Polychromatic Sparse Image Reconstruction and Mass Attenuation Spectrum Estimation via B-Spline Basis Function Expansion

Renliang Gu[a] and Aleksandar Dogandžić[b]

Iowa State University, Center for Nondestructive Evaluation, 1915 Scholl Road, Ames, IA 50011, USA

[a]renliang@iastate.edu [b]corresponding author: ald@iastate.edu

Abstract. We develop a sparse image reconstruction method for polychromatic computed tomography (CT) measurements under the blind scenario where the material of the inspected object and the incident energy spectrum are unknown. To obtain a parsimonious measurement model parameterization, we first rewrite the measurement equation using our mass-attenuation parameterization, which has the Laplace integral form. The unknown mass-attenuation spectrum is expanded into basis functions using a B-spline basis of order one. We develop a block coordinate-descent algorithm for constrained minimization of a penalized negative log-likelihood function, where constraints and penalty terms ensure nonnegativity of the spline coefficients and sparsity of the density map image in the wavelet domain. This algorithm alternates between a Nesterov's proximal-gradient step for estimating the density map image and an active-set step for estimating the incident spectrum parameters. Numerical simulations demonstrate the performance of the proposed scheme.

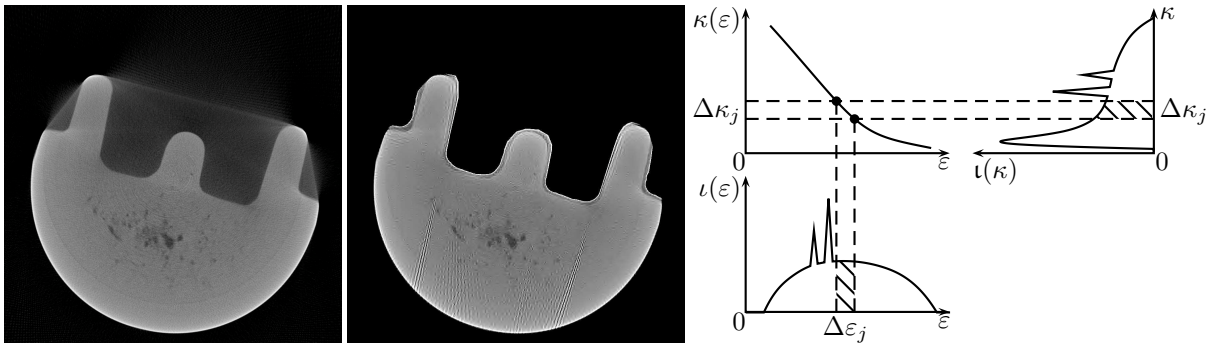
BACKGROUND AND INTRODUCTION

X-ray computed tomography (CT) measurement systems are important in modern nondestructive evaluation (NDE) and medical diagnostics. Therefore, improving reconstruction accuracy and speed of data collection in these systems could have a significant impact on these broad areas. 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, linear reconstructions such as filtered backprojection (FBP) exhibit beam hardening artifacts, e.g., cupping and streaking [1]: the FBP reconstruction in Fig. 1a shows decreasing material density towards the center of the inspected object and existence of nonzero object density in the 'bay area' of the object where the true density is zero. These artifacts limit the quantitative analysis of the reconstruction. In medical CT application, severe artifacts can even look similar to certain pathologies and further mislead the diagnosis [2, 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 as well. 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 via sparse signal reconstruction. Fig. 1b shows the reconstructed image by the mask double overrelaxation (mask DORE) method that is also based on the linear measurement model, but with additional geometric object shape information incorporated, see [3] for details and description of other geometric mask methods that employ the linear measurement model. (Here, we show mask DORE for illustration: other mask approaches in [3] exhibit similar artifacts.) Generally, adding geometric information will help the reconstruction. However, in Fig. 1b, the non-convex mask cuts into the convex hull of the object, and the nonzero signal outside the object is pushed inside the mask, which results in energy accumulation along the mask border and artifacts inside the object as well. Hence, to reap the benefits of signal sparsity and object contour information, we must correct the beam hardening effect first by accounting for the polychromatic source model.

In this paper, we adopt the nonlinear measurement model resulting from the polychromatic X-ray source, and simplify it by exploiting the relation between the *mass attenuation coefficients*, *X-ray photon energy* and *incident spectrum*, see Fig. 1c. Based on this idea, the optimization problem formulated from the image reconstruction becomes feasible and gives the estimation of the density map and the mass attenuation spectrum as the result.

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 (replaced by $I, \mathbf{1}$, and $\mathbf{0}$ when the dimensions can be inferred easily); $|\cdot|$, $\|\cdot\|_2$, $\otimes$, and "^{T}" are the absolute value, ℓ_p norm, Kronecker product, and transpose, respectively. Furthermore, $\lfloor x \rfloor$ is the largest integer smaller than or equal to x,



(a) (b) (c)

FIGURE 1. (a)–(b) FBP and mask DORE linear reconstructions from polychromatic X-ray CT measurements and (c) relation between mass attenuation κ, incident spectrum ι and photon energy ε.

"$\succeq$" is the elementwise version of "$\geq$", $\mathbb{I}_+(\boldsymbol{\alpha})$ is the nonnegativity indicator function:

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

and $\boldsymbol{a}^{\mathrm{L}}(s)$ is the *Laplace transform* of a vector function $\boldsymbol{a}(\kappa)$: $\boldsymbol{a}^{\mathrm{L}}(s) \triangleq \int \boldsymbol{a}(\kappa)\mathrm{e}^{-s\kappa}\,\mathrm{d}\kappa$. Note that $(\kappa\boldsymbol{a})^{\mathrm{L}}(s) = \int \kappa\boldsymbol{a}(\kappa)\mathrm{e}^{-s\kappa}\,\mathrm{d}\kappa$. Define also the elementwise logarithm $\ln_\circ(\boldsymbol{x}) = [\ln x_1, \ldots, \ln x_N]^T$, nonnegativity projector $[(\boldsymbol{x})_+]_i = \max\{x_i, 0\}$ where $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]^T$, and Laplace transforms $\boldsymbol{a}^{\mathrm{L}}_\circ(\boldsymbol{s}) = \left[[\boldsymbol{a}^{\mathrm{L}}(s_1)]^T \quad [\boldsymbol{a}^{\mathrm{L}}(s_2)]^T \quad \cdots \quad [\boldsymbol{a}^{\mathrm{L}}(s_N)]^T\right]^T$ and $(\kappa\boldsymbol{a})^{\mathrm{L}}_\circ(\boldsymbol{s}) = \left[[(\kappa\boldsymbol{a})^{\mathrm{L}}(s_1)]^T \quad [(\kappa\boldsymbol{a})^{\mathrm{L}}(s_2)]^T \quad \cdots \quad [(\kappa\boldsymbol{a})^{\mathrm{L}}(s_N)]^T\right]^T$ where $\boldsymbol{s} = [s_1, s_2, \ldots, s_N]^T$.

In the following section, we review the noiseless polychromatic X-ray CT measurement model and our mass-attenuation parameterization of this model.

POLYCHROMATIC X-RAY CT MODEL

To describe the polychromatic X-ray source, assume that its incident intensity $\mathcal{I}^{\mathrm{in}}$ spreads along photon energy ε following the density $\iota(\varepsilon) \geq 0$, i.e.,

$$\int \iota(\varepsilon)\,\mathrm{d}\varepsilon = \mathcal{I}^{\mathrm{in}}. \tag{2a}$$

see Fig. 1c, which shows a typical density $\iota(\varepsilon)$. The noiseless measurement collected by an energy integral detector upon traversing a straight line $\ell = \ell(x, y)$ has the familiar superposition-integral form:

$$\mathcal{I}^{\mathrm{out}} = \int \iota(\varepsilon) \exp\left[-\int_\ell \mu(x, y, \varepsilon)\,\mathrm{d}\ell\right]\mathrm{d}\varepsilon = \int \iota(\varepsilon) \exp\left[-\kappa(\varepsilon)\int_\ell \alpha(x, y)\,\mathrm{d}\ell\right]\mathrm{d}\varepsilon \tag{2b}$$

where we model the attenuation $\mu(x, y, \varepsilon)$ of the inspected object consisting of a single material using the following *separable form* [4, Sec. 6]:

$$\mu(x, y, \varepsilon) = \kappa(\varepsilon)\alpha(x, y). \tag{3}$$

Here, $\kappa(\varepsilon)$ is the *mass attenuation coefficient* of the material, a function of the photon energy ε (illustrated in Fig. 1c) and $\alpha(x, y)$ 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 [2, 5] 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 $\kappa(\varepsilon)$.

Mass Attenuation Parameterization

We now review our parsimonious parameterization of (2b) for signal reconstruction [6, 7]. Since the mass attenuation $\kappa(\varepsilon)$ and incident spectrum density $\iota(\varepsilon)$ are both functions of ε (see Fig. 1c), we combine the variations of these two

functions and write (2a) and (2b) as integrals of κ rather than ε, with goal 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 [6]. Hence, we express (2a) and (2b) as

$$\mathcal{I}^{\text{in}} = \iota^{\text{L}}(0), \qquad \mathcal{I}^{\text{out}} = \iota^{\text{L}}\left(\int_{\ell} \alpha(x,y)\,\mathrm{d}\ell\right) \tag{4}$$

where $\iota^{\text{L}}(s)$ is the Laplace transform of $\iota(\kappa)$. Here, $s > 0$, in contrast with the traditional Laplace transform where s is generally complex and $\iota(\kappa)$ represents the density of the incident X-ray energy at attenuation κ; we refer to it as incident *mass attenuation spectrum*. The mass attenuation spectrum depends both on the measurement system (through the incident energy spectrum) and inspected object (through the mass attenuation of the inspected material). As shown in Fig. 1c, with respect to this particular material described by the mass attenuation curve, the area $\iota(\varepsilon_j)\Delta\varepsilon_j$ depicting the X-ray energy within the $\Delta\varepsilon_j$ slot is the same as the corresponding area $\iota(\kappa_j)\Delta\kappa_j$, the amount of X-ray energy that attenuated at a rate within $\Delta\kappa_j$ slot.

In the blind scenario where the inspected material and incident signal spectrum are unknown, the above parameterization allows us to estimate two functions [$\iota(\kappa)$ and $\alpha(x,y)$] rather than three [$\iota(\varepsilon),\kappa(\varepsilon)$, and $\alpha(x,y)$].

MEASUREMENT MODEL AND PARAMETER CONSTRAINTS

Observe that the mass attenuation spectrum $\iota(\kappa)$ and inspected object's density $\alpha(x,y)$ are nonnegative for all κ, x and y and assume that the shadow of the inspected object is completely covered by the receiver array and that the upper bound $\mathcal{I}^{\text{in}}_{\text{MAX}}$ on incident X-ray energy is known (obtained, e.g., from energy measurements at the detectors that have a line-of-sight view of the X-ray source); hence,

$$\iota(\kappa) \geq 0, \qquad \alpha(x,y) \geq 0 \qquad \mathcal{I}^{\text{in}} \leq \mathcal{I}^{\text{in}}_{\text{MAX}}. \tag{5}$$

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

$$\int_{\ell}\alpha(x,y)\,\mathrm{d}\ell \approx \boldsymbol{\phi}^T\boldsymbol{\alpha} \tag{6}$$

where $\boldsymbol{\alpha}$ is a $p\times 1$ column vector representing the 2D image that we wish to reconstruct [i.e., discretized $\alpha(x,y)$] and $\boldsymbol{\phi}$ is a $p\times 1$ vector of 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. For the ith measurement, define its discretized line integral as $\boldsymbol{\phi}_i^T\boldsymbol{\alpha}$; stacking all N such integrals into a vector yields $\Phi\boldsymbol{\alpha}$, where $\Phi = [\boldsymbol{\phi}_1\boldsymbol{\phi}_2\ldots\boldsymbol{\phi}_N]^T$ is the $N\times p$ Radon transform matrix for our imaging system.

Express $\iota(\kappa)$ as a linear combination of J basis functions

$$\iota(\kappa) = \boldsymbol{b}(\kappa)\boldsymbol{\mathcal{I}} \tag{7}$$

where $\boldsymbol{\mathcal{I}} = [\mathcal{I}_1,\mathcal{I}_2,\ldots,\mathcal{I}_J]^T$ is the $J\times 1$ vector of corresponding basis function coefficients and substitute (6) and (7) into (4) for each of the N measurements, which yields the following expressions for the incident energy and the $N\times 1$ vector of noiseless measurements:

$$\mathcal{I}^{\text{in}}(\boldsymbol{\mathcal{I}}) = \boldsymbol{b}^{\text{L}}(0)\boldsymbol{\mathcal{I}}, \qquad \boldsymbol{\mathcal{I}}^{\text{out}}(\boldsymbol{\alpha},\boldsymbol{\mathcal{I}}) = \boldsymbol{b}^{\text{L}}_{\circ}(\Phi\boldsymbol{\alpha})\boldsymbol{\mathcal{I}}. \tag{8}$$

Here, we select the row vector function $\boldsymbol{b}(\kappa) = \left[b_1(\kappa), b_2(\kappa), \ldots, b_J(\kappa)\right]$ as B-splines [8] of order one (termed B1 splines, illustrated in Fig. 2a) because, in this case, the decomposition (7) yields nonnegative elements of the spline coefficients $\boldsymbol{\mathcal{I}}$ [based on (5)] and thus allows us to impose the physically meaningful nonnegativity constraint when estimating $\boldsymbol{\mathcal{I}}$. Note that $\boldsymbol{b}^{\text{L}}_{\circ}(\Phi\boldsymbol{\alpha})$ is an $N\times J$ matrix. In this paper, we adopt a geometric series with common ratio $q > 1$ as the knots of our B1 splines:

$$b_j(\kappa) = \begin{cases} \dfrac{\kappa - q^{j-1}\kappa_0}{(q-1)q^{j-1}\kappa_0}, & q^{j-1}\kappa_0 \leq \kappa < q^j\kappa_0 \\ \dfrac{-\kappa + q^{j+1}\kappa_0}{(q-1)q^j\kappa_0}, & q^j\kappa_0 \leq \kappa < q^{j+1}\kappa_0 \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

so that the jth basis function can be obtained simply by q-scaling the $(j-1)$th basis function (as illustrated in Fig. 2a):

$$b_j(\kappa) = b_{j-1}\left(\frac{\kappa}{q}\right), \qquad q > 1. \tag{10}$$

Ambiguity of the density map and mass attenuation spectrum. We now discuss density map scaling ambiguity under the blind scenario where both the density map $\alpha(x,y)$ and incident spectrum parameters $\boldsymbol{\mathcal{I}}$ are unknown. By noting (10) and the 'time scaling' property of the Laplace transform, $\boldsymbol{b}(\kappa/q) \overset{\mathcal{L}}{\rightarrow} q\boldsymbol{b}^{\mathrm{L}}(qs)$ for $q > 0$, we conclude that selecting q times narrower basis functions $\left[b_0(\kappa), b_1(\kappa), \ldots, b_{J-1}(\kappa)\right]$ than those in $\boldsymbol{b}(\kappa)$ and q times larger density map and spectral parameters [$q\alpha(x,y)$ and $q\boldsymbol{\mathcal{I}}$] yields the same output photon energy $\mathcal{I}^{\mathrm{out}}$.

Linear inequality constraints on $\boldsymbol{\mathcal{I}}$. Combining the nonnegativity of the elements of $\boldsymbol{\mathcal{I}}$ and the incident-energy upper bound constraint $\boldsymbol{b}^{\mathrm{L}}(0)\boldsymbol{\mathcal{I}} \le \mathcal{I}^{\mathrm{in}}_{\mathrm{MAX}}$ [see (5)] yields

$$C\boldsymbol{\mathcal{I}} \succeq \boldsymbol{c}, \qquad C \triangleq \begin{bmatrix} I_J & -\dfrac{\left[\boldsymbol{b}^{\mathrm{L}}(0)\right]^T}{\|\boldsymbol{b}^{\mathrm{L}}(0)\|_2} \end{bmatrix}^T, \quad \boldsymbol{c} \triangleq \begin{bmatrix} \mathbf{0}_J^T & -\dfrac{\mathcal{I}^{\mathrm{in}}_{\mathrm{MAX}}}{\|\boldsymbol{b}^{\mathrm{L}}(0)\|_2} \end{bmatrix}^T. \tag{11}$$

The rows of C are normalized, for numerical stability.

PARAMETER ESTIMATION

Consider an $N \times 1$ vector $\boldsymbol{\mathcal{E}}$ of energy measurements corrupted by lognormal noise and the corresponding squared-error cost function [see (8)]:

$$\mathcal{L}(\Phi\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) = \frac{1}{2}\left\| \ln_{\circ}(\boldsymbol{\mathcal{E}}) - \ln_{\circ}\left[\mathcal{I}^{\mathrm{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})\right] \right\|_2^2 \tag{12a}$$

Note that (12a) is bi-convex [9] with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\mathcal{I}}$ respectively under the following sufficient condition:

$$-\frac{2q^2}{(q^2-1)^2}\mathbf{1} \preceq \ln_{\circ}\left[\mathcal{I}^{\mathrm{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})\right] - \ln_{\circ}\boldsymbol{\mathcal{E}} \preceq \mathbf{1} \tag{12b}$$

where the inequalities on the left and right sides are for the convexity over $\boldsymbol{\alpha}$ and $\boldsymbol{\mathcal{I}}$, respectively.

We formulate the following constrained penalized optimization problem:

$$\min_{\substack{\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}} \\ C\boldsymbol{\mathcal{I}} \succeq \boldsymbol{c}}} \mathcal{L}(\Phi\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}) + ur(\boldsymbol{\alpha}) \tag{13a}$$

where $u > 0$ is a scalar tuning constant and the regularization term

$$r(\boldsymbol{\alpha}) = \|\Psi^T\boldsymbol{\alpha}\|_1 + \mathbb{I}_+(\boldsymbol{\alpha}) \tag{13b}$$

enforces nonnegativity [based on (5)] and sparsity (in the wavelet-transform domain) of the signal $\boldsymbol{\alpha}$, with sparsifying transform matrix Ψ^T that satisfies

$$\Psi\Psi^T = I_p. \tag{14}$$

In this paper, we use the full circular mask in [3] and select Ψ accordingly as a submatrix of the inverse discrete wavelet transform (DWT) matrix [3].

Our goal is to estimate the image and incident energy density parameters $(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}})$. We descend the objective function (13a) by alternating between Step 1) and Step 2), where Iteration $i+1$ proceeds as follows:

Step 1) *Nesterov's proximal-gradient (NPG) step* [10] for $\boldsymbol{\alpha}$ with fixed $\boldsymbol{\mathcal{I}} = \boldsymbol{\mathcal{I}}^{(i)}$, yielding $\boldsymbol{\alpha}^{(i+1)}$,

Step 2) if the right side hand of condition (12b) holds, use the *active-set (AS)* step [11, Ch. 5.2] for $\boldsymbol{\mathcal{I}}$ with fixed $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(i+1)}$ to yield $\boldsymbol{\mathcal{I}}^{(i+1)}$, otherwise let $\boldsymbol{\mathcal{I}}^{(i+1)} = \boldsymbol{\mathcal{I}}^{(i)}$.

Iterate between Steps 1) and 2) until convergence; we select the following convergence criterion:

$$\max\{\delta_{\boldsymbol{\alpha}}^{(i)}, \delta_{\boldsymbol{\mathcal{I}}}^{(i)}\} < \epsilon \tag{15a}$$

where ϵ is the convergence threshold and

$$\delta_{\boldsymbol{\alpha}}^{(i)} \triangleq \frac{\|\boldsymbol{\alpha}^{(i)} - \boldsymbol{\alpha}^{(i-1)}\|_2}{\|\boldsymbol{\alpha}^{(i)}\|_2}, \qquad \delta_{\boldsymbol{\mathcal{I}}}^{(i)} \triangleq \frac{\|\boldsymbol{\mathcal{I}}^{(i)} - \boldsymbol{\mathcal{I}}^{(i-1)}\|_2}{\|\boldsymbol{\mathcal{I}}^{(i)}\|_2}. \tag{15b}$$

The above iteration is the first physical-model based image reconstruction method (in addition to [6, 7]) for simultaneous *blind* (assuming unknown incident X-ray spectrum and unknown materials) sparse image reconstruction from polychromatic measurements. In [6, 7], we applied a piecewise-constant (B0 spline) expansion of the mass attenuation spectrum, approximated Laplace integrals with Riemann sums, and used a smooth approximation of both the ℓ_1 and nonnegativity penalties in (13b), see also [12] for the ℓ_1 norm smoothing. Here, we use B1 splines, evaluate the Laplace integrals exactly, and employ the exact ℓ_1 norm instead of its approximation, thus eliminating one tuning constant and improving the reconstruction performance compared with [6, 7]. Unlike [6, 7], we impose the nonnegativity constraint on the density map $\boldsymbol{\alpha}$ *directly* and hence eliminate one more tuning constant compared with the methods in these references.

Step 2) aims at minimizing the squared-error cost function $\mathcal{L}(\Phi\boldsymbol{\alpha}^{(i+1)}, \boldsymbol{\mathcal{I}})$ with respect to $\boldsymbol{\mathcal{I}}$ subject to the linear inequality constraints (11). [Note that the regularization term in (13a) is constant with respect to $\boldsymbol{\mathcal{I}}$, and is thus omitted in our optimization problem.] Here, we apply the *gradient-projection active-set method* to perform this optimization, described in detail in [11].

We initialize the above algorithm by

$$\boldsymbol{\alpha}^{(-1)} = \widehat{\boldsymbol{\alpha}}_{\text{FBP}} \quad \text{and} \quad \mathcal{I}_j^{(0)} = \begin{cases} \mathcal{I}_{\text{MAX}}^{\text{in}} / b_j^{\text{L}}(0), & j = \lfloor J/2 \rfloor + 1 \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

where $\widehat{\boldsymbol{\alpha}}_{\text{FBP}}$ is the standard FBP reconstruction (without linearization [5, Ch. 3], see also the numerical examples). Plugging in the initialization (16) into (8), we have $\mathcal{I}^{\text{in}}(\boldsymbol{\mathcal{I}}^{(-1)}) = \mathcal{I}_{\text{MAX}}^{\text{in}}$ and the initial estimate $(\boldsymbol{\alpha}^{(-1)}, \boldsymbol{\mathcal{I}}^{(0)})$ corresponds to the monochromatic X-ray model with only one spectrum component. It is also desirable to have the main lobe of the estimated spectrum at the center, which is why the nonzero element of $\boldsymbol{\mathcal{I}}^{(0)}$ is placed in the middle position.

Step 1) in Iteration $i+1$: Updating $\boldsymbol{\alpha}$ Using NPG with Restart

Step 1) aims at minimizing $l^{(i)}(\boldsymbol{\alpha}) + u\,r(\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$, where

$$l^{(i)}(\boldsymbol{\alpha}) \triangleq \mathcal{L}(\Phi\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}^{(i)}) \tag{17}$$

with gradient $\nabla l^{(i)}(\boldsymbol{\alpha}) = \Phi^T \operatorname{diag}\Big[\big[(\kappa\boldsymbol{b})_{\circ}^{\text{L}}(\Phi\boldsymbol{\alpha})\big]\boldsymbol{\mathcal{I}}\Big] \operatorname{diag}^{-1}\big[\boldsymbol{\mathcal{I}}^{\text{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}^{(i)})\big] \Big(\ln_{\circ}(\boldsymbol{\mathcal{E}}) - \ln_{\circ}\big[\boldsymbol{\mathcal{I}}^{\text{out}}(\boldsymbol{\alpha}, \boldsymbol{\mathcal{I}}^{(i)})\big]\Big)$. Here, $(\kappa\boldsymbol{b})_{\circ}^{\text{L}}(\Phi\boldsymbol{\alpha})$ is an $N \times J$ matrix. We implement Step 1) via the *NPG scheme* [13]:

$$\theta^{(i+1)} = \frac{1}{2}\left[1 + \sqrt{1 + 4\big(\theta^{(i)}\big)^2}\right] \tag{18a}$$

$$\overline{\boldsymbol{\alpha}}^{(i+1)} = \boldsymbol{\alpha}^{(i)} + \frac{\theta^{(i)} - 1}{\theta^{(i+1)}}\big(\boldsymbol{\alpha}^{(i)} - \boldsymbol{\alpha}^{(i-1)}\big) \tag{18b}$$

$$\boldsymbol{\alpha}^{(i+1)} = \operatorname{prox}_{\beta^{(i)} u r}\Big(\overline{\boldsymbol{\alpha}}^{(i+1)} - \beta^{(i)} \nabla l^{(i)}\big(\overline{\boldsymbol{\alpha}}^{(i+1)}\big)\Big) \tag{18c}$$

where

$$\operatorname{prox}_{\lambda r}(\boldsymbol{a}) = \arg\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\boldsymbol{\alpha} - \boldsymbol{a}\|_2^2 + \lambda r(\boldsymbol{\alpha}) \tag{19}$$

is the proximal operator for scaled (by $\lambda > 0$) regularization term (13b) [14]. We select the step size $\beta^{(i)}$ to satisfy the majorization condition:

$$l^{(i)}(\boldsymbol{\alpha}^{(i+1)}) \le l^{(i)}(\overline{\boldsymbol{\alpha}}^{(i+1)}) + \big(\boldsymbol{\alpha}^{(i+1)} - \overline{\boldsymbol{\alpha}}^{(i+1)}\big)^T \nabla l^{(i)}\big(\overline{\boldsymbol{\alpha}}^{(i+1)}\big) + \frac{1}{2\beta^{(i)}}\big\|\boldsymbol{\alpha}^{(i+1)} - \overline{\boldsymbol{\alpha}}^{(i+1)}\big\|_2^2 \tag{20}$$

using the following adaptation scheme [15]:

a) In Iteration i: if there has been no step size reductions for n consecutive iterations (implying $i > n$), i.e., $\beta^{(i-1)} = \beta^{(i-2)} = \cdots = \beta^{(i-n-1)}$, start with a larger step size $\beta^{(i)} = \beta^{(i-1)}/\xi$, $\xi \in (0,1)$; otherwise start with the previous step size $\beta^{(i)} = \beta^{(i-1)}$.

b) *Backtrack* using the same multiplicative scaling constant ξ, with goal to find the largest $\beta^{(i)}$ that satisfies (20).

We select the initial step size $\beta^{(0)}$ using the Barzilai-Borwein method [16].

(Re)start. We initialize the NPG step (18) as follows and [13, 17]:

$$\theta^{(0)} = 0, \qquad \boldsymbol{\alpha}^{(0)} = \mathbf{0} \tag{21a}$$

which implies that $\bar{\boldsymbol{\alpha}}^{(1)} = \boldsymbol{\alpha}^{(-1)}$ [see (16)] and leads to two consecutive proximal-gradient (PG) steps:

$$\boldsymbol{\alpha}^{(1)} = \mathrm{prox}_{\beta^{(0)}ur}\left(\boldsymbol{\alpha}^{(-1)} - \beta^{(0)}\nabla l^{(0)}\big(\boldsymbol{\alpha}^{(-1)}\big)\right) \quad \text{and} \quad \boldsymbol{\alpha}^{(2)} = \mathrm{prox}_{\beta^{(1)}ur}\left(\boldsymbol{\alpha}^{(1)} - \beta^{(1)}\nabla l^{(1)}\big(\boldsymbol{\alpha}^{(1)}\big)\right) \tag{21b}$$

ensuring the monotonicity of our initial iterates: $l^{(0)}(\boldsymbol{\alpha}^{(1)}) \le l^{(0)}(\boldsymbol{\alpha}^{(-1)})$ and $l^{(1)}(\boldsymbol{\alpha}^{(2)}) \le l^{(1)}(\boldsymbol{\alpha}^{(1)})$ [13]. Here, $\boldsymbol{\alpha}^{(0)}$ is needed solely for determining the convergence criterion of the corresponding inner proximal-mapping iteration, discussed in the following section.

Direct NPG iteration is not guaranteed to decrease monotonically the objective function that it minimizes due to the momentum term, see examples in [18]. O'Donoghue and Candès [18] restart the acceleration scheme by resetting the momentum term to zero and take the current $\boldsymbol{\alpha}^{(i)}$ as the new initial value for the acceleration scheme. Here we adopt the *function restart* version in [18], i.e., restart the NPG iteration steps where the objective function (17) increases. Suppose that the objective function increase occurs at the ith iteration; then, the restart is equivalent to setting $\theta^{(i)} = 0$, consistent with (21a) where 0 is replaced by i.

Proximal Mapping

Rewrite the objective function in (19) as $\min\limits_{\substack{\boldsymbol{\alpha},\boldsymbol{z}\\ \boldsymbol{\alpha}=\boldsymbol{z}}} \frac{1}{2}\|\boldsymbol{\alpha}-\boldsymbol{a}\|_2^2 + \mathbb{I}_+(\boldsymbol{\alpha}) + \lambda\|\Psi^T\boldsymbol{z}\|_1$, which leads to the following (inner) alternating direction method of multipliers (ADMM) iteration:

$$\boldsymbol{z}^{(k+1)} = \Psi\mathcal{T}_{\lambda/\rho}\big(\Psi^T(\boldsymbol{\alpha}^{(k)} - \boldsymbol{v}^{(k)})\big) \tag{22a}$$

$$\boldsymbol{\alpha}^{(k+1)} = \frac{1}{1+\rho}\big(\boldsymbol{a} + \rho(\boldsymbol{z}^{(k+1)} + \boldsymbol{v}^{(k)})\big)_+ \tag{22b}$$

$$\boldsymbol{v}^{(k+1)} = \boldsymbol{v}^{(k)} + \boldsymbol{z}^{(k+1)} - \boldsymbol{\alpha}^{(k+1)} \tag{22c}$$

where $k \ge 0$ indexes the (inner) iteration steps and $\rho > 0$ is a quadratic penalty parameter, usually set to one [19, Sec. 3.4]. Here, the orthonormality of the rows of the transform matrix Ψ (14) is needed to derive (22). If (14) does not hold, (22) can be replaced by linearized ADMM [14, Ch. 4.4.2]. We initialize the above iterations using $\boldsymbol{\alpha}^{(0)} = (\boldsymbol{a})_+$ and $\boldsymbol{v}^{(0)} = \boldsymbol{a} - (\boldsymbol{a})_+$.

Convergence criterion. Denote by i and k the outer and inner iteration indices corresponding to the NPG and ADMM iterations, respectively, and by $\boldsymbol{\alpha}^{(i,k)}$ and $\boldsymbol{z}^{(i,k)}$ the iterates of $\boldsymbol{\alpha}$ and $\boldsymbol{z}$ in the kth (inner) ADMM iteration step within the ith step of the (outer) NPG iteration (18). We set the following convergence criterion for the inner iteration:

$$\max\left\{\frac{\|\boldsymbol{z}^{(i,k)} - \boldsymbol{z}^{(i,k-1)}\|_2}{\|\boldsymbol{z}^{(i,k)}\|_2}, \frac{\|\boldsymbol{\alpha}^{(i,k)} - \boldsymbol{\alpha}^{(i,k-1)}\|_2}{\|\boldsymbol{\alpha}^{(i,k)}\|_2}\right\} < \eta\delta_{\boldsymbol{\alpha}}^{(i-1)}, \quad i \ge 1 \tag{23}$$

where the tuning constant $0 < \eta < 1$ trades the accuracy and speed of the inner iteration, with goal to provide sufficiently accurate PG steps (18c), see also (15b). Once (23) holds for some index $K_i \ge 1$, the ADMM algorithm returns $\boldsymbol{\alpha}^{(i)} \triangleq \boldsymbol{\alpha}^{(i,K_i)}$.

NUMERICAL EXAMPLES

We construct a simulation example using the binary 1024×1024 image in Fig. 2b (denoted $\boldsymbol{\alpha}_{\text{true}}$) as the 'ground truth'; here, the inspected object contains irregularly shaped inclusions. We simulated the polychromatic sinogram using the

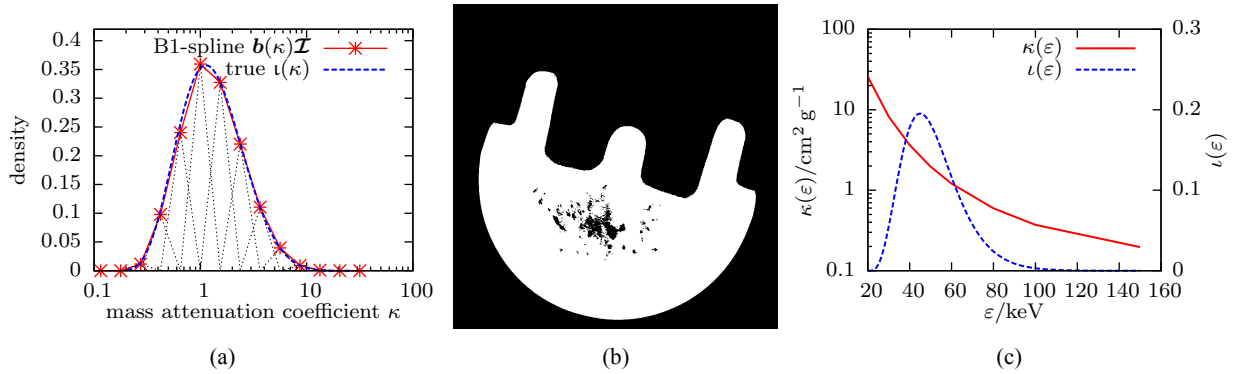


FIGURE 2. (a) B1-spline approximation of $\iota(\kappa)$, (b) 'ground truth' image used to generate the sinogram, and (c) mass attenuation and incident X-ray spectrum as functions of the photon energy ε.

interpolated mass attenuation of iron for the inspected object [20] and incident spectrum in Fig. 2c:

$$\iota(\varepsilon) = \begin{cases} \mathrm{Gamma}\left(\frac{4(\varepsilon-20)}{25} \,\middle|\, 5, 1\right), & 20 \le \varepsilon \le 150 \\ 0, & \text{otherwise} \end{cases} \tag{24}$$

where $\mathrm{Gamma}(x \mid \alpha, \beta)$ denotes the Gamma probability density function [21, Sec. 3.3]. Our simulated approximation of the integral (2b) uses 130 equi-spaced discretization points over the range 20 keV to 150 keV. We construct the Radon transform matrix Φ directly on GPU (multi-thread version on CPU is also available) with full circular mask [3]. For each of the 180 projections, we obtain the measurements from an energy detector array of size 1024 and reconstruct images of size 1024×1024.

Linearization. Observe that $\mathcal{I}^{\mathrm{out}}$ in (2b) and (4) is a *decreasing function* of the line integral $\int_\ell \alpha(x,y)\,\mathrm{d}\ell$, which is a key insight behind various linearization methods popular in the literature and practice, see [4, Sec. 6]. If we know exactly the incident spectrum and mass attenuation (material), we can determine the mass attenuation spectrum $\iota(\kappa)$ and its Laplace transform $\iota^{\mathrm{L}}(s)$ exactly as well. Here, we refer to applying the inverse function of $\iota^{\mathrm{L}}(s)$ to the elements of the noisy measurement vector $\boldsymbol{\mathcal{E}}$ as *linearization*. However, such a zero-forcing approach to reconstruction ignores noise and therefore leads to noise enhancement. In this example, we use noiseless measurements, i.e., $(\iota^{\mathrm{L}})_\circ^{-1}(\boldsymbol{\mathcal{E}}) = \Phi\alpha$ and the linear reconstruction methods with linearization are equivalent to these methods applied directly to noiseless linear measurements $\Phi\alpha$.

We compare

- the traditional FBP method without [5, Ch. 3] and with linearization [22], i.e., based on the 'data'

$$\boldsymbol{y} = -\ln_\circ(\boldsymbol{\mathcal{E}}) \ \text{(without linearization)} \qquad \text{and} \qquad \boldsymbol{y} = \left(\iota^{\mathrm{L}}\right)_\circ^{-1}(\boldsymbol{\mathcal{E}}) \ \text{(with linearization)} \tag{25}$$

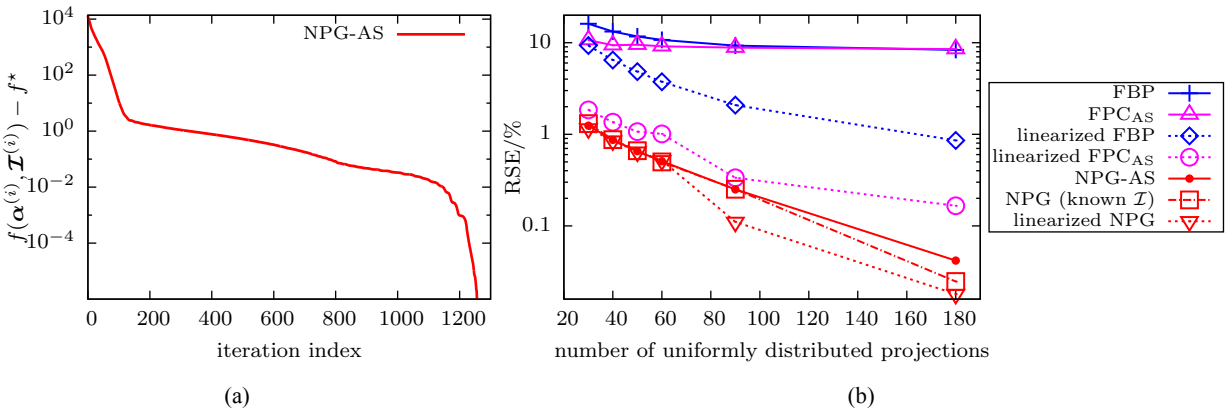
  respectively,
- the fixed-point continuation active set ($\mathrm{FPC_{AS}}$) methods [23] without and with linearization [see (25)], which solve the synthesis basis pursuit denoising (BPDN) problem [24]:

$$\min_{\boldsymbol{s}} \frac{1}{2}\|\boldsymbol{y} - \Phi\Psi\boldsymbol{s}\|_2^2 + u'\|\boldsymbol{s}\|_1 \tag{26}$$

  and obtain the signal estimates as $\hat{\boldsymbol{\alpha}} = \Psi\boldsymbol{s}^{(+\infty)}$, where $\boldsymbol{s}^{(+\infty)}$ is the vector of the transform signal coefficients obtained upon convergence;
- our [Matlab implementation available at `https://github.com/isucsp/npg`]
  - NPG-AS alternating descent method with step adaptation parameters n, ξ and regularization constant u chosen as

$$n = 4, \qquad \xi = 0.5, \qquad u = 10^a q^{J/2} \|\Psi^T \Phi^T \ln_\circ(\boldsymbol{\mathcal{E}}/\mathcal{I}^{\mathrm{in}}_{\mathrm{MAX}})\|_\infty \tag{27}$$

    where a is tuned in the set $\{-1,-2,-3,-4,-5,-6,-7,-8,-9\}$ for good reconstruction performance in terms of relative square error (RSE) [defined in (31)] for each number of projections,
  - NPG for known mass attenuation spectrum $\iota(\kappa)$ that iterates Step 1) only and uses the exact sampled $\iota(\kappa)$ as $\boldsymbol{\mathcal{I}}$ and the step adaptation parameters in (27).



(a) (b)

FIGURE 3. (a) Centered objective function $f - f^\star$ versus the iteration index i for NPG-AS from 180 equally spaced projections, and (b) RSEs of various methods versus the number of equally spaced projections.

- linearized NPG [15], which aims at solving

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\boldsymbol{y} - \Phi\boldsymbol{\alpha}\|_2^2 + u'' r(\boldsymbol{\alpha}) \tag{28}$$

where $\boldsymbol{y}$ are the linearized measurements in (25) and the penalty $r(\boldsymbol{\alpha})$ has been defined in (13b).

The regularization tuning constants u' and u'' for sparse regularization terms of FPC$_{\text{AS}}$ and linearized NPG have the form $10^a \|\Psi^T \Phi^T \boldsymbol{y}\|_\infty$, where a is tuned for good reconstruction performance in terms of RSE. Before applying the reconstruction algorithms, we normalize the energy measurements by their largest value and set $\mathcal{I}^{\text{in}}_{\text{MAX}} = 1$. We initialize iterative reconstruction methods with or without linearization using corresponding FBP reconstructions. NPG-AS employs the convergence criterion in (15a) with convergence threshold

$$\epsilon = 10^{-6} \tag{29}$$

FPC$_{\text{AS}}$ runs on its default options and all other iterative methods use the relative change of $\boldsymbol{\alpha}$ between two consecutive steps as the convergence criterion:

$$\delta^{(i)}_{\boldsymbol{\alpha}} < \epsilon \tag{30}$$

with the convergence threshold ϵ in (29). The maximum number of iterations is $M = 2000$ for all iterative methods.

For NPG-AS and NPG, we set $J = 17$ and $q^J = 10^3$ to ensure sufficiently wide coverage of the spline basis. Note that (12a) is convex with respect to $\boldsymbol{\alpha}$ under condition (12b).

Our reconstruction performance metric for an estimator $\widehat{\boldsymbol{\alpha}}$ is the RSE:

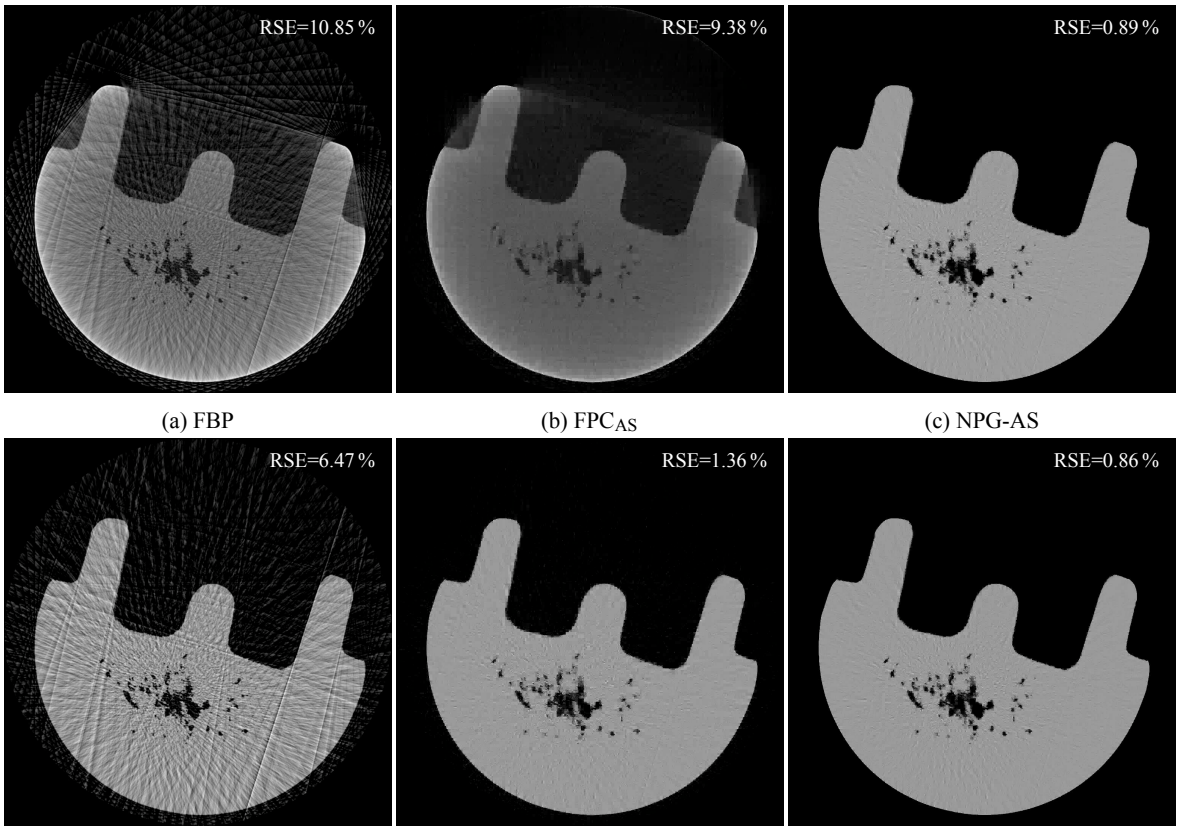
$$\text{RSE}\{\widehat{\boldsymbol{\alpha}}\} = 1 - \left(\frac{\widehat{\boldsymbol{\alpha}}^T \boldsymbol{\alpha}_{\text{true}}}{\|\widehat{\boldsymbol{\alpha}}\|_2 \|\boldsymbol{\alpha}_{\text{true}}\|_2} \right)^2 \tag{31}$$

which is invariant to scaling $\widehat{\boldsymbol{\alpha}}$ by a nonzero constant.

In Fig. 3a, we show the centered objective as a function of the iteration index i for NPG-AS from 180 parallel projections. Note that $f(\boldsymbol{\alpha},\mathcal{I}) = \mathcal{L}(\Phi\boldsymbol{\alpha},\mathcal{I}) + u r(\boldsymbol{\alpha})$ and $f^\star$ is its minimum value when optimized with respect to $\boldsymbol{\alpha}$ and $\mathcal{I}$.

Figure 3b shows RSEs as functions of the number of equally spaced parallel projections for various methods. RSEs of the methods that do not assume knowledge of the mass attenuation spectrum $\iota(\kappa)$ are shown using solid lines whereas dashed lines represent methods that assume known $\iota(\kappa)$ [i.e., known incident spectrum of the X-ray machine and mass attenuation (material)] are shown using dashed lines. Among the methods that do not assume knowledge of $\iota(\kappa)$, FBP and FPC$_{\text{AS}}$ ignore the polychromatic source effects whereas NPG-AS blindly corrects for the polychromatic effects.

The same line color is used to present the same general methods, e.g., FBP and linearized FBP, FPC$_{\text{AS}}$ and linearized FPC$_{\text{AS}}$, and NPG-AS, NPG, and linearized NPG. FBP and FPC$_{\text{AS}}$ ignore the polychromatic nature of the measurements and consequently perform poorly and do not improve as the number of projections increases. Linearized FBP, FPC$_{\text{AS}}$, and NPG, which assume perfect knowledge of the mass attenuation spectrum, perform much better than FBP and FPC$_{\text{AS}}$. Linearized FPC$_{\text{AS}}$, which imposes signal sparsity, is up to 5 times better in RSE than linearized FBP, demonstrating the importance of enforcing signal sparsity. NPG and linearized NPG, which impose *both* signal sparsity and nonnegativity,



(a) FBP (b) FPC_{AS} (c) NPG-AS

(d) linearized FBP (e) linearized FPC_{AS} (f) NPG, linearized NPG

FIGURE 4. Reconstructions from 40 equally spaced parallel projections (a)–(c) without and (d)–(f) with prior knowledge of the X-ray spectrum and the inspected material.

outperform linearized FPC_{AS}, particularly as the number of projections increases. The RSEs are cut two or more times by imposing the signal nonnegativity constraints.

The blind NPG-AS method is competitive with NPG and linearized NPG, indicating that good reconstruction is possible even without the information of mass attenuation spectrum $\iota(\kappa)$. Note that these three methods reach almost perfect reconstructions from 180 parallel-beam projections. In this example, linearized NPG is the best-performing method overall; however, we expect it to be sensitive to noise due to the zero-forcing nature of linearization.

Figure 4 shows reconstructions from 40 equally spaced parallel projections. The reconstructions in Figs. 4a–4c correspond to methods that to not have knowledge of the incident spectrum and material. FBP exhibits aliasing, due to the small number of projections, and beam hardening artifacts, including cupping and streaking artifacts. FPC_{AS} provides a smooth reconstruction and removes aliasing thanks to the fact that it imposes image sparsity in the DWT domain, but clearly cannot cope with the beam hardening effects because it is based on the linear model that ignores the nonlinearities due to the polychromatic source. The NPG-AS method accounts for the signal sparsity, nonnegativity, and, most importantly, uses the correct polychromatic model, which leads to over 10 times smaller RSE than those of FBP and FPC_{AS}.

Figures 4d-4f show the reconstructions with known incident spectrum and material [i.e., known $\iota(\kappa)$]. By comparing linearized FPC_{AS}, linearized FBP, and (linearized) NPG, we observe the improvements brought by signal sparsity and nonnegativity constraints, respectively.

CONCLUSION

Future work will include the adaptation of the number and positions of B1-spline knots used for modeling the mass attenuation spectrum and accounting for characteristic lines in the incident spectrum. We will also adapt the proposed approach to the Poisson measurements model, useful in low-dose scenarios, which are of great practical interest in biomedical applications [4].

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (NSF) Industry-University Cooperative Research Program, CNDE, Iowa State University and the NSF Award CCF-1421480.

REFERENCES

1. J. Barrett and N. Keat, "Artifacts in CT recognition and avoidance," Radiographics **24**, 1679–1691 (2004).
2. J. Hsieh, *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*, 2nd ed. (SPIE, Bellingham, WA, 2009).
3. 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. (Nov. 2011), pp. 2111–2116.
4. J. Nuyts et al., "Modelling the physics in the iterative reconstruction for transmission computed tomography," Phys. Med. Biol. **58**, R63–R96 (2013).
5. A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging* (IEEE Press, New York, 1988).
6. R. Gu and A. Dogandžić, "Beam hardening correction via mass attenuation discretization," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (May 2013), pp. 1085–1089.
7. R. Gu and A. Dogandžić, "Sparse signal reconstruction from polychromatic X-ray CT measurements via mass attenuation discretization," in Review of Progress in Quantitative Nondestructive Evaluation, edited by D. O. Thompson, D. E. Chimenti, and L. J. Bond, AIP Conf. Proc. Vol. 33 1581 (published by AIP, Melville, NY, 2014), pp. 1848–1855.
8. L. L. Schumaker, *Spline Functions: Basic Theory*, 3rd ed. (Cambridge Univ. Press, New York, 2007).
9. J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: A survey and extensions," Math. Methods Oper. Res. **66**, 373–407 (2007).
10. Y. Nesterov, "Gradient methods for minimizing composite functions," Math. Program. **140**, 125–161 (2013).
11. P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization* (Academic Press, San Diego, CA, 1981).
12. M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," Magnetic Resonance in Medicine **58**, 1182–1195 (2007).
13. Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," in Sov. Math. Dokl. Vol. 27, 2 (1983), pp. 372–376.
14. N. Parikh and S. Boyd, "Proximal algorithms," Found. Trends Optim. **1**, 123–231 (2013).
15. R. Gu and A. Dogandžić, "A fast proximal gradient algorithm for reconstructing nonnegative signals with sparse transform coefficients," in Proc. Asilomar Conf. Signals, Syst. Comput. (Nov. 2014), in press.
16. J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," IMA J. Numer. Anal. **8**, 141–148 (1988).
17. A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imag. Sci. **2**, 183–202 (2009).
18. B. O'Donoghue and E. Candès, "Adaptive restart for accelerated gradient schemes," Found. Comput. Math., 1–18 (2013).
19. S. Boyd et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," Found. Trends Machine Learning **3**, 1–122 (2011).
20. 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*, tech. rep. NISTIR 5632 (National Inst. Standards Technol., Ionizing Radiation Div., Gaithersburg, MD, 1995).
21. G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. (Duxbury Press, Pacific Grove, CA, 2002).
22. G. T. Herman, "Correction for beam hardening in computed tomography," Phys. Med. Biol. **24**, 81–106 (1979).
23. Z. Wen et al., "A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization, and continuation," SIAM J. Sci. Comput. **32**, 1832–1857 (2010).
24. E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," Commun. Pur. Appl. Math. **59**, 1207–1223 (2006).