convergence of block-coordinate descent schemes in biconvex optimization problems. Our framework applies to other NLLs and signal-sparsity penalties, such as lognormal NLL and ℓ_1 norm of 2D discrete wavelet transform (DWT) image coefficients. Numerical experiments with simulated and real X-ray CT data demonstrate the performance of the proposed scheme.**

***Index Terms*—X-ray CT, statistical model-based iterative reconstruction (MBIR), beam-hardening correction.**

I. Introduction

X-ray computed tomography (CT) measurement systems are important in modern nondestructive evaluation (NDE) and medical diagnostics. The past decades have seen great progress in CT hardware and (reconstruction) software development. CT sees into the interior of the inspected object and gives 2D and 3D reconstruction at a high resolution. It is a fast, high-resolution method that can distinguish density differences as small as 1 %. As it shows the finest interior detail, it has been one of the most important techniques in medical diagnosis, material analysis and characterization, and NDE [1, 2]. Thanks to recent computational and theoretical advances, such as graphics processing units (GPUs) and sparse signal reconstruction theory and methods, it is now possible to design iterative reconstruction methods that incorporate accurate nonlinear physical models into sparse signal reconstructions from significantly undersampled measurements.

Due to the polychromatic nature of the X-ray source and the fact that mass attenuation generally decreases as a function of photon energy, the center of the spectrum shifts to higher energy as X-rays traverse the object, an effect known as "hardening" [3]. This effect destroys the linearity between the attenuation coefficient and the logarithm of the noiseless measurements. Therefore, linear reconstructions such as filtered backprojection (FBP) exhibit beam-hardening artifacts, e.g., cupping and streaking [4, Ch. 7.6], which limit the quantitative analysis of the reconstruction. In medical CT applications, severe artifacts can look similar to certain pathologies and further mislead the diagnosis [4, Sec. 7.6.2]. Fulfilling the promise of compressed sensing and sparse signal reconstruction in X-ray CT depends on accounting for the polychromatic measurements, in addition to other effects such as ring artifacts, metal artifacts in medical applications, X-ray scatter, and detector crosstalk and afterglow [5, 6]. It is not clear how aliasing and beam-hardening artifacts interact, and our experience is that we cannot achieve great undersampling when applying sparse linear reconstruction to polychromatic measurements. Indeed, the error caused by the model mismatch may well be larger than the aliasing error that we wish to correct using sparse signal reconstruction.

Beam-hardening correction methods can be categorized into pre-filtering, linearization, dual-energy, and post-reconstruction approaches [7]. Reconstruction methods have recently been developed in [8–10] that aim to optimize nonlinear objective functions based on the underlying physical model; [8, 9] assume known incident polychromatic source spectrum and imaged materials, whereas [10] considers a blind scenario for a lognormal measurement model with *unknown* incident spectrum and imaged materials, but employs a photon-energy discretization [11, eq. (2)], [4, Sec. 8.4] with an excessive number of parameters (which leads to permutation and scaling ambiguities; see [11] for details) and suffers from numerical instability [12]. The methods in [10] do not impose sparsity of the reconstructed density-map image, only its nonnegativity, and they have been tested in [10] using real and noiseless simulated data.

It is often expensive to determine the X-ray spectrum and the materials of the object. X-ray spectrum measurements based on semiconductor detectors are usually distorted by charge trapping, escape events, and other effects [13], and the corresponding correction requires a highly collimated beam and special procedures [14]. Even after measuring the spectrum, it is not feasible to scan

The authors are with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: {renliang,ald}@iastate.edu). Portions of this work have been presented at 2015 Rev. Prog. Quant. Nondestr. Eval. (QNDE) and 2016 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP). This work was supported by the U.S. National Science Foundation under Grant CCF-1421480.

different objects with fixed scanning configurations, e.g., X-ray tube voltage, current, prefiltrations, and scanning time. Knowing the mass-attenuation function can be challenging as well when the inspected material is unknown or the inspected object is made of a compound or a mixture with an unknown percentage of each constituent.

In this paper (see also [11, 12, 15]), we adopt the nonlinear measurement scenario resulting from the polychromatic X-ray source and formulate a parsimonious measurement-model parameterization by exploiting the relationship between the *mass-attenuation coefficients*, *X-ray photon energy*, and *incident spectrum*; see Fig. 1a. This simplified model allows *blind* density-map reconstruction and estimation of the composite *mass-attenuation spectrum* $\iota(\kappa)$ in the case for which both the mass attenuation and incident spectrum are unknown. We develop a blind sparse density-map reconstruction scheme from measurements corrupted by Poisson noise, where the signal sparsity in the density-map domain is enforced using a total-variation (TV) norm penalty. The Poisson noise model is appropriate for measurements from photon-counting detectors and a good approximation for the more precise compound Poisson distribution for measurements from energy-integrating detectors [16, 17].

Although we focus on Poisson noise and gradient-map image sparsity in this paper, our framework is general and easy to adapt to, for example, lognormal noise and image sparsity in a 2D discrete wavelet transform (DWT) domain; see [12, 15].

We introduce the notation: I_N, $\mathbf{1}_{N\times 1}$, and $\mathbf{0}_{N\times 1}$ are the identity matrix of size N and the $N \times 1$ vectors of ones and zeros, respectively (replaced by I, $\mathbf{1}$, and $\mathbf{0}$ when the dimensions can be inferred easily); $|\cdot|$, $\|\cdot\|_p$, and "^{T}" are the absolute value, ℓ_p norm, and transpose, respectively. Denote by $\lceil x \rceil$ the smallest integer larger than or equal to $x \in \mathbb{R}$. For a vector $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_p]^T \in \mathbb{R}^p$, define the nonnegativity indicator function

$$\mathbb{I}_{[0,+\infty)}(\boldsymbol{\alpha}) \triangleq \begin{cases} 0, & \boldsymbol{a} \succeq \mathbf{0} \\ +\infty, & \text{otherwise} \end{cases} \tag{1}$$

where "$\succeq$" and "$\succ$"are the elementwise versions of "$\geq$" and "$>$", respectively. Furthermore, $\boldsymbol{a}^{\mathrm{L}}(s) \triangleq \int \boldsymbol{a}(\kappa) \mathrm{e}^{-s\kappa}\, \mathrm{d}\kappa$ is the *Laplace transform* of a vector function $\boldsymbol{a}(\kappa)$ and

$$\big((-\kappa)^m \boldsymbol{a}\big)^{\mathrm{L}}(s) = \int (-\kappa)^m \boldsymbol{a}(\kappa) \mathrm{e}^{-s\kappa}\, \mathrm{d}\kappa = \tfrac{\mathrm{d}^m \boldsymbol{a}^{\mathrm{L}}(s)}{\mathrm{d}s^m} \tag{2}$$

is the mth derivative of $\boldsymbol{a}^{\mathrm{L}}(s)$. Define also the set of nonnegative real numbers as $\mathbb{R}_+ = [0, +\infty)$, the elementwise logarithm $\ln_\circ \boldsymbol{x} = [\ln x_1, \ldots, \ln x_N]^T$ where $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]^T$, and Laplace transforms $\boldsymbol{a}^{\mathrm{L}}_\circ(\boldsymbol{s}) = \big(\boldsymbol{a}^{\mathrm{L}}(s_n)\big)_{n=1}^N$ and $(\kappa\boldsymbol{a})^{\mathrm{L}}_\circ(\boldsymbol{s}) = \big((\kappa\boldsymbol{a})^{\mathrm{L}}(s_n)\big)_{n=1}^N$ obtained by stacking $\boldsymbol{a}^{\mathrm{L}}(s_n)$ and $(\kappa\boldsymbol{a})^{\mathrm{L}}(s_n)$ columnwise, where $\boldsymbol{s} = [s_1, s_2, \ldots, s_N]^T$. We define the proximal operator for function $r(\boldsymbol{\alpha})$ scaled by λ [18]:

$$\operatorname{prox}_{\lambda r}\, \boldsymbol{a} = \arg\min_{\boldsymbol{\alpha}} \tfrac{1}{2}\|\boldsymbol{\alpha} - \boldsymbol{a}\|_2^2 + \lambda r(\boldsymbol{\alpha}). \tag{3}$$

Finally, $\operatorname{supp}(\iota(\cdot))$ is the support set of a function $\iota(\cdot)$, $\operatorname{dom}(f) = \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) < +\infty\}$ is the domain of function $f(\cdot)$, and $\operatorname{diag}(\boldsymbol{x})$ is the diagonal matrix with diagonal elements defined by the corresponding elements of vector $\boldsymbol{x}$.

A. Polychromatic X-ray CT Model

We review the standard noiseless polychromatic X-ray CT measurement model.

Assume that the incident intensity $\mathcal{I}^{\mathrm{in}}$ of a polychromatic X-ray source spreads along photon energy ε following the density $\iota(\varepsilon) \geq 0$:

$$\mathcal{I}^{\mathrm{in}} = \int \iota(\varepsilon)\, \mathrm{d}\varepsilon; \tag{4a}$$

see Fig. 1a, which shows a typical $\iota(\varepsilon)$. The noiseless measurement collected by an energy-integrating detector upon traversing a straight line $\ell = \ell(x, y)$ in a Cartesian coordinate system has the superposition-integral form [3, Ch. 4.1], [5, Sec. 6]:

$$\begin{aligned} \mathcal{I}^{\mathrm{out}} &= \int \iota(\varepsilon) \exp\left[-\int_\ell \mu(x, y, \varepsilon)\, \mathrm{d}\ell\right] \mathrm{d}\varepsilon \\ &= \int \iota(\varepsilon) \exp\left[-\kappa(\varepsilon) \int_\ell \alpha(x, y)\, \mathrm{d}\ell\right] \mathrm{d}\varepsilon, \end{aligned} \tag{4b}$$

where we model the attenuation coefficients $\mu(x, y, \varepsilon)$ of the inspected object consisting of a *single* material using the following *separable form* [5, Sec. 6]:

$$\mu(x, y, \varepsilon) = \kappa(\varepsilon)\alpha(x, y). \tag{5}$$

Here, $\kappa(\varepsilon) > 0$ is the mass-attenuation coefficient of the material, a function of the photon energy ε (illustrated in Fig. 1a), and $\alpha(x, y) \geq 0$ is the density-map of the object. For a monochromatic source at photon energy ε, $\ln[\mathcal{I}^{\mathrm{in}}(\varepsilon)/\mathcal{I}^{\mathrm{out}}(\varepsilon)]$ is a linear function of $\alpha(x, y)$, which is a basis for traditional linear reconstruction. However, X-rays generated by vacuum tubes are not monochromatic [3, 4], and we cannot transform the underlying noiseless measurements to a linear model unless we know perfectly the incident energy spectrum $\iota(\varepsilon)$ and mass attenuation of the inspected material $\kappa(\varepsilon)$.

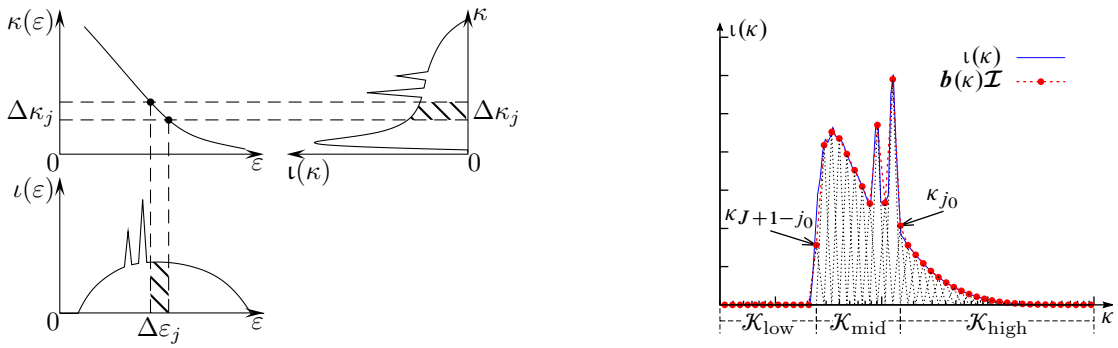
In Section II, we introduce our parsimonious parameterization of the measurement model (4b) tailored for signal reconstruction. In Section III, we define the parameters to be estimated and discuss their identifiability. Section IV presents the measurement model and establishes biconvexity of the underlying negative log-likelihood (NLL) function with respect to the density-map and mass-attenuation parameters. Section V introduces the penalized NLL function that incorporates the parameter constraints, establishes its properties, and describes a block coordinate-descent algorithm for its minimization. In Section VI, we show the performance of the proposed method using simulated and real X-ray CT data. Concluding remarks are given in Section VII.

II. Mass-Attenuation Parameterization

Since the mass attenuation $\kappa(\varepsilon)$ and incident spectrum density $\iota(\varepsilon)$ are both functions of ε (see Fig. 1a), we combine the variations of these two functions and write (4a) and (4b) as integrals of κ rather than ε, seeking to represent our model using two functions $\iota(\kappa)$ (defined below) and $\alpha(x, y)$ instead of three $(\iota(\varepsilon), \kappa(\varepsilon)$, and $\alpha(x, y))$; see also [11]. Hence, we rewrite (4a) and (4b) as (see Appendix A)

$$\mathcal{I}^{\mathrm{in}} = \iota^{\mathrm{L}}(0) \tag{6a}$$

$$\mathcal{I}^{\mathrm{out}} = \iota^{\mathrm{L}}\left(\int_\ell \alpha(x, y)\, \mathrm{d}\ell\right), \tag{6b}$$



(a) (b)

Fig. 1: (a) Mass-attenuation spectrum $\iota(\kappa)$ obtained by combining the mass attenuation $\kappa(\varepsilon)$ and incident spectrum $\iota(\varepsilon)$ and (b) its B1-spline expansion, with κ-axis in log scale.

where $\iota^{\mathrm{L}}(s) = \int \iota(\kappa) \mathrm{e}^{-s\kappa} \, \mathrm{d}\kappa$ is the Laplace transform of the mass-attenuation spectrum $\iota(\kappa)$, which represents the density of the incident X-ray energy at attenuation κ; here, $s > 0$, in contrast with the traditional Laplace transform where s is generally complex. For invertible $\kappa(\varepsilon)$ with differentiable inverse function $\varepsilon(\kappa)$,

$$\iota(\kappa) \triangleq \iota(\varepsilon(\kappa)) |\varepsilon'(\kappa)| \geq 0 \tag{7}$$

with $\varepsilon'(\kappa) = \mathrm{d}\varepsilon(\kappa) / \mathrm{d}\kappa$. In Fig. 1a, the area $\iota(\varepsilon_j)\Delta\varepsilon_j$ depicting the X-ray energy within the $\Delta\varepsilon_j$ slot is the same as area $\iota(\kappa_j)\Delta\kappa_j$, the amount of X-ray energy attenuated within the corresponding $\Delta\kappa_j$ slot. In Appendix A, we generalize (7) to non-invertible $\kappa(\varepsilon)$ with K-edges.

The mass-attenuation spectrum $\iota(\kappa)$ is nonnegative for all κ; see (7) and its generalization (A1) in Appendix A. Due to its nonnegative support and range, $\iota^{\mathrm{L}}(s)$ is a decreasing function of s. Here, $s > 0$, in contrast with the traditional Laplace transform where s is generally complex. The function $(\iota^{\mathrm{L}})^{-1}$ maps the noiseless measurement $\mathcal{I}^{\mathrm{out}}$ in (6), which is a nonlinear function of the density-map $\alpha(x, y)$, into a noiseless linear "measurement" $\int_\ell \alpha(x, y) \, \mathrm{d}\ell$. The $(\iota^{\mathrm{L}})^{-1} \circ \exp(-\cdot)$ mapping corresponds to the *linearization function* in [19] (where it was defined through (4b) rather than the mass-attenuation spectrum) and converts $-\ln \mathcal{I}^{\mathrm{out}}$ into a noiseless linear "measurement" $\int_\ell \alpha(x, y) \, \mathrm{d}\ell$.

The mass-attenuation spectrum depends on the measurement system (through the incident energy spectrum) and inspected object (through the mass attenuation of the inspected material). In the blind scenario with unknown inspected material and incident signal spectrum, parameterization (6) allows us to estimate two functions: $\iota(\kappa)$ and $\alpha(x, y)$ rather than three: $\iota(\varepsilon), \kappa(\varepsilon)$, and $\alpha(x, y)$. This blind scenario is the focus of this paper.

III. Discrete Parameter Definition and Ambiguity

We first define the discrete density map and mass-attenuation spectrum parameters and then discuss their identifiability.

A. Density-map discretization and mass-attenuation spectrum basis-function expansion

Upon spatial-domain discretization into p pixels, approximate the integral $\int_\ell \alpha(x, y) \, \mathrm{d}\ell$ with $\boldsymbol{\phi}^T \boldsymbol{\alpha}$:

$$\int_\ell \alpha(x, y) \, \mathrm{d}\ell = \boldsymbol{\phi}^T \boldsymbol{\alpha}, \tag{8}$$

where $\boldsymbol{\alpha} \succeq \mathbf{0}$ is a $p \times 1$ vector representing the 2D image that we wish to reconstruct and $\boldsymbol{\phi} \succeq \mathbf{0}$ is a $p \times 1$ vector of known weights quantifying how much each element of $\boldsymbol{\alpha}$ contributes to the X-ray attenuation on the straight-line path ℓ. An X-ray CT scan consists of hundreds of projections with the beam intensity measured by thousands of detectors for each projection. Denote by N the total number of measurements from all projections collected at the detector array. For the nth measurement, define its discretized line integral as $\boldsymbol{\phi}_n^T \boldsymbol{\alpha}$. Stacking all N such integrals into a vector yields $\Phi\boldsymbol{\alpha}$, where

$$\Phi = \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_N \end{bmatrix}^T \in \mathbb{R}^{N \times p} \tag{9}$$

is the *projection matrix*, also known as the Radon transform matrix in a parallel-beam X-ray tomographic imaging system. We call the corresponding transformation, $\Phi\boldsymbol{\alpha}$, the *monochromatic projection* of $\boldsymbol{\alpha}$.

Approximate $\iota(\kappa)$ with a linear combination of J ($J \ll N$) basis functions:

$$\iota(\kappa) = \boldsymbol{b}(\kappa)\boldsymbol{\mathcal{I}}, \tag{10a}$$

where

$$\boldsymbol{\mathcal{I}} \triangleq [\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_J]^T \succeq \mathbf{0} \tag{10b}$$

is the $J \times 1$ vector of corresponding basis-function coefficients, and the $1 \times J$ row-vector function

$$\boldsymbol{b}(\kappa) \triangleq \left[b_1(\kappa), b_2(\kappa), \ldots, b_J(\kappa) \right] \tag{11}$$

consists of B-splines [20] of order one (termed B1 splines, illustrated in Fig. 1b). In this case, the decomposition (10a) yields nonnegative elements of the spline coefficients $\boldsymbol{\mathcal{I}}$ (based on (7)) and thus allows us to impose the physically meaningful nonnegativity constraint (10b) when estimating $\boldsymbol{\mathcal{I}}$. Substituting (8) and (10a) into (6a)–(6b) for each of the N measurements

yields the following expressions for the incident energy and the $N \times 1$ vector of noiseless measurements:

$$\mathcal{I}^{\text{in}}(\boldsymbol{\mathcal{I}}) = \boldsymbol{b}^{\text{L}}(0)\boldsymbol{\mathcal{I}} \tag{12a}$$

$$\boldsymbol{\mathcal{I}}^{\text{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) = \boldsymbol{b}^{\text{L}}_{\circ}(\Phi\boldsymbol{\alpha})\boldsymbol{\mathcal{I}} \tag{12b}$$

where, following the notation introduced in Section I, $\boldsymbol{b}^{\text{L}}_{\circ}(\boldsymbol{s}) = \left(\boldsymbol{b}^{\text{L}}(s_n)\right)_{n=1}^{N}$ is an *output basis-function matrix* obtained by stacking the $1 \times J$ vectors $\boldsymbol{b}^{\text{L}}(s_n)$ columnwise, and $\boldsymbol{s} = \Phi\boldsymbol{\alpha}$ is the monochromatic projection. Since the Laplace transform of (11) (see also (13b)) can be computed analytically, $\boldsymbol{b}^{\text{L}}(s)$ has a closed-form expression.

1) Spline selection: We select the spline knots from a growing geometric series $(\kappa_j)_{j=0}^{J+1}$ with $\kappa_0 > 0$:

$$\kappa_j = q^j \kappa_0 \tag{13a}$$

and common ratio $q > 1$, which yields the B1 splines

$$b_j(\kappa) = \begin{cases} \frac{\kappa - \kappa_{j-1}}{\kappa_j - \kappa_{j-1}}, & \kappa_{j-1} \le \kappa < \kappa_j \\ \frac{-\kappa + \kappa_{j+1}}{\kappa_{j+1} - \kappa_j}, & \kappa_j \le \kappa < \kappa_{j+1} \\ 0, & \text{otherwise} \end{cases} \tag{13b}$$

that satisfy the q-scaling property:

$$b_j(\kappa) = b_{j+1}(q\kappa) \tag{13c}$$

see also Fig. 1b. The geometric-series knots (13a) appear uniformly spaced in Fig. 1b because the κ-axis in this figure is shown in the log scale. When computing $b_j^{\text{L}}(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})$, larger j implies exponentially smaller $\text{e}^{-\boldsymbol{\phi}_n^T\boldsymbol{\alpha}\kappa}$ terms within the integral range $[\kappa_{j-1}, \kappa_{j+1})$. The geometric-series knot selection (13a) compensates for larger j with a geometrically wider integral range $[\kappa_{j-1}, \kappa_{j+1})$, which results in a more effective approximation of (6). In particular, this knot selection leads to $\left(b_j^{\text{L}}(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})\right)_{j=1}^{J}$ with similar values for different values of j, which allows us to balance the weight of each $\left(\mathcal{I}_j\right)_{j=1}^{J}$ in $\boldsymbol{b}^{\text{L}}(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})\boldsymbol{\mathcal{I}}$. Furthermore, the geometric-series knots (13a) span a range from κ_0 to κ_{J+1}, which can be made wide with a moderate number of knots J.

The common ratio q determines the resolution of the B1-spline approximation. Here, we select q and J so that the range of κ spanning the mass-attenuation spectrum is constant:

$$\frac{\kappa_{J+1}}{\kappa_0} = q^{J+1} = \text{const.} \tag{13d}$$

In summary, the following three tuning constants define our B1-spline basis functions $\boldsymbol{b}(\kappa)$:

$$(q, \kappa_0, J). \tag{13e}$$

B. Density-map and mass-attenuation spectrum ambiguities

By noting (13c) and the κ-scaling property of the Laplace transform,

$$b_j(q\kappa) \xrightarrow{\mathcal{L}} \frac{1}{q} b_j^{\text{L}}\left(\frac{s}{q}\right), \qquad q > 0 \tag{14}$$

we conclude that selecting basis functions $\left[b_0(\kappa), b_1(\kappa), \ldots, b_{J-1}(\kappa)\right]$ that are q times narrower than those in $\boldsymbol{b}(\kappa)$ and density-map and spectral parameters q times larger than $\boldsymbol{\alpha}$ and $\boldsymbol{\mathcal{I}}$: $q\boldsymbol{\alpha}$ and $q\boldsymbol{\mathcal{I}}$, yields the same mean output photon energy. Consequently,

$$\boldsymbol{\mathcal{I}}^{\text{out}}\left(\boldsymbol{\alpha}, [0, \mathcal{I}_2, \ldots, \mathcal{I}_J]^T\right) = \boldsymbol{\mathcal{I}}^{\text{out}}\left(q\boldsymbol{\alpha}, q[\mathcal{I}_2, \ldots, \mathcal{I}_J, 0]^T\right). \tag{15}$$

We refer to this property as the *shift ambiguity* of the mass-attenuation spectrum, which allows us to rearrange leading or trailing zeros in the mass-attenuation coefficient vector $\boldsymbol{\mathcal{I}}$ and position the central nonzero part of $\boldsymbol{\mathcal{I}}$.

C. Rank of $\boldsymbol{b}^{\text{L}}_{\circ}(\Phi\boldsymbol{\alpha})$ and selection of the number of splines J

If $\boldsymbol{b}^{\text{L}}_{\circ}(\Phi\boldsymbol{\alpha})$ does not have full column rank, then $\boldsymbol{\mathcal{I}}$ is not identifiable even if $\boldsymbol{\alpha}$ is known; see (12b). The estimation of $\boldsymbol{\mathcal{I}}$ may be numerically unstable if $\boldsymbol{b}^{\text{L}}_{\circ}(\Phi\boldsymbol{\alpha})$ is poorly conditioned and has small minimum singular values. We can think of the noiseless X-ray CT measurements as $\boldsymbol{b}^{\text{L}}(s)\boldsymbol{\mathcal{I}}$ sampled at different $s = \boldsymbol{\phi}_n^T\boldsymbol{\alpha} \in \left[0, \max_n(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})\right]$. If we could collect all $s \in [0, a]$, $a > 0$ (denoted $\boldsymbol{s}$), the corresponding $\boldsymbol{b}^{\text{L}}_{\circ}(\boldsymbol{s})$ would be a full-rank matrix; see [12, Lemma 1 in Sec. II-B]. If our data collection system can sample over $\left[0, \max_n(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})\right]$ sufficiently densely, we expect $\boldsymbol{b}^{\text{L}}_{\circ}(\Phi\boldsymbol{\alpha})$ to have full column rank.

As the number of splines J increases for fixed support $[\kappa_0, \kappa_{J+1}]$ (see (13d)), we achieve better resolution of the mass-attenuation spectrum, but $\boldsymbol{b}^{\text{L}}_{\circ}(\Phi\boldsymbol{\alpha})$ becomes poorly conditioned with its smallest singular values approaching zero. To estimate this spectrum well, we should choose a J that provides both good resolution *and* sufficiently large smallest singular value of $\boldsymbol{b}^{\text{L}}_{\circ}(\Phi\boldsymbol{\alpha})$. Fortunately, we focus on the reconstruction of $\boldsymbol{\alpha}$, which is affected by $\boldsymbol{\mathcal{I}}$ only through the function $\boldsymbol{b}^{\text{L}}(s)\boldsymbol{\mathcal{I}}$, and $\boldsymbol{b}^{\text{L}}(s)\boldsymbol{\mathcal{I}}$ is stable as we increase J. Indeed, we observe that when we choose a J significantly larger than the rank of $\boldsymbol{b}^{\text{L}}_{\circ}(\Phi\boldsymbol{\alpha})$, the estimation of $\boldsymbol{\alpha}$ will be good and $\boldsymbol{b}^{\text{L}}(s)\boldsymbol{\mathcal{I}}$ stable, even though the estimation of $\boldsymbol{\mathcal{I}}$ is poor due to its non-identifiability. The increase of J will also increase the computational complexity of signal reconstruction under the blind scenario for which the mass-attenuation spectrum is unknown.

IV. Measurement Model and Its Properties

For an $N \times 1$ vector $\boldsymbol{\mathcal{E}}$ of independent Poisson measurements, the NLL in the form of generalized Kullback-Leibler divergence [21] is (see also (12b))

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) = \mathbf{1}^T\left[\boldsymbol{\mathcal{I}}^{\text{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) - \boldsymbol{\mathcal{E}}\right] - \sum_{n, \mathcal{E}_n \neq 0} \mathcal{E}_n \ln \frac{\mathcal{I}_n^{\text{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})}{\mathcal{E}_n}. \tag{16}$$

In the following, we express the NLL (16) as a function of $\boldsymbol{\alpha}$ with $\boldsymbol{\mathcal{I}}$ fixed and vice versa, and derive conditions for its convexity under the two scenarios. These results will then be used to establish biconvexity conditions for this NLL.

NLL of $\boldsymbol{\alpha}$. Recall (10a) and define

$$\iota^{\text{L}}_{\circ}(\Phi\boldsymbol{\alpha}) = \boldsymbol{b}^{\text{L}}_{\circ}(\Phi\boldsymbol{\alpha})\boldsymbol{\mathcal{I}} \tag{17}$$

obtained by stacking $\left(\iota^{\text{L}}(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})\right)_{n=1}^{N}$ columnwise. The NLL of $\boldsymbol{\alpha}$ for fixed $\boldsymbol{\mathcal{I}}$ is

$$\mathcal{L}_{\iota}(\boldsymbol{\alpha}) = \mathbf{1}^T\left[\iota^{\text{L}}_{\circ}(\Phi\boldsymbol{\alpha}) - \boldsymbol{\mathcal{E}}\right] - \sum_{n, \mathcal{E}_n \neq 0} \mathcal{E}_n \ln \frac{\iota^{\text{L}}(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})}{\mathcal{E}_n}, \tag{18}$$

which corresponds to the Poisson generalized linear model (GLM) with design matrix Φ and link function equal to the inverse of $\iota^{\mathrm{L}}(\cdot)$. See [22] for an introduction to GLMs.

To establish convexity of the NLL (18), we enforce monotonicity of the mass-attenuation spectrum $\iota(\kappa)$ in low- and high-κ regions and also assume that the mid-κ region has higher spectrum than the low-κ region. Note that we do not require here that $\iota(\kappa)$ satisfy the basis-function expansion (10a); however, (10a) will be needed to establish the biconvexity of the NLL in (16). Hence, we define the three κ regions using the spline parameters (13e) as well as an additional integer constant

$$j_0 \geq \lceil (J+1)/2 \rceil. \tag{19a}$$

In particular, κ_{J+1-j_0} and κ_{j_0} partition the range $[\kappa_0, \kappa_{J+1}]$ into the low-, mid-, and high-κ regions: $\mathcal{K}_{\mathrm{low}}$, $\mathcal{K}_{\mathrm{mid}}$, and $\mathcal{K}_{\mathrm{high}}$, respectively, see Fig. 1b.

Assumption 1: The mass-attenuation spectrum satisfies

$$\begin{aligned}\mathcal{A} = \Big\{ \iota : \ [\kappa_0, \kappa_{J+1}] \to \mathbb{R}_+ \,\Big|\, & \iota \text{ non-decreasing in } \mathcal{K}_{\mathrm{low}}, \\ & \text{non-increasing in } \mathcal{K}_{\mathrm{high}}, \text{ and} \\ & \iota(\kappa) \geq \iota\big(\kappa_{J+1-j_0}\big) \ \forall \kappa \in \mathcal{K}_{\mathrm{mid}} \Big\}. \end{aligned} \tag{20a}$$

If the basis-function expansion (10a) holds, (20a) reduces to

$$\begin{aligned}\mathcal{A} = \Big\{ \boldsymbol{\mathcal{I}} \in \mathbb{R}_+^J \,\Big|\, & \mathcal{I}_1 \leq \mathcal{I}_2 \leq \cdots \leq \mathcal{I}_{J+1-j_0}, \ \mathcal{I}_{j_0} \geq \cdots \geq \mathcal{I}_J, \\ & \text{and } \mathcal{I}_j \geq \mathcal{I}_{J+1-j_0}, \ \forall\, j \in [J+1-j_0, j_0] \Big\}. \end{aligned} \tag{20b}$$

Here, the monotonic low- and high-κ regions each contain $J - j_0$ knots, whereas the central region contains $2j_0 - J$ knots in the B1-spline representation.

In practice, the X-ray spectrum $\iota(\varepsilon)$ starts at the lowest effective energy that can penetrate the object, vanishes at the tube voltage (the highest photon energy), and has a region in the center higher than the two ends; see Fig. 1a. When the support of $\iota(\varepsilon)$ is free of K-edges (see the discussion in Appendix A), the mass-attenuation coefficient $\kappa(\varepsilon)$ is a monotonic function of ε; thus $\iota(\kappa)$ as a function of κ has similar shape as $\iota(\varepsilon)$ as a function of ε, which justifies Assumption 1. If a K-edge is present within the support of $\iota(\varepsilon)$, it is difficult to infer the shape of $\iota(\kappa)$. In most cases, Assumption 1 holds.

For the approximation of $\iota(\kappa)$ using a B1-spline basis expansion, as long as $[\kappa_0, \kappa_{J+1}]$ is sufficiently large to cover the range of $\kappa(\varepsilon)$ with $\varepsilon \in \mathrm{supp}(\iota(\varepsilon))$, we can always meet Assumption 1 by the appropriate selection of j_0.

Multiple different $(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ share the same noiseless output $\boldsymbol{\mathcal{I}}^{\mathrm{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ and thus the same NLL; see Section III-B. In particular, equivalent $(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ can be constructed by left- or right-shifting the mass attenuation spectrum and properly rescaling it and the density-map; see (15).

Lemma 1: Provided that Assumption 1 holds, the Poisson NLL $\mathcal{L}_\iota(\boldsymbol{\alpha})$ is a convex function of $\boldsymbol{\alpha}$ over the following region:

$$\left\{ \boldsymbol{\alpha} \,\middle|\, \iota_\circ^{\mathrm{L}}(\Phi\boldsymbol{\alpha}) \succeq (1-V)\boldsymbol{\mathcal{E}}, \ \boldsymbol{\alpha} \in \mathbb{R}_+^p \right\} \tag{21a}$$

where

$$V \triangleq \frac{2q^{j_0}}{q^{2j_0}+1}. \tag{21b}$$

Proof: See Appendix B. ∎

Note that (21a) is only a subset of the region where $\mathcal{L}_\iota(\boldsymbol{\alpha})$ is convex and that Lemma 1 does not assume a basis-function expansion of the mass-attenuation spectrum, only that it satisfies (20a).

The condition in (21a) corresponds to lower-bounding $\mathcal{I}_n^{\mathrm{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) \big/ \mathcal{E}_n$ by $1 - V$ for all n. The constant V is a function of q^{j_0}, which is the ratio of the point where $\iota(\kappa)$ starts to be monotonically decreasing to the point where the support of $\iota(\kappa)$ starts; see Fig. 1b.

NLL of $\boldsymbol{\mathcal{I}}$. The NLL of $\boldsymbol{\mathcal{I}}$ for fixed $\boldsymbol{\alpha}$ reduces to a Poisson GLM with design matrix

$$A = \boldsymbol{b}_\circ^{\mathrm{L}}(\Phi\boldsymbol{\alpha}) \tag{22a}$$

all of whose elements are positive, and the identity link function:

$$\mathcal{L}_A(\boldsymbol{\mathcal{I}}) = \mathbf{1}^T(A\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{E}}) - \sum_{n, \mathcal{E}_n \neq 0} \mathcal{E}_n \ln \frac{[A\boldsymbol{\mathcal{I}}]_n}{\mathcal{E}_n}. \tag{22b}$$

We now prove the convexity of $\mathcal{L}_A(\boldsymbol{\mathcal{I}})$.

Lemma 2: The NLL $\mathcal{L}_A(\boldsymbol{\mathcal{I}})$ in (22b) is a convex function of $\boldsymbol{\mathcal{I}}$ for all $\boldsymbol{\mathcal{I}} \in \mathbb{R}_+^J$.

Proof: The Hessian of the NLL in (22b)

$$\frac{\partial^2 \mathcal{L}_A(\boldsymbol{\mathcal{I}})}{\partial \boldsymbol{\mathcal{I}} \partial \boldsymbol{\mathcal{I}}^T} = A^T \operatorname{diag}(\boldsymbol{\mathcal{E}}) \operatorname{diag}^{-2}(A\boldsymbol{\mathcal{I}}) A \tag{23}$$

is positive semidefinite. Thus, $\mathcal{L}_A(\boldsymbol{\mathcal{I}})$ is convex on $\mathbb{R}_+^J$. ∎

The Hessian expression in (23) implies that $\mathcal{L}_A(\boldsymbol{\mathcal{I}})$ in (22b) is strongly convex if the design matrix A has full rank. Combining the convexity results in Lemmas 1 and 2 yields the biconvexity region for the NLL $\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ in (16).

Theorem 1 (Biconvexity of the NLL): Suppose that Assumption 1 in (20b) holds. Then, the Poisson NLL (16) is biconvex [23] with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\mathcal{I}}$ in the following set:

$$\mathcal{P} = \left\{ (\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) \,\middle|\, \boldsymbol{\mathcal{I}}^{\mathrm{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) \succeq (1-V)\boldsymbol{\mathcal{E}}, \boldsymbol{\mathcal{I}} \in \mathcal{A}, \boldsymbol{\alpha} \in \mathbb{R}_+^p \right\}, \tag{24}$$

which bounds $\mathcal{I}_n^{\mathrm{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) \big/ \mathcal{E}_n$ from below by $1 - V$ for all n; see also (21b).

Proof: We first show the convexity of $\mathcal{P}$ with respect to each variable ($\boldsymbol{\alpha}$ and $\boldsymbol{\mathcal{I}}$) with the other fixed. We then show the convexity of the NLL (16) for each variable.

Region $\mathcal{A}$ in (20b) is a subspace, thus a convex set. Since $\boldsymbol{\mathcal{I}}^{\mathrm{out}}$ in (12b) is a linear function of $\boldsymbol{\mathcal{I}}$, the inequalities comparing $\boldsymbol{\mathcal{I}}^{\mathrm{out}}$ to constants specify a convex set. Therefore, $\mathcal{P}_{\boldsymbol{\alpha}} = \left\{ \boldsymbol{\mathcal{I}} \,|\, (\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) \in \mathcal{P} \right\}$ is convex for fixed $\boldsymbol{\alpha} \in \mathbb{R}_+^p$, for it is the intersection of the subspace $\mathcal{A}$ and a convex set via $\boldsymbol{\mathcal{I}}^{\mathrm{out}}$. Since $b_j(\kappa) \geq 0$, $\left(b_j^{\mathrm{L}}(s)\right)_{j=1}^J = \int_{\kappa_{j-1}}^{\kappa_j+1} b_j(\kappa) \mathrm{e}^{-s\kappa} \, \mathrm{d}\kappa$ are decreasing functions of s, which, together with the fact that $\boldsymbol{\mathcal{I}} \succeq \mathbf{0}$, implies that $\boldsymbol{b}^{\mathrm{L}}(\boldsymbol{s})\boldsymbol{\mathcal{I}}$ is a decreasing function of $\boldsymbol{s}$. Since the linear transform $\Phi\boldsymbol{\alpha}$ preserves convexity, $\mathcal{P}_{\boldsymbol{\mathcal{I}}} = \left\{ \boldsymbol{\alpha} \,|\, (\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) \in \mathcal{P} \right\}$ is convex with respect to $\boldsymbol{\alpha}$ for fixed $\boldsymbol{\mathcal{I}} \in \mathcal{A}$. Therefore, $\mathcal{P}$ is biconvex with respect to $\boldsymbol{\mathcal{I}}$ and $\boldsymbol{\alpha}$.

Observe that $\mathcal{P}$ in (24) is the intersection of the regions specified by Assumption 1 and Lemmas 1 and 2. Thus, within $\mathcal{P}$, the Poisson NLL (16) is a convex function of $\boldsymbol{\alpha}$ for fixed $\boldsymbol{\mathcal{I}}$ and a convex function of $\boldsymbol{\mathcal{I}}$ for fixed $\boldsymbol{\alpha}$, respectively.

By combining the above region and function convexity results, we conclude that (16) is biconvex within $\mathcal{P}$. ∎

In [12], we establish conditions for biconvexity of the NLL under the lognormal noise model.

V. Parameter Estimation

Our goal is to compute penalized maximum-likelihood estimates of the density-map and mass-attenuation spectrum parameters $(\boldsymbol{\alpha}, \mathcal{I})$ by solving the following minimization problem:

$$\min_{\boldsymbol{\alpha}, \mathcal{I}} f(\boldsymbol{\alpha}, \mathcal{I}) \tag{25a}$$

where

$$f(\boldsymbol{\alpha}, \mathcal{I}) = \mathcal{L}(\boldsymbol{\alpha}, \mathcal{I}) + u r(\boldsymbol{\alpha}) + \mathbb{I}_{[0,+\infty)}(\mathcal{I}) \tag{25b}$$

$$r(\boldsymbol{\alpha}) = \sum_{i=1}^{p} \sqrt{\sum_{j \in \mathcal{N}_i} (\alpha_i - \alpha_j)^2} + \mathbb{I}_{[0,+\infty)}(\boldsymbol{\alpha}) \tag{25c}$$

are the penalized NLL objective function and the density-map regularization term that enforces nonnegativity and sparsity of the image $\boldsymbol{\alpha}$; $u > 0$ is a scalar tuning constant. We impose the nonnegativity of the mass-attenuation coefficients (10b) using the indicator-function term in (25b). Here, $\mathcal{N}_i$ is the index set of neighbors of α_i, where the elements of $\boldsymbol{\alpha}$ are arranged to form a 2D image: Each set $\mathcal{N}_i$ consists of two pixels at most, with one on the top and the other on the right of the ith pixel, if possible [24].

A. Properties of the objective function $f(\boldsymbol{\alpha}, \mathcal{I})$

Since $r(\boldsymbol{\alpha})$ in (25c) and $\mathbb{I}_{[0,+\infty)}(\mathcal{I})$ in (25b) are convex functions of $\boldsymbol{\alpha}$ and $\mathcal{I}$ for all $\boldsymbol{\alpha} \succeq \mathbf{0}$ and $\mathcal{I} \succeq \mathbf{0}$, the following holds:

Corollary 1: The objective $f(\boldsymbol{\alpha}, \mathcal{I})$ in (25b) is biconvex with respect to $\boldsymbol{\alpha}$ and $\mathcal{I}$ under the conditions specified by Theorem 1.

Although the NLL may have multiple local minima of the form $q^{\ell}\hat{\boldsymbol{\alpha}}$ with integer ℓ (see Section III-B), those with large ℓ can be eliminated by the regularization penalty, see the discussion in [12, Sec. IV-A].

We now show that the objective function (25b) satisfies the Kurdyka-Łojasiewicz (KL) property [25], which is important for establishing local convergence of block-coordinate schemes in biconvex optimization problems. The KL property [25] regularizes the (sub)gradient of a function through its value at a certain point or over the whole domain and also ensures the steepness of the function around the optimum so that the length of the gradient trajectory is bounded.

Theorem 2 (KL Property): The objective function $f(\boldsymbol{\alpha}, \mathcal{I})$ satisfies the KL property in any compact subset $\mathbb{C} \subseteq \mathrm{dom}(f)$.

Proof: See Appendix C. ∎

Note that all $(\boldsymbol{\alpha}, \mathcal{I})$ that lead to positive noiseless measurements, i.e. $\mathcal{I}^{\mathrm{out}}(\boldsymbol{\alpha}, \mathcal{I}) \succ \mathbf{0}$, are in the domain of f, which excludes the case $\mathcal{I} = \mathbf{0}$ when no incident X-ray is applied; see also (12b).

B. Minimization algorithm

The parameters that we wish to estimate are naturally divided into two blocks, $\boldsymbol{\alpha}$ and $\mathcal{I}$. The large size of $\boldsymbol{\alpha}$ prohibits effective second-order methods under sparsity regularization, whereas $\mathcal{I}$ has much smaller size and only nonnegative constraints, thus allowing for more sophisticated solvers, such as the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) approach [26, Sec. 4.3.3.4] that we adopt here. In addition, the scaling difference between $\boldsymbol{\alpha}$ and $\mathcal{I}$ can be significant, so that the joint gradient method for $\boldsymbol{\alpha}$ and $\mathcal{I}$ together would converge slowly. Therefore, we adopt a block coordinate-descent algorithm to minimize $f(\boldsymbol{\alpha}, \mathcal{I})$ in (25b), where the Nesterov's proximal-gradient (NPG) [27, 28] and limited-memory BFGS with box constraints (L-BFGS-B) [29] methods are employed to update estimates of the density-map and mass-attenuation spectrum parameters, respectively. The choice of block coordinate-descent optimization is also motivated by the related alternate convex search (ACS) and block coordinate-descent schemes in [23] and [30], respectively, both with convergence guarantees under certain conditions.

We minimize the objective function (25b) by alternatively updating $\boldsymbol{\alpha}$ and $\mathcal{I}$ using Steps 1 and 2, respectively, where Iteration i proceeds as follows:

1) (NPG) Set the mass-attenuation spectrum $\iota(\kappa) = \boldsymbol{b}(\kappa)\mathcal{I}^{(i-1)}$, treat it as known[1], and descend the regularized NLL function $f(\boldsymbol{\alpha}, \mathcal{I}^{(i-1)}) = \mathcal{L}_{\iota}(\boldsymbol{\alpha}) + u r(\boldsymbol{\alpha})$ by applying an *NPG step* for $\boldsymbol{\alpha}$, which yields $\boldsymbol{\alpha}^{(i)}$:

$$\theta^{(i)} = \frac{1}{2}\left[1 + \sqrt{1 + 4\left(\theta^{(i-1)}\right)^2}\right] \tag{26a}$$

$$\overline{\boldsymbol{\alpha}}^{(i)} = \boldsymbol{\alpha}^{(i-1)} + \frac{\theta^{(i-1)} - 1}{\theta^{(i)}}\left(\boldsymbol{\alpha}^{(i-1)} - \boldsymbol{\alpha}^{(i-2)}\right) \tag{26b}$$

$$\boldsymbol{\alpha}^{(i)} = \mathrm{prox}_{\beta^{(i)} u r}\left(\overline{\boldsymbol{\alpha}}^{(i)} - \beta^{(i)} \nabla \mathcal{L}_{\iota}\left(\overline{\boldsymbol{\alpha}}^{(i)}\right)\right) \tag{26c}$$

where the minimization (26c) is computed using an inner iteration that employs the TV-based denoising method in [24, Sec. IV], and $\beta^{(i)} > 0$ is an adaptive step size chosen to satisfy the *majorization condition*:

$$\mathcal{L}_{\iota}\left(\boldsymbol{\alpha}^{(i)}\right) \le \mathcal{L}_{\iota}\left(\overline{\boldsymbol{\alpha}}^{(i)}\right) + \left(\boldsymbol{\alpha}^{(i)} - \overline{\boldsymbol{\alpha}}^{(i)}\right)^T \nabla \mathcal{L}_{\iota}\left(\overline{\boldsymbol{\alpha}}^{(i)}\right) + \frac{1}{2\beta^{(i)}}\left\|\boldsymbol{\alpha}^{(i)} - \overline{\boldsymbol{\alpha}}^{(i)}\right\|_2^2 \tag{26d}$$

using an adaptation scheme [31] that aims at finding the largest $\beta^{(i)}$ that satisfies (26d):

i) • if there have been no step-size backtracking events or increase attempts for $\mathtt{m}$ consecutive iterations ($i - \mathtt{m}$ to $i-1$), start with a larger step size $\overline{\beta}^{(i)} = \beta^{(i-1)}/\xi$ where $\xi \in (0,1)$ is a *step-size adaptation parameter*;
   • otherwise start with $\overline{\beta}^{(i)} = \beta^{(i-1)}$;

ii) (backtracking search) select

$$\beta^{(i)} = \xi^{t_i}\overline{\beta}^{(i)} \tag{27a}$$

where $t_i \ge 0$ is the smallest integer such that (27a) satisfies the majorization condition (26d); *backtracking event* corresponds to $t_i > 0$.

We select the initial step size $\overline{\beta}^{(0)}$ using the Barzilai-Borwein (BB) method [32]. We also apply the *function restart* [33] to restore the monotonicity and improve convergence; see the following discussion.

2) (BFGS) Set the design matrix $A = \boldsymbol{b}_{\circ}^{\mathrm{L}}\left(\Phi\boldsymbol{\alpha}^{(i)}\right)$, treat it as known[2], and minimize the regularized NLL function

[1]This selection corresponds to $\mathcal{L}_{\iota}(\boldsymbol{\alpha}) = \mathcal{L}(\boldsymbol{\alpha}, \mathcal{I}^{(i-1)})$; see also (18).

[2]This selection corresponds to $\mathcal{L}_A(\mathcal{I}) = \mathcal{L}(\boldsymbol{\alpha}^{(i)}, \mathcal{I})$; see also (22b).

$f\left(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\mathcal{I}}\right)$ with respect to $\boldsymbol{\mathcal{I}}$; i.e.,

$$\boldsymbol{\mathcal{I}}^{(i)} = \arg\min_{\boldsymbol{\mathcal{I}} \geq \mathbf{0}} \mathcal{L}_A(\boldsymbol{\mathcal{I}}) \tag{28}$$

using the inner L-BFGS-B iteration, initialized by $\boldsymbol{\mathcal{I}}^{(i-1)}$.

Iterate between Steps 1 and 2 until the relative distance of consecutive iterates of the density map $\boldsymbol{\alpha}$ does not change significantly:

$$\left\|\boldsymbol{\alpha}^{(i)} - \boldsymbol{\alpha}^{(i-1)}\right\|_2 < \epsilon \left\|\boldsymbol{\alpha}^{(i)}\right\|_2, \tag{29}$$

where $\epsilon > 0$ is the convergence threshold. The convergence criteria for the inner TV-denoising and L-BFGS-B iterations in Steps 1 and 2 are chosen to trade off the accuracy and speed of the inner iterations and provide sufficiently accurate solutions to (26c) and (28); see [12, Sec. IV-B2] for details.

We refer to the iteration between Steps 1 and 2 as the *NPG-BFGS algorithm*: it is the first physical-model–based image reconstruction method for simultaneous *blind* sparse image reconstruction and mass-attenuation spectrum estimation from polychromatic measurements; see also our preliminary work in [11]. In [11], we approximated Laplace integrals with Riemann sums, used a smooth approximation of the nonnegativity penalties in (25c), and did not employ signal-sparsity regularization.

If the mass-attenuation spectrum $\iota(\kappa)$ is *known* and we iterate Step 1 only to estimate the density-map image $\boldsymbol{\alpha}$, we refer to this iteration as the *NPG algorithm (known $\iota(\kappa)$)*.

If we *do not* apply the Nesterov's acceleration (26a)–(26b) and use only the proximal-gradient (PG) step (26c) to update the density-map iterates $\boldsymbol{\alpha}$, i.e., assign (31c) instead of (26b) *in every iteration*, we refer to the corresponding iteration as the *PG-BFGS algorithm*.

Scale-and-shift adjustment of the NPG-BFGS and PG-BFGS estimates. Denote by $\widehat{\boldsymbol{\mathcal{I}}}$ and $\widehat{\boldsymbol{\alpha}}$ the mass-attenuation spectrum parameter and density-map image estimates upon convergence of the NPG-BFGS iteration. To emphasize the dependence of the objective function (25b) on u, we denote it here by $f_u(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$. If the last element $\widehat{\mathcal{I}}_J$ of $\widehat{\boldsymbol{\mathcal{I}}}$ is zero, we can trivially improve this objective function by using the shift ambiguity: remove this zero element by circularly shifting $\widehat{\boldsymbol{\mathcal{I}}}$ and divide $\widehat{\boldsymbol{\mathcal{I}}}$ and $\widehat{\boldsymbol{\alpha}}$ by q; after this adjustment, we would need to continue the NPG-BFGS iteration and seek the new local minimum. However, we can avoid additional iteration and simply adjust the regularization constant u *as well as* $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\mathcal{I}}}$ by assigning new values to them: $(u, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\mathcal{I}}}) \leftarrow \left(qu, \widehat{\boldsymbol{\alpha}}/q, [0, \widehat{\mathcal{I}}_1, \ldots, \widehat{\mathcal{I}}_{J-1}]^T/q\right)$. Apply this adjustment sequentially until the last element of the new $\widehat{\boldsymbol{\mathcal{I}}}$ is nonzero, which yields a local minimum $\left(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\mathcal{I}}}\right)$ of the new objective function $f_u(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ that is *not possible* to improve on by a simple shift adjustment. Our empirical experience is that scale-and-shift adjustment is either not needed (no zero elements at the end of $\widehat{\boldsymbol{\mathcal{I}}}$) or minor (very few zero elements): it slightly changes the grid of u over which we search for the best reconstructions, see also Section VI for discussion on selection of u.

C. Function restart and monotonicity

If $f\left(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}^{(i-1)}\right)$ is a convex function of $\boldsymbol{\alpha}$, apply [28, Lemma 2.3] to establish that the iterate $\boldsymbol{\alpha}^{(i)}$ attains lower (or equal) objective function than the intermediate signal $\bar{\boldsymbol{\alpha}}^{(i)}$

$$f\left(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\mathcal{I}}^{(i-1)}\right) \leq f\left(\bar{\boldsymbol{\alpha}}^{(i)}, \boldsymbol{\mathcal{I}}^{(i-1)}\right) - \frac{1}{2\beta^{(i)}}\left\|\boldsymbol{\alpha}^{(i)} - \bar{\boldsymbol{\alpha}}^{(i)}\right\|_2^2, \tag{30}$$

where we have used the fact that step size $\beta^{(i)}$ satisfies the majorization condition (26d). However, (30) *does not* guarantee monotonicity of Step 1. We apply the function restart [33] to ensure this monotonicity and improve convergence. In particular, we apply the function restart as follows: if monotonicity of Step 1 is violated in Iteration i, i.e., if

$$f\left(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\mathcal{I}}^{(i-1)}\right) > f\left(\boldsymbol{\alpha}^{(i-1)}, \boldsymbol{\mathcal{I}}^{(i-1)}\right) \qquad \text{(restart cond.)} \tag{31a}$$

set

$$\theta^{(i-1)} = 1 \tag{31b}$$

and *repeat* Step 1 using this selection. In this repeated step, the *momentum term* $\frac{\theta^{(i-1)}-1}{\theta^{(i)}}\left(\boldsymbol{\alpha}^{(i-1)} - \boldsymbol{\alpha}^{(i-2)}\right)$ in (26b) becomes zero, and

$$\bar{\boldsymbol{\alpha}}^{(i)} = \boldsymbol{\alpha}^{(i-1)} \tag{31c}$$

holds. Consequently, the new Step 1 is *monotonic*:

$$f\left(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\mathcal{I}}^{(i-1)}\right) \leq f\left(\boldsymbol{\alpha}^{(i-1)}, \boldsymbol{\mathcal{I}}^{(i-1)}\right), \tag{31d}$$

which follows by substituting (31c) into (30).

Once we can guarantee the monotonicity of Step 1 in every Iteration i, it is easy to establish the monotonicity of the entire NPG-BFGS iteration:

Remark 1 (Monotonicity): Under condition (24) of Theorem 1, the NPG-BFGS iteration with function restart is monotonically non-increasing:

$$f\left(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\mathcal{I}}^{(i)}\right) \leq f\left(\boldsymbol{\alpha}^{(i-1)}, \boldsymbol{\mathcal{I}}^{(i-1)}\right) \tag{32}$$

for all i.

Proof: Under condition (24), $f\left(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}\right)$ is a convex function of $\boldsymbol{\alpha}$. In this case, we have established that (31d) holds and Step 1 is monotonic. By Step 2, $f\left(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\mathcal{I}}^{(i-1)}\right) \geq f\left(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\mathcal{I}}^{(i)}\right)$ and (32) follows. ∎

Clearly, PG-BFGS and NPG (for known $\iota(\kappa)$) are monotonic as well under the convexity condition (24). To derive the monotonicity results, we have used only the fact that step size $\beta^{(i)}$ satisfies the majorization condition (26d), rather than using any specific details of the step-size selection.

In the following, we show that our PG-BFGS algorithm converges to a critical point of the objective function; interestingly, this convergence analysis *does not* require convexity of the objective function with respect to $\boldsymbol{\alpha}$. Unfortunately, these theoretical convergence properties do not carry over to the NPG-BFGS iteration, which empirically outperforms the PG-BFGS method; see Figs. 5 and 10 in Section VI.

D. Convergence analysis of the PG-BFGS iteration

We analyze the convergence of the PG-BFGS iteration using arguments similar to those in [30]. Although NPG-BFGS converges faster than PG-BFGS empirically, it is not easy to analyze its convergence due to NPG's Nesterov's acceleration step and adaptive step size. In this section, we denote the sequence of PG-BFGS iterates by $\left\{\left(\boldsymbol{\alpha}^{(i)}, \boldsymbol{\mathcal{I}}^{(i)}\right)\right\}_{i=0}^{\infty}$.

We have established the monotonicity of the PG-BFGS iteration for step sizes $\beta^{(i)}$ that satisfy the majorization condition, which includes the above step-size selection as well.

Since our $f(\boldsymbol{\alpha}, \mathcal{I})$ are lower bounded (which is easy to argue; see Appendix C), the sequence $f(\boldsymbol{\alpha}^{(i)}, \mathcal{I}^{(i)})$ converges. It is also easy to conclude that the sequence $a_i \triangleq \left\|\boldsymbol{\alpha}^{(i)} - \boldsymbol{\alpha}^{(i-1)}\right\|_2^2 / \beta^{(i)}$ is Cauchy by showing $\sum_{i=0}^{\infty} a_i < +\infty$ according to (30) when (31c) holds. Thus $\boldsymbol{\alpha}^{(i)}$ converges if $\{\beta^{(i)}\}_{i=1}^{\infty}$ is upper bounded.

A better result $\sum_{i=0}^{\infty} \left\|\boldsymbol{\alpha}^{(i)} - \boldsymbol{\alpha}^{(i+1)}\right\|_2 < +\infty$ [30] can be established because $f(\boldsymbol{\alpha}, \mathcal{I})$ satisfies the KL property. This property has been first used in [25] to establish the critical-point convergence for an alternating proximal-minimization method, which is then extended in [30] to the more general block coordinate-descent method. Using the analysis in [25], [34] shows the convergence of the alternating proximal-minimization algorithm by applying the KL property to a biconvex objective function.

Next, we make the following claim on the convergence of the PG-BFGS iteration.

Theorem 3: Consider the sequence $\left\{\left(\boldsymbol{\alpha}^{(i)}, \mathcal{I}^{(i)}\right)\right\}_{i=0}^{\infty}$ of PG-BFGS iterates, with step size $\beta^{(i)}$ satisfying the majorization condition (26d). Assume

1) bounded step size: there exist positive $\beta_+ > \beta_- > 0$ such that $\beta^{(i)} \in [\beta_-, \beta_+]$ for all i,
2) $\mathcal{L}(\boldsymbol{\alpha}, \mathcal{I})$ is a strongly convex function of $\mathcal{I}$, and
3) the gradient of $\mathcal{L}(\boldsymbol{\alpha}, \mathcal{I})$ with respect to $(\boldsymbol{\alpha}, \mathcal{I})$ is Lipschitz continuous.

Then $\left(\boldsymbol{\alpha}^{(i)}, \mathcal{I}^{(i)}\right)$ converges to one of the critical points $\left(\boldsymbol{\alpha}^{\star}, \mathcal{I}^{\star}\right)$ of $f(\boldsymbol{\alpha}, \mathcal{I})$ and

$$\sum_{i=1}^{\infty} \left\|\boldsymbol{\alpha}^{(i+1)} - \boldsymbol{\alpha}^{(i)}\right\|_2 < +\infty, \quad \sum_{i=1}^{\infty} \left\|\mathcal{I}^{(i+1)} - \mathcal{I}^{(i)}\right\|_2 < +\infty. \tag{33}$$

Proof: We apply [30, Lemma 2.6] to establish the convergence of $\left\{\left(\boldsymbol{\alpha}^{(i)}, \mathcal{I}^{(i)}\right)\right\}_{i=1}^{+\infty}$. Since $r(\boldsymbol{\alpha})$ in (25c) and $\mathbb{I}_{[0,+\infty)}(\mathcal{I})$ are lower-bounded, we need to prove only that (16) is lower-bounded. By using the fact that $\ln x \le x - 1$, we have

$$\mathcal{L}(\boldsymbol{\alpha}, \mathcal{I}) \ge 0. \tag{34}$$

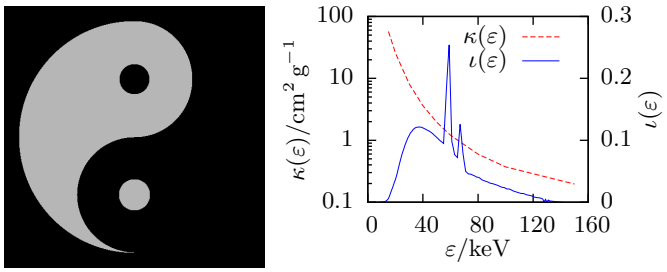
According to the assumption, $f(\boldsymbol{\alpha}, \mathcal{I})$ is strongly convex over $\mathcal{I}$ and the step size $\beta^{(i)}$ is bounded. Hence, there exist constants $0 < \ell < L < +\infty$ such that

$$f\left(\boldsymbol{\alpha}^{(i+1)}, \mathcal{I}^{(i)}\right) - f\left(\boldsymbol{\alpha}^{(i+1)}, \mathcal{I}^{(i+1)}\right) \ge \frac{\ell}{2} \left\|\mathcal{I}^{(i)} - \mathcal{I}^{(i+1)}\right\|_2^2$$

$$L \ge \frac{1}{\beta^{(i)}} \ge \ell.$$

In addition, $f(\boldsymbol{\alpha}, \mathcal{I})$ satisfies the KL property according to Theorem 2. We have now verified all conditions of [30, Lemma 2.6]. ∎

The conditions for strong convexity of $\mathcal{L}(\boldsymbol{\alpha}, \mathcal{I})$ as a function of $\mathcal{I}$ are discussed in Section IV; see also Section III-C. The KL property can provide guarantees on the convergence rate under additional assumptions; see [25, Theorem 3.4]. The convergence properties of NPG-BFGS are of great interest because NPG-BFGS converges faster than PG-BFGS; establishing these properties is left as future work.



(a) (b)

Fig. 2: (a) Density-map image used to generate the sinogram, and (b) mass attenuation and incident X-ray spectrum as functions of the photon energy ε.

VI. Numerical Examples

We now evaluate the proposed algorithms using simulated and real-data examples.

We construct the fan-beam X-ray projection transform matrix Φ and its adjoint operator Φ^T directly on GPU with circular masks [35]; the multi-thread version on CPU is also available; see https://github.com/isucsp/imgRecSrc, which also contains Matlab implementation of the proposed algorithms.

A. Simulation example

Consider the reconstruction of the 512×512 image in Fig. 2a of an iron object with density-map $\boldsymbol{\alpha}_{\text{true}}$. We generated a fan-beam polychromatic sinogram, with distance from the X-ray source to the rotation center equal to 2000 times the pixel size, using the interpolated mass attenuation $\kappa(\varepsilon)$ of iron [36] and the incident spectrum $\iota(\varepsilon)$ from tungsten anode X-ray tubes at 140 keV with 5% relative voltage ripple [37]; see Fig. 2b. The mass-attenuation spectrum $\iota(\kappa)$ is constructed by combining $\kappa(\varepsilon)$ and $\iota(\varepsilon)$ and shown in Fig. 1b, see also Fig. 1a. Our simulated approximation of the noiseless measurements uses 130 equi-spaced discretization points over the range 20 keV to 140 keV. We simulated independent Poisson measurements $(\mathcal{E}_n)_{n=1}^N$ with means $(\mathrm{E}\,\mathcal{E}_n)_{n=1}^N = \mathcal{I}^{\text{out}}(\boldsymbol{\alpha}, \mathcal{I})$. We mimic real X-ray CT system calibration by scaling projection matrix Φ and spectrum $\iota(\varepsilon)$ so that the maximum and minimum of the noiseless measurements $(\mathrm{E}\,\mathcal{E}_n)_{n=1}^N$ are 2^{16} and 20, respectively. Here, the scale of Φ corresponds to the real size that each image pixel represents, and the scale of $\iota(\varepsilon)$ corresponds to the current of the electrons hitting the tungsten anode as well as the overall scanning time.

Our goal is to reconstruct a 512×512 density-map using the measurements from an energy-integrating detector array of size 512 for each projection.

Since the true density-map is known, we adopt relative square error (RSE) as the main metric to assess the performance of the compared algorithms:

$$\mathrm{RSE}\{\hat{\boldsymbol{\alpha}}\} = 1 - \left(\frac{\hat{\boldsymbol{\alpha}}^T \boldsymbol{\alpha}_{\text{true}}}{\|\hat{\boldsymbol{\alpha}}\|_2 \|\boldsymbol{\alpha}_{\text{true}}\|_2}\right)^2 \tag{36}$$

where $\boldsymbol{\alpha}_{\text{true}}$ and $\hat{\boldsymbol{\alpha}}$ are the true and reconstructed signals, respectively. Note that (36) is invariant to scaling $\hat{\boldsymbol{\alpha}}$ by a nonzero

constant, which is needed because the magnitude level of $\boldsymbol{\alpha}$ is not identifiable due to the ambiguity of the density-map and mass-attenuation spectrum; see Section III-B.

We compare

- the traditional FBP methods
  - without linearization [3, Ch. 3] (termed FBP) and
  - with linearization to correct for the polychromatic source [19] (linearized FBP)

  based on the 'data'

$$\boldsymbol{y} = -\ln_{\circ} \boldsymbol{\mathcal{E}} \qquad \text{(without linearization)} \tag{37a}$$

$$\boldsymbol{y} = \left(\iota^{\mathrm{L}}\right)_{\circ}^{-1}(\boldsymbol{\mathcal{E}}) \qquad \text{(with linearization)} \tag{37b}$$

  respectively;
- linearized basis pursuit denoising (linearized BPDN), which applies the NPG approach to solve the analysis BPDN problem [24]: $\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\boldsymbol{y} - \Phi\boldsymbol{\alpha}\|_2^2 + u' r(\boldsymbol{\alpha})$, where $\boldsymbol{y}$ are the linearized measurements in (37b) and the penalty $r(\boldsymbol{\alpha})$ has been defined in (25c);
- our
  - NPG-BFGS algorithm with the B1-spline tuning constants (13e) chosen to satisfy

$$q^J = 10^3, \qquad \kappa_{\lceil 0.5(J+1)\rceil} = 1, \qquad J = 30 \tag{38}$$

  - NPG (known $\iota(\kappa)$) algorithm for estimating $\boldsymbol{\alpha}$

  with $\mathbb{n} = 4$; see Section V-B.

The linearizing transform (37b) assumes knowledge of the mass-attenuation spectrum $\iota(\kappa)$ and, in the absence of noise, leads to the linear model $\boldsymbol{y} = \Phi\alpha$ under the general polychromatic-source scenario. In contrast, the standard logarithm transformation of the X-ray measurements (37a) *ignores* the hardening effect and can possibly lead to the linear model only for monochromatic X-ray sources. If the X-ray source is monochromatic, (37a) and (37b) *coincide* up to a known additive constant, and the two FBP methods are identical; in this case, linearized BPDN also coincides with the standard analysis BPDN approach applied to X-ray CT data.

For all methods that use sparsity and nonnegativity regularization (NPG-BFGS, NPG, and linearized BPDN), the regularization constants u and u' have been tuned manually for the best average RSE performance for each number of projections using a 9-point grid spanning 9 orders of magnitude.

All iterative algorithms employ the convergence criterion (29) with the threshold $\epsilon = 10^{-6}$ and the maximum number of iterations set to 4000. We initialize iterative reconstruction schemes with or without linearization using the corresponding FBP reconstructions; see also [12, Sec. IV-B4] for details on NPG-BFGS initialization.

Here, the *non-blind* linearized FBP, NPG (known $\iota(\kappa)$), and linearized BPDN methods assume known $\iota(\kappa)$ (which requires knowledge of the incident spectrum of the X-ray machine and mass attenuation (material)), computed using (10a), with $\boldsymbol{\mathcal{I}}$ equal to the exact sampled $\iota(\kappa)$ and $J = 100$ spline basis functions spanning three orders of magnitude.

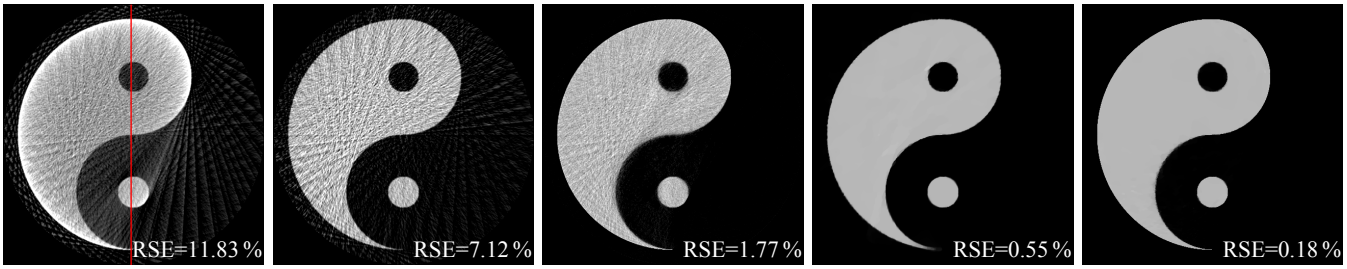
Neither FBP nor NPG-BFGS assumes knowledge of the mass-attenuation spectrum $\iota(\kappa)$: FBP *ignores* the polychromatic-source effects whereas NPG-BFGS corrects *blindly* for these effects *without* knowledge of $\iota(\kappa)$.

Figs. 3 and 4 show the reconstructed density-map images and profiles of different methods from 60 equi-spaced fan-beam projections with spacing 6°, using one realization of noisy Poisson measurements. Fig. 5 shows the RSEs of several methods as functions of the iteration index i and demonstrates that RSE of NPG-BFGS decreases significantly faster with increasing i than the RSE of PG-BFGS; NPG-BFGS also converges faster than PG-BFGS. The FBP reconstruction in Fig. 3a is corrupted by both aliasing and beam-hardening (cupping and streaking) artifacts. Linearized FBP removes the beam-hardening artifacts but retains the aliasing artifacts and enhances noise due to the zero-forcing nature of linearization; see Fig. 3b. Linearized BPDN enforces the signal nonnegativity and sparsity constraints and achieves a smooth reconstruction in Fig. 3d with a 0.55 % RSE. Thanks to the superiority of the proposed model that accounts for both the polychromatic X-ray source and Poisson noise, NPG-BFGS and NPG achieve the best (and nearly the same) reconstructions; see Fig. 3e.

We also show in Fig. 3c the reconstruction by the NPG-BFGS method with very small u (labeled NPG-BFGS$_0$), which effectively removes the signal sparsity constraint and imposes only the signal nonnegativity constraint; consequently, Step 1 in NPG-BFGS$_0$ iteration has a closed form and reduces to simple nonnegativity thresholding. Hence, NPG-BFGS$_0$ is a maximum-likelihood (ML) approach that aims at minimizing the NLL (16) subject to the physical parameter constraints $\boldsymbol{\alpha} \succeq \mathbf{0}$ and $\boldsymbol{\mathcal{I}} \succeq \mathbf{0}$. As NPG-BFGS$_0$ iterates, its RSE decreases, reaches a minimum, and then increases; see Fig. 5. This is a common behavior for unregularized ML image reconstruction approaches [38]. Fig. 3c shows this method's reconstruction at iteration step $i = 500$, which gives the best RSE; see also Fig. 5. Since it terminates early and has a simple Step 1, NPG-BFGS$_0$ running only 500 iterations is roughly 8 times faster than NPG-BFGS. The NPG-BFGS$_0$ method can be thought of as an improved version of [11], which also imposes only signal nonnegativity. A comparison of NPG-BFGS$_0$ and NPG-BFGS shows the benefit of signal-sparsity regularization.

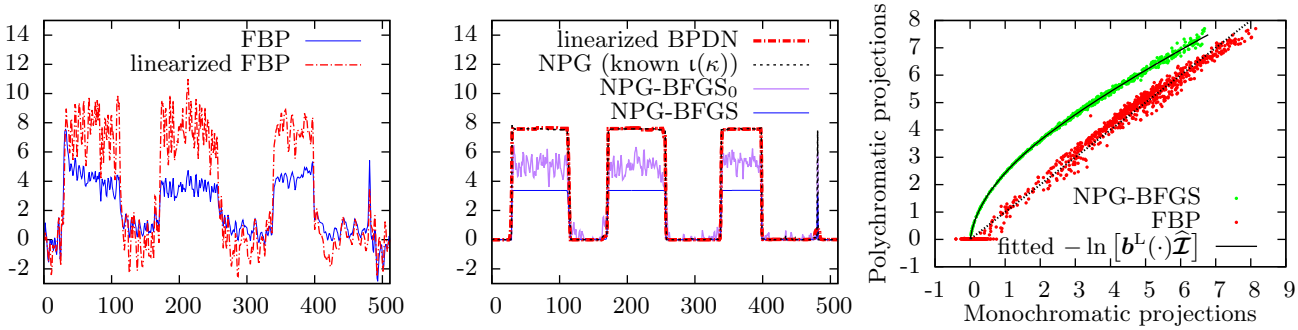
Figs. 4a and 4b show the reconstruction profiles of the 250th column, indicated by the red line in Fig. 3a. Recall that NPG-BFGS *cannot* identify the magnitude level of the density-map image $\boldsymbol{\alpha}$, which explains the corresponding magnitude discrepancy between NPG-BFGS, NPG-BFGS$_0$, and the non-blind methods in Fig. 4b. We have corrected this discrepancy manually in Fig. 3 because we wish to show visual quality and ability of different methods to remove artifacts and suppress noise, rather than the trivial difference in image contrast.

In Fig. 4c, we show the scatter plots with 1000 randomly selected points representing FBP and NPG-BFGS reconstructions from 60 fan-beam projections. Denote by $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\mathcal{I}}})$ the estimate of $(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$ obtained upon convergence of the NPG-BFGS iteration. The y-coordinates in the scatter plots in Fig. 4c are the *noisy* measurements in log scale $-\ln \mathcal{E}_n$, and the corresponding x-coordinates are the monochromatic projections $\boldsymbol{\phi}_n^T \widehat{\boldsymbol{\alpha}}_{\mathrm{FBP}}$ (red) and $\boldsymbol{\phi}_n^T \widehat{\boldsymbol{\alpha}}$ (green) of the estimated density-maps. $-\ln\left[\boldsymbol{b}^{\mathrm{L}}(\cdot)\widehat{\boldsymbol{\mathcal{I}}}\right]$ is the inverse linearization function that maps monochromatic projections to fitted *noiseless* polychromatic projections $-\ln \mathcal{I}_n^{\mathrm{out}}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\mathcal{I}}})$. Since FBP assumes a linear relation between $-\ln_{\circ} \boldsymbol{\mathcal{I}}^{\mathrm{out}}$ and $\Phi\boldsymbol{\alpha}$, its scatter plot (red) can be fitted



(a) FBP (b) linearized FBP (c) NPG-BFGS$_0$ (d) linearized BPDN (e) NPG-BFGS and NPG (known $\iota(\kappa)$)

Fig. 3: Reconstructions from 60 projections.



(a) 250th column (b) 250th column (c)

Fig. 4: (a)–(b) Reconstruction profiles of different methods from 60 projections and (c) the polychromatic measurements as function of the monochromatic projections and corresponding fitted inverse linearization curves.

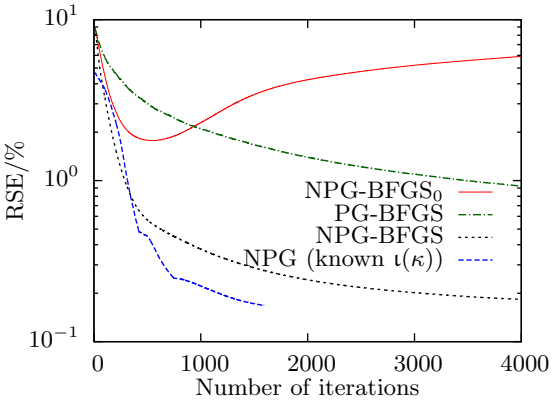


Fig. 5: The RSEs as functions of the iteration index i.

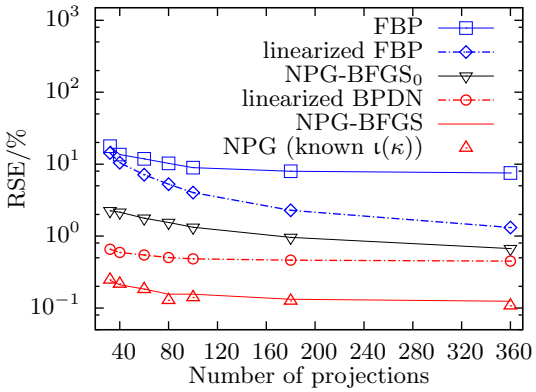


Fig. 6: Average RSEs as functions of the number of projections.

by a straight line $y = x$, as shown in Fig. 4c. A few points in the FBP scatter plot with $\ln \mathcal{E}_n = 0$ and positive monochromatic projections indicate severe streaking artifacts. Observe relatively large residuals with bias, which remain even when more sophisticated linear models, e.g. iterative algorithms with sparsity and nonnegativity constraints, were adopted, thereby necessitating the need for accounting for the polychromatic source. The nonnegativity constraints on $\boldsymbol{\alpha}$ are particularly important for good estimation of $\boldsymbol{b}^{\mathrm{L}}(\cdot)\boldsymbol{\mathcal{I}}$.

Fig. 6 shows the average RSEs (over 5 Poisson noise realizations) of different methods as functions of the number of fan-beam projections in the range from 0° to 359°. Average RSEs of the methods that do not assume knowledge of the mass-attenuation spectrum $\iota(\kappa)$ are shown using solid lines; dashed lines represent non-blind methods that assume known mass-attenuation spectrum $\iota(\kappa)$. Red color represents methods that employ both signal-sparsity regularization and nonnegativity image constraints, black is for the method that employs the nonnegativity image constraints only, and blue marks the methods that apply neither signal-sparsity regularization nor nonnegativity image constraints.

FBP ignores the polychromatic nature of the measurements; consequently, it performs poorly and does not improve as the number of projections increases. Linearized FBP, which assumes

perfect knowledge of the mass-attenuation spectrum, performs much better than FBP, as shown in Fig. 6. Thanks to the signal nonnegativity and sparsity that it imposes, linearized BPDN achieves up to 20 times smaller RSEs compared with the linearized FBP. However, due to its zero-forcing nature, linearized BPDN enhances noise and breaches the Poisson measurement model, which explains its inferior performance compared with NPG (known $\iota(\kappa)$).

As expected, NPG (known $\iota(\kappa)$) performs slightly better than NPG-BFGS because it uses perfect knowledge of $\iota(\kappa)$. NPG (known $\iota(\kappa)$) and NPG-BFGS attain RSEs that are 24% to 37% of that achieved by linearized BPDN, which can be attributed to optimal statistical processing by these methods, in contrast with the suboptimal linearization. It is remarkable that the blind NPG-BFGS method effectively matches the performance of NPG (known $\iota(\kappa)$).

B. Real-data examples

We compare the NPG-BFGS and linear FBP methods by applying them to reconstruct two industrial objects containing defects, labeled `C-I` and `C-II`, from real fan-beam projections. Here, NPG-BFGS achieves visually good reconstructions for $u = 10^{-5}$, presented in Fig. 7, where we also show its reconstruction for $u = 10^{-4}$.

The `C-I` data set consists of 360 equi-spaced fan-beam projections with 1° separation collected using an array of 694 detectors, with X-ray source to rotation center distance equal to 3492 times the detector size. Figs. 7a and 7b show 512×512 density-map image reconstructions of object `C-I` using the FBP and NPG-BFGS methods, respectively. The linear FBP reconstruction, which does not account for the polychromatic nature of the X-ray source, suffers from severe streaking and cupping artifacts, whereas the NPG-BFGS reconstruction removes these artifacts by accounting for the polychromatic X-ray source.

The `C-II` data set consists of 360 equi-spaced fan-beam projections with 1° separation collected using an array of 1380 detectors, with X-ray source to rotation center distance equal to 8696 times the detector size. Figs. 7c–7e show 1024×1024 density-map image reconstructions of object `C-II` by the FBP, NPG-BFGS$_0$, and NPG-BFGS methods, respectively. The NPG-BFGS and NPG-BFGS$_0$ reconstructions do not have streaking and cupping artifacts exhibited by FBP. NPG-BFGS$_0$ terminates after 500 iterations and is 2 to 3 times faster than NPG-BFGS.

Figs. 7g–7i show the FBP, NPG-BFGS$_0$ (terminated at $i = 500$ iterations), and NPG-BFGS reconstructions from a downsampled `C-II` data set with 120 equi-spaced fan-beam projections with 3° separation. The FBP reconstruction in Fig. 7g exhibits both beam-hardening and aliasing artifacts. In contrast, the NPG-BFGS reconstruction in Fig. 7i does not have these artifacts because it accounts for the polychromatic X-ray source and employs signal-sparsity regularization in (25c). Indeed, if we reduce regularization constant u sufficiently, the aliasing effect will occur in the NPG-BFGS reconstruction in Fig. 7i as well. A comparison of NPG-BFGS$_0$ and NPG-BFGS shows the benefit of signal-sparsity regularization, particularly its ability to reduce noise. If we run NPG-BFGS$_0$ beyond $i = 500$ iterations, it will exhibit aliasing artifacts, in addition to noise.

Fig. 8 shows the reconstruction profiles of the 337th and 531th rows highlighted by the red horizontal lines across Figs. 7c and 7e. Noise in the NPG-BFGS reconstructions can be reduced by increasing regularization parameter u: Figs. 8c and 8d show the corresponding NPG-BFGS reconstruction profiles for $u = 10^{-4}$, which is 10 times that in Figs. 8a and 8b.

The NPG-BFGS reconstructions of `C-I` and `C-II` have higher contrast around the inner region where cracks reside, which may be due to the detector saturation that leads to measurement truncation, scattering, noise-model mismatch, or the bowtie filter applied to the X-ray source. We leave further verification of causes and potential correction of this problem to future work and note that this issue does not occur in the simulated-data examples that we constructed; see Section VI-A.

In Fig. 9, we show the scatter plots with 1000 randomly selected points representing FBP and NPG-BFGS reconstructions of the `C-II` object from 360 projections. A few points in the FBP scatter plot with $\ln \mathcal{E}_n = 0$ and positive monochromatic projections indicate severe streaking artifacts, which we also observed in the simulation example; see Fig. 4c.

We now illustrate the advantage of using Nesterov's acceleration in Step 1 of NPG-BFGS. Fig. 10 shows the centered objective $f(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) - f_{\text{MIN}}$ with $u = 10^{-5}$ as a function of the iteration index i for the NPG-BFGS and PG-BFGS methods applied to the `C-II` reconstruction from 360 projections; here $f_{\text{MIN}} = \min_{\boldsymbol{x}} f(\boldsymbol{x})$. Thanks to the Nesterov's acceleration (26b), NPG-BFGS is 2 to 3 times faster than PG-BFGS.

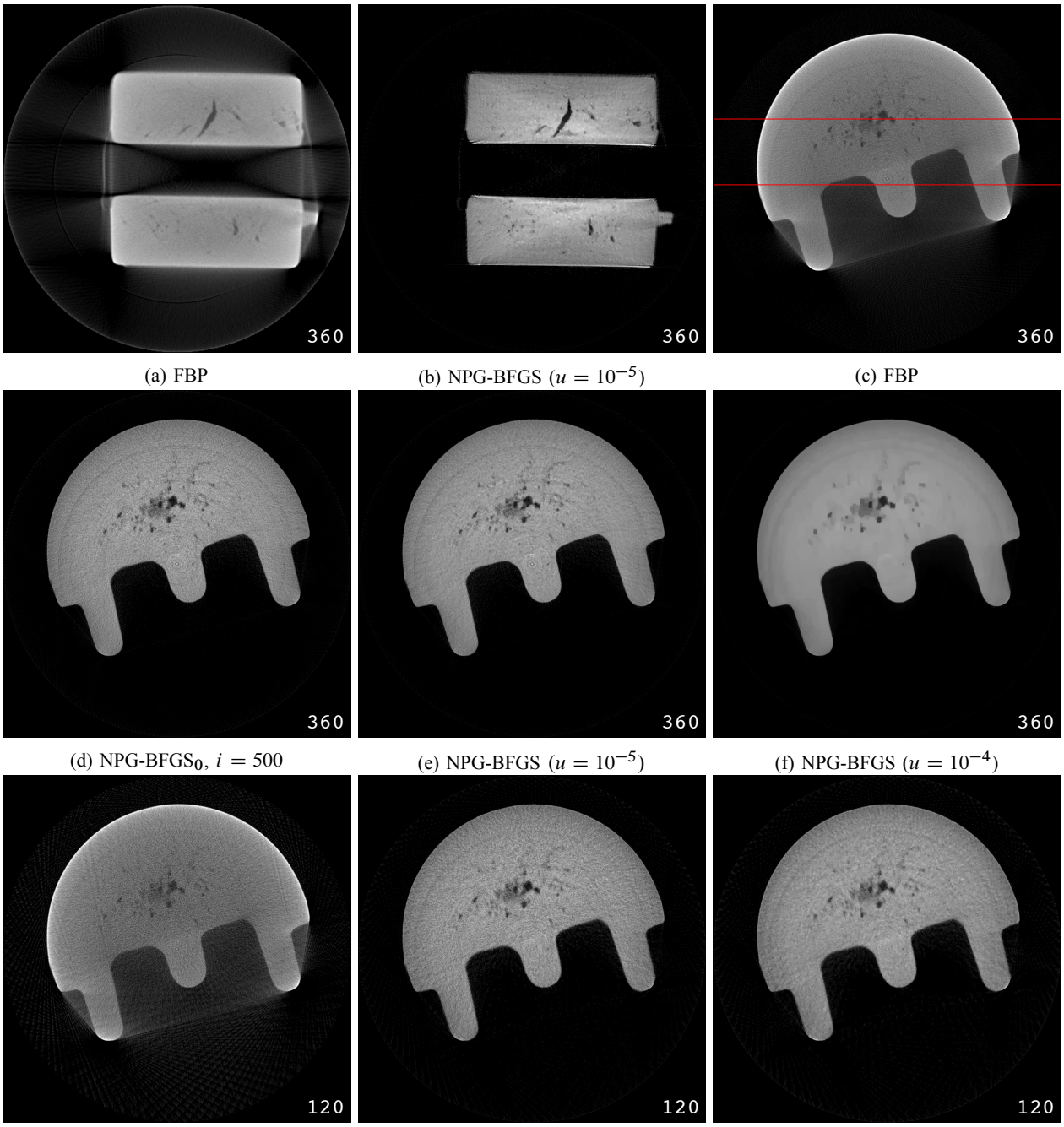
VII. Conclusion

We developed a model for single-material beam-hardening artifact correction that requires no more information than the conventional FBP method. The proposed model relies on separability of the attenuation to combine the variations of the mass attenuation and X-ray spectrum into the *mass-attenuation spectrum*. Numerical experiments on both simulated and real X-ray CT data were presented. Our *blind* method for sparse X-ray CT reconstruction matches or outperforms non-blind linearization methods that assume perfect knowledge of the X-ray source and material properties. Future work will include extending our parsimonious polychromatic measurement-model parameterization to multiple materials [39] and developing corresponding reconstruction algorithms.

Appendix A
Mass-Attenuation Parameterization

All mass-attenuation functions $\kappa(\varepsilon)$ encountered in practice can be divided into piecewise-continuous segments, where each segment is a differentiable monotonically decreasing function of ε, see [36, Tables 3 and 4] and [40, Sec. 2.3]. The points of discontinuity in $\kappa(\varepsilon)$ are referred to as K-edges and are caused by the interaction between photons and K shell electrons, which occurs only when ε reaches the binding energy of the K shell electron. One example in Fig. 11 is the mass attenuation coefficient curve of iron with a single K-edge at 7.11 keV.

We define the domain $\mathcal{E}$ of ε and partition it into $M+1$ intervals $\big((e_m, e_{m+1})\big)_{m=0}^{M}$ with $e_0 = \min(\mathcal{E})$ and $e_{M+1} = \max(\mathcal{E})$, such that in each interval $\kappa(\varepsilon)$ is invertible and differentiable.



(a) FBP (b) NPG-BFGS ($u = 10^{-5}$) (c) FBP

(d) NPG-BFGS$_0$, $i = 500$ (e) NPG-BFGS ($u = 10^{-5}$) (f) NPG-BFGS ($u = 10^{-4}$)

(g) FBP (h) NPG-BFGS$_0$, $i = 500$ (i) NPG-BFGS ($u = 10^{-5}$)

Fig. 7: Real X-ray CT reconstructions of objects `C-I` and `C-II` from (a)–(f) 360 and (g)–(h) 120 projections.

Here, $\mathcal{E}$ is the support set of the incident X-ray spectrum $\iota(\varepsilon)$ and $(e_m)_{m=1}^M$ are the M K-edges in $\mathcal{E}$. Taking Fig. 11 as an example, there is only one K-edge at e_1, given that the incident spectrum has its support as (e_0, e_2). The range and inverse of $\kappa(\varepsilon)$ within (e_m, e_{m+1}) are (u_m, v_m) and $\varepsilon_m(\kappa)$, respectively, with $u_m \triangleq \inf_{\varepsilon \nearrow e_{m+1}} \kappa(\varepsilon) < v_m \triangleq \sup_{\varepsilon \searrow e_m} \kappa(\varepsilon)$. Then, the noiseless measurement in (4b) can be written as

$$\mathcal{I}^{\text{out}} = \int \sum_{m=0}^{M} 1_{(u_m, v_m)}(\kappa) \iota(\varepsilon_m(\kappa)) \left|\varepsilon_m'(\kappa)\right| \mathrm{e}^{-\kappa \int \alpha(x,y)\, \mathrm{d}\ell}\, \mathrm{d}\kappa,$$

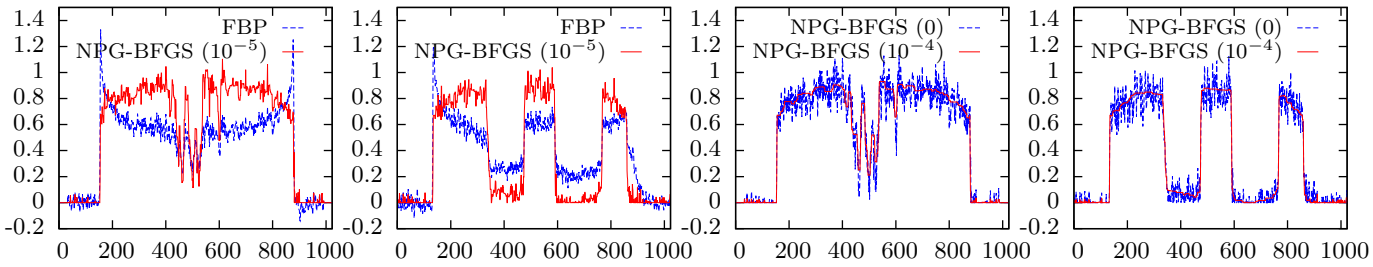
and (6b) and (6a) follow by noting that

$$\iota(\kappa) = \sum_{m=0}^{M} 1_{(u_m, v_m)}(\kappa) \iota(\varepsilon_m(\kappa)) \left|\varepsilon_m'(\kappa)\right| \geq 0 \qquad \text{(A1)}$$

and that $\mathcal{I}^{\text{out}}$ equals $\mathcal{I}^{\text{in}}$ when $\alpha(x, y) = 0$. Here, $1_{(u_m, v_m)}(\kappa)$ is an indicator function that takes value 1 when $\kappa \in (u_m, v_m)$ and 0 otherwise. Observe that (A1) reduces to (7) when $M = 0$.

Appendix B
Proof of Lemma 1

We first introduce a lemma.



(a) 337th row (b) 531st row (c) 337th row (d) 531st row

Fig. 8: `C-II` object reconstruction profiles from 360 projections with (a)–(b) $u = 10^{-5}$ and (c)–(d) $u = 10^{-4}$ used by the NPG-BFGS method.

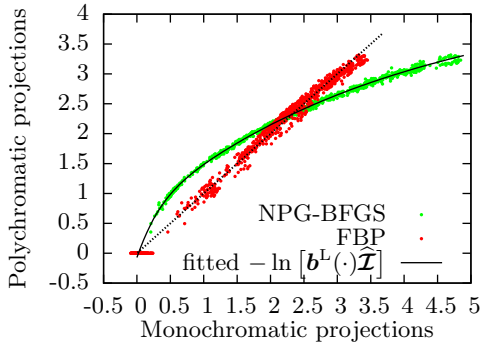


Fig. 9: Polychromatic measurements as functions of monochromatic projections and corresponding inverse linearization curves.

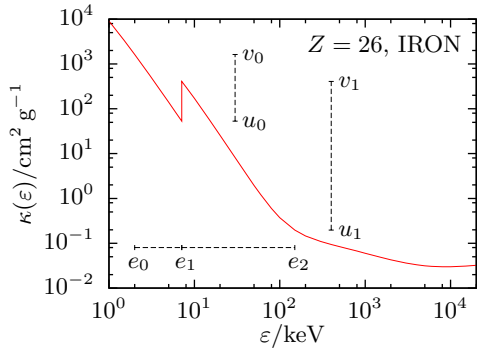


Fig. 11: The mass attenuation coefficients κ of iron versus the photon energy ε with a K-edge at 7.11 keV.

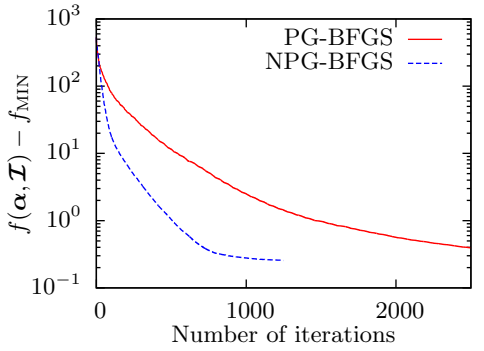


Fig. 10: Centered objectives as functions of the iteration index i.

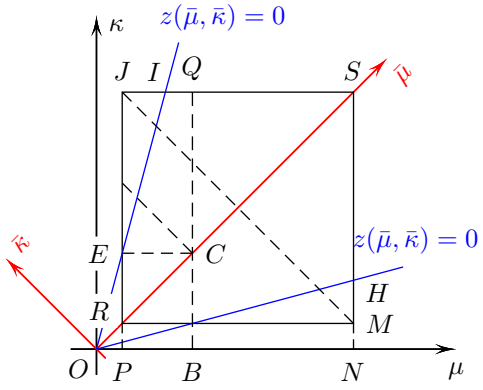


Fig. 12: Integral region illustration.

Lemma 3: For $\iota(\kappa)$ that satisfy Assumption 1, the following holds:

$$\mathrm{w} \triangleq \iint \left[\mu\kappa - \tfrac{q^{j_0}}{(q^{j_0}+1)^2}(\mu+\kappa)^2\right] \iota(\kappa)\iota(\mu) h(\kappa+\mu)\,\mathrm{d}\mu\,\mathrm{d}\kappa \geq 0 \tag{B1}$$

for $q > 1$ and any nonnegative function $h : \mathbb{R} \to \mathbb{R}_+$.

Proof: In Fig. 12, the (μ, κ) coordinates of P, B and N are $(\kappa_0, 0)$, $(\kappa_{j_0}, 0)$ and $(\kappa_{J+1}, 0)$, respectively; the line OS is defined by $\kappa = \mu$.

Considering the finite support set of $\iota(\kappa)$, the effective integral range is $[\kappa_0, \kappa_{J+1}]^2$, which is the rectangle $RMSJ$ in Fig. 12.

Using the symmetry between κ and μ in (B1), we change the integral variables of (B1) by rotating the coordinates by 90°:

$$\mu = \tfrac{\bar{\mu}-\bar{\kappa}}{\sqrt{2}}, \qquad \kappa = \tfrac{\bar{\mu}+\bar{\kappa}}{\sqrt{2}} \tag{B2}$$

which yields

$$\mathrm{w} = \int_{\sqrt{2}\kappa_0}^{\sqrt{2}\kappa_{J+1}} \int_0^{g(\bar{\mu})} \bar{w}(\bar{\mu}, \bar{\kappa})\,\mathrm{d}\bar{\kappa}\; h\big(\sqrt{2}\bar{\mu}\big)\,\mathrm{d}\bar{\mu} \tag{B3a}$$

where

$$\bar{w}(\bar{\mu},\bar{\kappa}) \triangleq z(\bar{\mu},\bar{\kappa})\iota\Big(\tfrac{\bar{\mu}+\bar{\kappa}}{\sqrt{2}}\Big)\iota\Big(\tfrac{\bar{\mu}-\bar{\kappa}}{\sqrt{2}}\Big) \tag{B3b}$$

$$z(\bar{\mu},\bar{\kappa}) \triangleq \left(\frac{q^{j_0}-1}{q^{j_0}+1}\right)^2 \bar{\mu}^2 - \bar{\kappa}^2 \tag{B3c}$$

$$g(\bar{\mu}) \triangleq \begin{cases} \bar{\mu} - \sqrt{2}\kappa_0, & \bar{\mu} \le \frac{1}{\sqrt{2}}(\kappa_0 + \kappa_{J+1}) \\ \sqrt{2}\kappa_{J+1} - \bar{\mu}, & \bar{\mu} > \frac{1}{\sqrt{2}}(\kappa_0 + \kappa_{J+1}) \end{cases} \tag{B3d}$$

and (B3a) follows because (B3b) is even-symmetric with respect to $\bar{\kappa}$. Hence, the integration region is reduced to the triangle RSJ.

Note that $z(\bar{\mu},\bar{\kappa}) \ge 0$ in the cone between lines OH and OI, [both of which are specified by $z(\bar{\mu},\bar{\kappa}) = 0$], which implies that $\bar{w}(\bar{\mu},\bar{\kappa}) \ge 0$ within RCE and CSQ; hence, the integrals of $\bar{w}(\bar{\mu},\bar{\kappa})h\big(\sqrt{2}\bar{\mu}\big)$ over RCE and CSQ are nonnegative and, consequently,

$$\mathrm{w} \ge \iint_{\mathcal{R}} \bar{w}(\bar{\mu},\bar{\kappa})\,\mathrm{d}\bar{\kappa}\; h\big(\sqrt{2}\bar{\mu}\big)\,\mathrm{d}\bar{\mu}. \tag{B4}$$

Now

$$\mathcal{R} \triangleq \left\{(\bar{\mu},\bar{\kappa}) \;\middle|\; \frac{\bar{\mu}-\bar{\kappa}}{\sqrt{2}} \in [\kappa_0,\kappa_{j_0}],\ \frac{\bar{\mu}+\bar{\kappa}}{\sqrt{2}} \in [\kappa_{j_0},\kappa_{J+1}]\right\} \tag{B5}$$

is our new integration region, which is the rectangle $ECQJ$.

Next, we split the inner integral over $\bar{\kappa}$ on the right-hand side of (B4) for fixed $\bar{\mu}$ into two regions: $z(\bar{\mu},\bar{\kappa}) \ge 0$ and $z(\bar{\mu},\bar{\kappa}) < 0$, i.e., trapezoid $ECQI$ and triangle EIJ, and prove that the positive contribution of the integral over $ECQI$ is larger than the negative contribution of the integral over the EIJ.

The line OI is specified by $z(\bar{\mu},\bar{\kappa}) = 0$, and the (μ,κ)-coordinate of I in Fig. 12 is thus $(\kappa_{J+1-j_0},\kappa_{J+1})$. Define

$$c \triangleq \frac{\sqrt{2}}{1+q^{j_0}} \tag{B6}$$

and note that $ECQI \subseteq (\mathcal{K}_{\text{low}} \cup \mathcal{K}_{\text{mid}}) \times \mathcal{K}_{\text{high}}$ and $EIJ \subseteq \mathcal{K}_{\text{low}} \times \mathcal{K}_{\text{high}}$. We now use Assumption 1 to conclude that the following hold within $\mathcal{R}$:

- When $z(\bar{\mu},\bar{\kappa}) \ge 0$, i.e., in region $ECQI$,

$$\iota(\kappa)\,|_{\kappa=\frac{\bar{\mu}+\bar{\kappa}}{\sqrt{2}}} \ge \iota\big(cq^{j_0}\bar{\mu}\big) \tag{B7a}$$

$$\iota(\mu)\,|_{\mu=\frac{\bar{\mu}-\bar{\kappa}}{\sqrt{2}}} \ge \iota\big(c\bar{\mu}\big) \tag{B7b}$$

  where (B7a) follows because $\kappa = \frac{\bar{\mu}+\bar{\kappa}}{\sqrt{2}}$ takes values between κ_{j_0} and $cq^{j_0}\bar{\mu} \in [\kappa_{j_0},\kappa_{J+1}]$; i.e., $\kappa \in \mathcal{K}_{\text{high}}$ and $\iota(\kappa)$ decreases in $\mathcal{K}_{\text{high}}$. (B7b) follows because $\mu = \frac{\bar{\mu}-\bar{\kappa}}{\sqrt{2}}$ takes values between $c\bar{\mu} \in [\kappa_0,\kappa_{J+1-j_0}]$ and κ_{j_0}; i.e., μ crosses $\mathcal{K}_{\text{low}}$ ($\iota(\kappa)$ increasing) and $\mathcal{K}_{\text{mid}}$ ($\iota(\kappa)$ high) regions.
- When $z(\bar{\mu},\bar{\kappa}) < 0$, i.e., in region EIJ,

$$\iota(\kappa)\,|_{\kappa=\frac{\bar{\mu}+\bar{\kappa}}{\sqrt{2}}} < \iota\big(cq^{j_0}\bar{\mu}\big) \tag{B7c}$$

$$\iota(\mu)\,|_{\mu=\frac{\bar{\mu}-\bar{\kappa}}{\sqrt{2}}} < \iota\big(c\bar{\mu}\big) \tag{B7d}$$

  where (B7c) follows because $\kappa = \frac{\bar{\mu}+\bar{\kappa}}{\sqrt{2}} > cq^{j_0}\bar{\mu}$, i.e., $\kappa \in \mathcal{K}_{\text{high}}$, and (B7d) follows because $\mu = \frac{\bar{\mu}-\bar{\kappa}}{\sqrt{2}} < c\bar{\mu}$, i.e., $\mu \in \mathcal{K}_{\text{low}}$.

By combining (B7) and (B4), we have

$$\mathrm{w} \ge \int_{(\kappa_0+\kappa_{j_0})/\sqrt{2}}^{(\kappa_{J+1}+\kappa_{J+1-j_0})/\sqrt{2}} \int_{\{\bar{\kappa}\,|\,(\bar{\mu},\bar{\kappa})\in\mathcal{R}\}} z(\bar{\mu},\bar{\kappa})\,\mathrm{d}\bar{\kappa}\,\bar{h}(\bar{\mu})\,\mathrm{d}\bar{\mu} \tag{B8}$$

where $\bar{h}(\bar{\mu}) \triangleq \iota\big(cq^{j_0}\bar{\mu}\big)\iota\big(c\bar{\mu}\big)h\big(\sqrt{2}\bar{\mu}\big) \ge 0$. It is easy to verify that $\int_{\{\bar{\kappa}\,|\,(\bar{\mu},\bar{\kappa})\in\mathcal{R}\}} z(\bar{\mu},\bar{\kappa})\,\mathrm{d}\bar{\kappa}$ is an increasing function of $\bar{\mu}$ over the range of the outer integral $[(\kappa_0+\kappa_{j_0})/\sqrt{2}, (\kappa_{J+1}+\kappa_{J+1-j_0})/\sqrt{2}]$, and, consequently,

$$\int_{\{\bar{\kappa}\,|\,(\bar{\mu},\bar{\kappa})\in\mathcal{R}\}} z(\bar{\mu},\bar{\kappa})\,\mathrm{d}\bar{\kappa} \ge 0, \tag{B9}$$

where the equality is attained for $\bar{\mu} = (\kappa_0+\kappa_{j_0})/\sqrt{2}$. Finally, (B1) follows from (B8) and (B9). ∎

This proof of convexity of Lemma 3 is conservative as we loosen the positive integrals in regions RCE and CSQ by replacing them with zeros.

We now use Lemma 3 to prove the convexity of $\mathcal{L}_\iota(\boldsymbol{\alpha})$ in Lemma 1. Note that the mass-attenuation spectrum $\iota(\kappa)$ is considered known in Lemma 1. We define $\xi(\cdot) \triangleq \iota^{\mathrm{L}}(\cdot)$ and the corresponding first and second derivatives: $\dot{\xi}(s) = (-\kappa\iota)^{\mathrm{L}}(s)$ and $\ddot{\xi}(s) = \big(\kappa^2\iota\big)^{\mathrm{L}}(s)$. Observe that $\boldsymbol{\mathcal{I}}^{\text{out}} = \big(\mathcal{I}_n^{\text{out}}\big)_{n=1}^N = \xi_\circ(\Phi\boldsymbol{\alpha}) = \big(\xi(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})\big)_{n=1}^N$. For notational simplicity, we omit the dependence of $\boldsymbol{\mathcal{I}}^{\text{out}}$ on $\boldsymbol{\alpha}$ and $\boldsymbol{\mathcal{I}}$ and use $\boldsymbol{\mathcal{I}}^{\text{out}}$ and $\xi_\circ(\Phi\boldsymbol{\alpha})$ interchangeably.

We use the identities

$$\frac{\partial \xi_\circ(\Phi\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^T} = \operatorname{diag}\big(\dot{\xi}_\circ(\Phi\boldsymbol{\alpha})\big)\Phi \tag{B10a}$$

$$\frac{\partial \xi(\boldsymbol{\phi}_n^T\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\,\partial\boldsymbol{\alpha}^T} = \ddot{\xi}\big(\boldsymbol{\phi}_n^T\boldsymbol{\alpha}\big)\boldsymbol{\phi}_n\boldsymbol{\phi}_n^T \tag{B10b}$$

to compute the gradient and Hessian of the Poisson NLL in (18):

$$\frac{\partial \mathcal{L}_\iota(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}} = \Phi^T \operatorname{diag}\Big(\dot{\xi}_\circ\big(\Phi\boldsymbol{\alpha}\big)\Big)\Big[\mathbf{1} - \operatorname{diag}^{-1}\big(\boldsymbol{\mathcal{I}}^{\text{out}}\big)\boldsymbol{\mathcal{E}}\Big] \tag{B11a}$$

$$\frac{\partial \mathcal{L}_\iota(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\,\partial\boldsymbol{\alpha}^T} = \Phi^T \operatorname{diag}^{-2}\big(\boldsymbol{\mathcal{I}}^{\text{out}}\big)\operatorname{diag}(\boldsymbol{\mathcal{E}})\operatorname{diag}(\boldsymbol{x})\Phi \tag{B11b}$$

where the $N\times 1$ vector $\boldsymbol{x} = (x_n)_{n=1}^N$ is defined as

$$x_n = \dot{\xi}^2\big(s\big) + \ddot{\xi}\big(s\big)\xi(s)\Big(\tfrac{\mathcal{I}_n^{\text{out}}}{\mathcal{E}_n} - 1\Big)\Big|_{s=\boldsymbol{\phi}_n^T\boldsymbol{\alpha}}. \tag{B11c}$$

Since $\mathcal{I}_n^{\text{out}} \ge (1-V)\mathcal{E}_n \ge 0$ according to (21a), we have

$$\frac{\mathcal{I}_n^{\text{out}}}{\mathcal{E}_n} - 1 \ge -V \tag{B12}$$

and

$$x_n \ge \iint \big(\mu\kappa - \kappa^2 V\big)\iota(\mu)\iota(\kappa)\mathrm{e}^{-(\mu+\kappa)\boldsymbol{\phi}_n^T\boldsymbol{\alpha}}\,\mathrm{d}\kappa\,\mathrm{d}\mu \tag{B13a}$$

$$= \iint \Big(\mu\kappa - \tfrac{\mu^2+\kappa^2}{2}V\Big)\iota(\mu)\iota(\kappa)\mathrm{e}^{-(\mu+\kappa)\boldsymbol{\phi}_n^T\boldsymbol{\alpha}}\,\mathrm{d}\kappa\,\mathrm{d}\mu \tag{B13b}$$

$$\ge \frac{(q^{j_0}+1)^2}{q^{2j_0}+1}\mathrm{w} \ge 0 \tag{B13c}$$

where (B13a) follows by applying inequality (B12) to (B11c), using the Laplace-transform identity for derivatives (2), and combining the multiplication of the integrals; and (B13b) is due to the symmetry with respect to μ and κ. Now, plug (21b) into (B13b) and apply Lemma 3 with $h(\kappa) = \mathrm{e}^{-\kappa \boldsymbol{\phi}_n^T \boldsymbol{\alpha}}$ to conclude (B13c). Therefore, the Hessian of $\mathcal{L}_\iota(\boldsymbol{\alpha})$ in (B11b) is positive semidefinite.

Appendix C
Proof of Theorem 2

According to [30], real-analytic and semialgebraic functions and their summations satisfy the KL property automatically. Therefore, the proof consists of showing the following two parts: (a) the NLL in (16) is a real-analytic function of $(\boldsymbol{\alpha}, \mathcal{I})$ on $\mathbb{C} \subseteq \mathrm{dom}(f)$ and (b) both $r(\boldsymbol{\alpha})$ in (25c) and $\mathbb{I}_{[0,+\infty)}(\mathcal{I})$ are semialgebraic functions.

Real-analytic NLL. The NLL in (16) is in the form of weighted summations of terms $\boldsymbol{b}^{\mathrm{L}}(\boldsymbol{\phi}_n^T \boldsymbol{\alpha})\mathcal{I}$, $\ln\big[\boldsymbol{b}^{\mathrm{L}}(\boldsymbol{\phi}_n^T \boldsymbol{\alpha})\mathcal{I}\big]$, and $\ln^2\big[\boldsymbol{b}^{\mathrm{L}}(\boldsymbol{\phi}_n^T \boldsymbol{\alpha})\mathcal{I}\big]$ for $n = 1, 2, \ldots, N$. Weighted summation of real-analytic functions is real-analytic; hence, we need to prove that $\ell_1(t) = \boldsymbol{b}^{\mathrm{L}}\big(\boldsymbol{\phi}^T(\boldsymbol{\alpha} + t\boldsymbol{\gamma})\big)(\mathcal{I} + t\mathcal{J}), \ell_2(t) = \ln \ell_1(t)$, and $\ell_3(t) = \ell_2^2(t)$ are real-analytic functions. Since $\big(\ell_i(t)\big)_{i=1}^3$ are smooth, it is sufficient to prove that the mth derivatives, $\ell_i^{(m)}(t)$, are bounded for all m, $(\boldsymbol{\alpha}, \mathcal{I})$, $(\boldsymbol{\gamma}, \mathcal{J})$, and t such that $(\boldsymbol{\alpha} + t\boldsymbol{\gamma}, \mathcal{I} + t\mathcal{J}) \in \mathrm{dom}(f)$.

The mth derivative of $\ell_1(t)$ is

$$\begin{aligned}\ell_1^{(m)} = {} & (\boldsymbol{\phi}^T\boldsymbol{\gamma})^m \big((-\kappa)^m \boldsymbol{b}\big)^{\mathrm{L}}(\boldsymbol{\alpha} + t\boldsymbol{\gamma})(\mathcal{I} + t\mathcal{J}) \\ & + m(\boldsymbol{\phi}^T\boldsymbol{\gamma})^{m-1}\big((-\kappa)^{m-1}\boldsymbol{b}\big)^{\mathrm{L}}(\boldsymbol{\alpha} + t\boldsymbol{\gamma})\mathcal{J} \quad \text{(C1)}\end{aligned}$$

which is bounded for any $\boldsymbol{\alpha}$, $\mathcal{I}$, $\boldsymbol{\gamma}$, $\mathcal{J}$, and t such that $(\boldsymbol{\alpha} + t\boldsymbol{\gamma}, \mathcal{I} + t\mathcal{J})$ is in one of compact subsets $\mathbb{C} \subseteq \mathrm{dom}(f)$.

For any compact set $\mathbb{C} \subseteq \mathrm{dom}(f)$, there exists $\epsilon > 0$ such that $\ell_1(t) \geq \epsilon$ for all $(\boldsymbol{\alpha} + t\boldsymbol{\gamma}, \mathcal{I} + t\mathcal{J}) \in \mathbb{C}$. $\ln(\cdot)$ and $(\cdot)^2$ are analytic on $[\epsilon, +\infty)$. Since the compositions and products of analytic functions are analytic [41, Ch. 1.4], both $\ell_2(t)$ and $\ell_3(t)$ are analytic. Therefore, the NLL in (16) is analytic.

Semialgebraic regularization terms. According to [30], i) the ℓ_2 norm $\|\cdot\|_2$ is semialgebraic, ii) the indicator function $\mathbb{I}_{[0,+\infty)}(\cdot)$ is semialgebraic, iii) finite sums and products of semialgebraic functions are semialgebraic, and iv) the composition of semialgebraic functions are semialgebraic. Therefore, $\mathbb{I}_{[0,+\infty)}(\boldsymbol{\alpha})$ and $\mathbb{I}_{[0,+\infty)}(\mathcal{I})$ are both semialgebraic. Since we can write

$$\sqrt{\sum_{j \in \mathcal{N}_i} (\alpha_i - \alpha_j)^2} = \|P_i \boldsymbol{\alpha}\|_2 \quad \text{(C2)}$$

for some matrix P_i, using i), iii), and iv) leads to semialgebraic (C2), thus semialgebraic $r(\boldsymbol{\alpha})$ in (25c). Finally, according to [30], the sum of real-analytic and semialgebraic functions satisfies the KL property. Therefore, $f(\boldsymbol{\alpha}, \mathcal{I})$ satisfies the KL property on a compact subset of dom $f(\boldsymbol{\alpha}, \mathcal{I})$.

Acknowledgment

The authors are grateful to Dr. Joseph N. Gray, Center for Nondestructive Evaluation, Iowa State University, for providing real X-ray CT data used in the numerical examples.

References

[1] W. Dewulf, Y. Tan, and K. Kiekens, "Sense and nonsense of beam hardening correction in CT metrology," *CIRP Annals*, vol. 61, no. 1, pp. 495–498, 2012.

[2] G. Wang, H. Yu, and B. De Man, "An outlook on x-ray CT research and development," *Med. Phys.*, vol. 35, no. 3, pp. 1051–1064, 2008.

[3] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. New York: IEEE Press, 1988.

[4] J. Hsieh, *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*, 2nd ed. Bellingham, WA: SPIE, 2009.

[5] J. Nuyts, B. De Man, J. A. Fessler, W. Zbijewski, and F. J. Beekman, "Modelling the physics in the iterative reconstruction for transmission computed tomography," *Phys. Med. Biol.*, vol. 58, no. 12, R63–R96, 2013.

[6] J. Barrett and N. Keat, "Artifacts in CT recognition and avoidance," *Radiographics*, vol. 24, no. 6, pp. 1679–1691, 2004.

[7] M. Krumm, S. Kasperl, and M. Franz, "Reducing nonlinear artifacts of multi-material objects in industrial 3D computed tomography," *NDT & E Int.*, vol. 41, no. 4, pp. 242–251, 2008.

[8] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 89–99, 2002.

[9] ——, "Segmentation-free statistical image reconstruction for polyenergetic X-ray computed tomography with experimental validation," *Phys. Med. Biol.*, vol. 48, no. 15, pp. 2453–2477, 2003.

[10] G. Van Gompel, K. Van Slambrouck, M. Defrise, K. Batenburg, J. de Mey, J. Sijbers, and J. Nuyts, "Iterative correction of beam hardening artifacts in CT," *Med. Phys.*, vol. 38, S36–S49, 2011.

[11] R. Gu and A. Dogandžić, "Beam hardening correction via mass attenuation discretization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Vancouver, Canada, May 2013, pp. 1085–1089.

[12] ——, (Sep. 2015). Polychromatic X-ray CT image reconstruction and mass-attenuation spectrum estimation. arXiv: 1509.02193 [stat.ME].

[13] R. Redus, J. Pantazis, T. Pantazis, A. Huber, and B. Cross, "Characterization of CdTe detectors for quantitative X-ray spectroscopy," *IEEE Trans. Nucl. Sci.*, vol. 56, no. 4, pp. 2524–2532, 2009.

[14] Y. Lin, J. C. Ramirez-Giraldo, D. J. Gauthier, K. Stierstorfer, and E. Samei, "An angle-dependent estimation of CT x-ray spectrum from rotational transmission measurements," *Med. Phys.*, vol. 41, no. 6, p. 062 104, 2014.

[15] R. Gu and A. Dogandžić, "Polychromatic sparse image reconstruction and mass attenuation spectrum estimation via B-spline basis function expansion," in *Rev. Prog. Quant. Nondestr. Eval.*, D. E. Chimenti and L. J. Bond, Eds., ser. AIP Conf. Proc. Vol. 34 1650, Melville, NY, 2015, pp. 1707–1716.

[16] J. Xu and B. M. Tsui, "Quantifying the importance of the statistical assumption in statistical X-ray CT image

reconstruction," *IEEE Trans. Med. Imag.*, vol. 33, no. 1, pp. 61–73, 2014.

[17] G. M. Lasio, B. R. Whiting, and J. F. Williamson, "Statistical reconstruction for X-ray computed tomography using energy-integrating detectors," *Phys. Med. Biol.*, vol. 52, no. 8, p. 2247, 2007.

[18] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 123–231, 2013.

[19] G. T. Herman, "Correction for beam hardening in computed tomography," *Phys. Med. Biol.*, vol. 24, no. 1, pp. 81–106, 1979.

[20] L. L. Schumaker, *Spline Functions: Basic Theory*, 3rd ed. New York: Cambridge Univ. Press, 2007.

[21] L. Zanni, A. Benfenati, M. Bertero, and V. Ruggiero, "Numerical methods for parameter estimation in Poisson data inversion," *J. Math. Imaging Vis.*, vol. 52, no. 3, pp. 397–413, 2015.

[22] P. McCullagh and J. Nelder, *Generalized Linear Models*, 2nd ed. New York: Chapman & Hall, 1989.

[23] J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: A survey and extensions," *Math. Methods Oper. Res.*, vol. 66, no. 3, pp. 373–407, 2007.

[24] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, 2009.

[25] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality," *Math. Oper. Res.*, vol. 35, no. 2, pp. 438–457, May 2010.

[26] R. A. Thisted, *Elements of Statistical Computing*. New York: Chapman & Hall, 1989.

[27] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," in *Sov. Math. Dokl.*, vol. 27, 1983, pp. 372–376.

[28] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[29] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, 1995.

[30] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.

[31] R. Gu and A. Dogandžić, "Projected Nesterov's proximal-gradient signal recovery from compressive Poisson measurements," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 2015, to appear.

[32] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. Numer. Anal.*, vol. 8, no. 1, pp. 141–148, 1988.

[33] B. O'Donoghue and E. Candès, "Adaptive restart for accelerated gradient schemes," *Found. Comput. Math.*, pp. 1–18, Jul. 2013.

[34] J. Liu, X. Zhang, X. Zhang, H. Zhao, Y. Gao, D. Thomas, D. A. Low, and H. Gao, "5D respiratory motion model based image reconstruction algorithm for 4D cone-beam computed tomography," *Inverse Prob.*, vol. 31, no. 11, p. 115 007, 2015.

[35] A. Dogandžić, R. Gu, and K. Qiu, "Mask iterative hard thresholding algorithms for sparse image reconstruction of objects with known contour," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 2011, pp. 2111–2116.

[36] J. H. Hubbell and S. M. Seltzer, "Tables of X-ray mass attenuation coefficients and mass energy-absorption coefficients 1 keV to 20 MeV for elements $Z = 1$ to 92 and 48 additional substances of dosimetric interest," National Inst. Standards Technol., Ionizing Radiation Div., Gaithersburg, MD, Tech. Rep. NISTIR 5632, 1995.

[37] J. M. Boone and J. A. Seibert, "An accurate method for computer-generating tungsten anode X-ray spectra from 30 to 140 kV," *Med. Phys.*, vol. 24, no. 11, pp. 1661–1670, 1997.

[38] J. Llacer and E. Veklerov, "Feasible images and practical stopping rules for iterative algorithms in emission tomography," *IEEE Trans. Med. Imag.*, vol. 8, no. 2, pp. 186–193, 1989.

[39] R. Zhang, J.-B. Thibault, C. Bouman, K. Sauer, and J. Hsieh, "Model-based iterative reconstruction for dual-energy X-ray CT using a joint quadratic likelihood model," *IEEE Trans. Med. Imag.*, vol. 33, no. 1, pp. 117–134, Jan. 2014.

[40] W. Huda, *Review of Radiologic Physics*, 3rd ed. Baltimore, MD: Lippincott Williams & Wilkins, 2010.

[41] S. Krantz and H. Parks, *A Primer of Real Analytic Functions*, 2nd ed. Boston, MA: Birkhäuser, 2002.