Optimization-Based Approach for Tomographic Inversion from Multiple Data Modalities

Zichao (Wendy) Di, Sven Leyffer, and Stefan M. Wild

Optimization-Based Approach for Tomographic Inversion from Multiple Data Modalities

Zichao (Wendy) Di*, Sven Leyffer*, and Stefan M. Wild*

Abstract. Fluorescence tomographic reconstruction, based on the detection of photons coming from fluorescent emission, can be used for revealing the internal elemental composition of a sample. On the other hand, conventional transmission tomography can be used for reconstructing the spatial distribution of the absorption coefficient inside a sample. In this work, we integrate both fluorescence and transmission data modalities and formulate a nonlinear optimization-based approach for reconstruction. This model provides a simultaneous reconstruction of both the quantitative spatial distribution of all elements and the absorption effect in the sample. Mathematically speaking, we show that compared with the single-modality inversion, the joint inversion provides a better-posed problem, which implies a better recovery. Therefore, the challenges in fluorescence tomography arising mainly from the self-absorption effects in the sample are partially mitigated. The use of this technique is demonstrated on the reconstruction of several synthetic samples.

Keywords. Tomographic reconstruction, X-ray fluorescence, X-ray transmission, Joint inversion, Nonlinear optimization, Truncated-Newton method

1. Introduction. Tomographic imaging refers to the reconstruction of a 3D object from its 2D projections by sectioning the object, through the use of any kind of penetrating wave, from many different directions. It has had a revolutionary impact in a number of fields ranging from biology, physics, and chemistry to astronomy. The technique requires an accurate image reconstruction, however, and the resulting reconstruction problem is ill-posed and does not have a unique solution because of insufficient measurements. The main reasons for the lack of measurements include time-consuming restrictions such as the low-dose computed tomography (CT) requirement for performing longitudinal studies, and a limitation on the number of projection angles, the field of view, and the signal-to-noise ratio. Different modalities of tomograms all have been derived by using different physical phenomena; these include CT and magnetic resonance. The problems associated with these mathematical and experimental implementations are similar, however. In particular, one of the most popular modalities is X-ray transmission (XRT) tomography, which measures the ratio of the incident beam intensity to the intensity of the beam transmitted through the sample. A complementary tomographic technique, X-ray fluorescence (XRF) tomography, which is a novel synchrotron-based imaging modality, has also received considerable attention in recent years because of its ability to trace the elemental content of samples [11, 20, 23]. However, the reconstruction problem for fluorescence tomography is much more difficult than it is for transmission tomography, because of the absorption of the photons on the excitation and detection paths.

*Mathematics and Computer Science Division, Argonne National Laboratory. (`wendydi@mcs.anl.gov`, `leyffer@mcs.anl.gov`, `wild@anl.gov`.)

Conventional tomographic reconstruction approaches can be classified in two main categories [12]: filtered backprojection (FBP) [26] and iterative methods. FBP includes the first reconstruction algorithms used for XRF imaging [3]. It is well known for its rapid processing with contrast enhancement; however, some artifacts can be caused by blurring effects and the reconstruction filter. Iterative methods include the algebraic reconstruction technique (ART), the simultaneous iterative reconstruction technique, and the iterative least squares technique. Iterative techniques generally reduce sensitivity to noise and can reconstruct a high-quality image in the case of incomplete data, such as an insufficient set of available projections, or if the projections are not distributed uniformly in angle or are sparse or missing at certain orientations. In particular, ART [8, 17, 23, 34] and the expectation maximization algorithm [9, 14, 15, 32] have been widely applied for XRF reconstruction.

Another research effort regarding XRF reconstruction has focused on achieving a more accurate approach with a correction for attenuation, which involves simultaneously estimating emission and attenuation maps in conventional emission tomography when no transmission scan is performed. The correction of fluorescent attenuation was first introduced by Hogan et al. [11]. In [14,15], La Rivière et al. propose a penalized-likelihood reconstruction for XRF with an approximate expression for the fluorescence attenuation map as a linear combination of known quantities and an element's own unknown distribution. In [16], Miqueles and De Pierro develop a novel alternating method for retrieving simultaneously the fluorescence density and the attenuation coefficients for XRF reconstruction. Recently, Sawatzky et al. [24] have proposed a general numerical framework for multichannel image reconstruction in spectral X-ray CT that enables exploration of statistical correlations between the decomposed material sinograms and improves the reconstruction quality.

A common feature of these XRF reconstruction approaches is that the linear attenuation coefficients of the sample either are assumed to be known or approximated from a single imaging modality. In this work, we take advantage of recent advances in simultaneous acquisition of different modality signals, and we propose a novel joint inversion framework in the context of XRF and XRT that improves the ill-posed nature of tomography problems. Our new approach has the potential to provide the distribution of both the heavy elements that emit detectable fluorescent lines as well as lighter elements.

1.1. Principles of X-Ray Transmission. XRT is an imaging technique that uses X-rays to view the internal structure of an opaque object, and provides the spatial distribution of the absorption coefficient inside the object. In order to create the image, a beam of X-rays produced from an X-ray generator is projected toward the object. A certain amount of the X-rays is absorbed by the object, depending on the object's density and composition. The X-rays that pass through the object are captured behind the object by a detector. The detector can then provide a 2D representation of the object's internal structures.

1.2. Principles of X-Ray Fluorescence. XRF is the emission of characteristic X-rays from a nonradioactive material that has been excited by bombardment with a pencil beam of intense, monochromatic synchrotron X-rays of energy greater than the principal binding energy of the elements of interest. These elements emit certain characteristic fluorescence X-rays in the form of photons isotropically, which are detected by an energy-dispersive detector. This detector is placed parallel to the direction of the incident beam in order to minimize

contamination by Compton scattering photons from the incident beam. The detector outputs a spectrum comprising multiple peaks [22]. These peaks correspond to the principal emission lines of the various elements present in the sample, since the number of detected fluorescence photons is linearly proportional to the quantity of chemical elements from which it originated. By measuring the height or area of these peaks, the chemical element from which it originated can be deduced, and its amount can be quantified.

2. Mathematical Model. The radiation intensity traveling toward a sample is attenuated along the path of an X-ray beam. Consequently, the Radon transform [21], which in 2D consists of integration over straight lines, forms the mathematical foundation for tomographic imaging. In this section, we describe the mathematical models for XRT and XRF.

Notation. We use Greek letters to denote the indices related to the experimental setup; all other quantities are denoted by Roman letters. In detail, we use Θ and $\mathcal{T}$, respectively, to denote the complete collection of $|\Theta|$ angles and $|\mathcal{T}|$ beamlets and $\theta \in \Theta$ and $\tau \in \mathcal{T}$ to denote the index of the X-ray beam angle and discretized beamlet, respectively. The set $\mathcal{V}$ denotes the complete collection of $|\mathcal{V}|$ spatial voxel indices, which we use to discretize the sample. By $\mathbf{L} = [L_v^{\theta,\tau}]$, we denote the tensor of intersection lengths of beamlet (θ, τ) with the voxel $v \in \mathcal{V}$. We use $\mathcal{E}$ to denote the collection of $|\mathcal{E}|$ possible element indices and μ_e^E to denote the linear attenuation coefficient of element e at beam incident energy E.

Our goal is to recover the tensor $\mathbf{W} = [W_{v,e}]$ denoting the mass of element $e \in \mathcal{E}$ in voxel $v \in \mathcal{V}$.

2.1. Discrete XRT Imaging Model. The geometry of the XRT imaging setup is illustrated in Figure 1 in the 2D case.

A traditional way (see, e.g., [12]) to model the transmission projection of a sample from beamlet (θ, τ) is

$$\tilde{F}_{\theta,\tau}^T(\boldsymbol{\mu}) = I_0 \exp\left\{ -\sum_v L_v^{\theta,\tau} \mu_v^E \right\},$$

where I_0 is the X-ray beam intensity and $\boldsymbol{\mu} = [\mu_v^E]$ is the attenuation/absorption coefficients at incident energy E.

We note that the coefficients $\boldsymbol{\mu}$ depend on $\mathbf{W}$ by way of $\mu_v^E = \sum_{e'} W_{v,e'} \mu_{e'}^E$ for all $v \in \mathcal{V}$. Since the goal is to recover $\mathbf{W}$, the forward model that we use is

$$(2.1) \qquad F_{\theta,\tau}^T(\mathbf{W}) = I_0 \exp\left\{ -\sum_{v,e} L_v^{\theta,\tau} \mu_e^E W_{v,e} \right\}.$$

For $e \in \mathcal{E}$ and $v \in \mathcal{V}$, the first-order derivative of (2.1) with respect to $W_{v,e}$ is

$$\frac{\partial}{\partial W_{v,e}} F_{\theta,\tau}^T(\mathbf{W}) = -I_0 \exp\left\{ -\sum_{v',e'} L_{v'}^{\theta,\tau} \mu_{e'}^E W_{v',e'} \right\} L_v^{\theta,\tau} \mu_e^E.$$
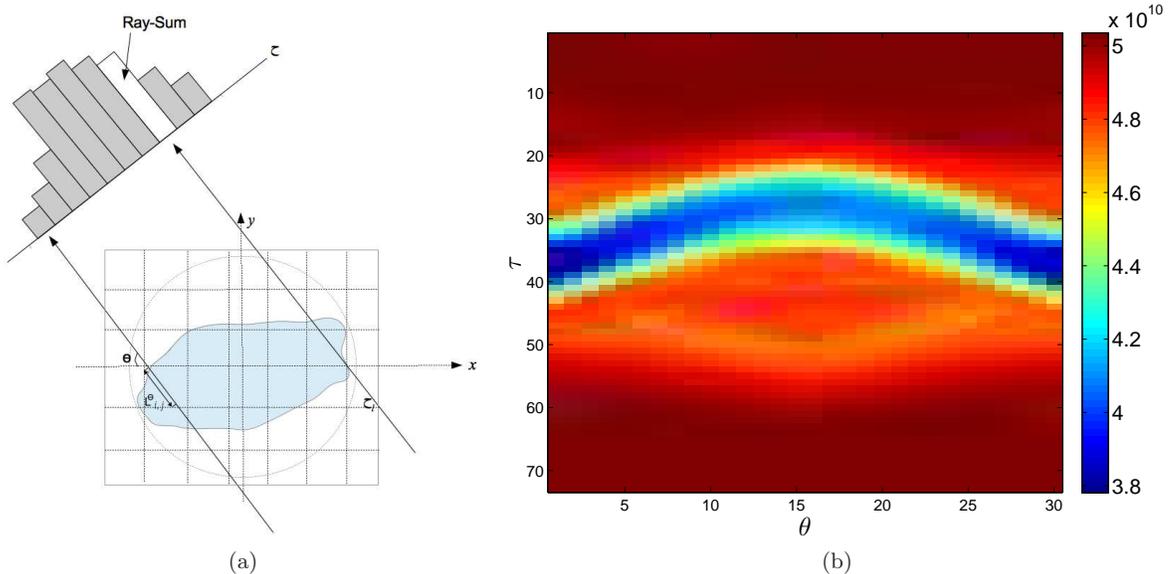
Figure 1: (a) Illustration of the discrete XRT projection geometry: the sample resides in space (discretized in 2D for simplicity of illustration), whereas the beam is parameterized by its angular and translation scan (θ and τ, respectively); (b) Corresponding structure of XRT data for all $|\Theta||\mathcal{T}|$ projections.

2.2. Discrete XRF Imaging Model. The discrete model we use to simulate XRF is built on the one proposed in [15], and the corresponding geometry is illustrated in Figure 2. Two main features distinguish our model from the one in [13, 15]. First, instead of modeling the XRF process in terms of both the elemental distribution and the attenuation map of fluorescence energy, we model the XRF only in terms of the elemental distribution. Second, rather than modeling the detected fluorescence photons directly, we follow an elemental approach, whereby we model the fluorescence emitted from an elemental atom by its corresponding elemental unit spectrum. Then, the total fluorescence spectrum detected from the given sample is modeled as a weighted sum of such unit spectra of the elements being recovered.

We approximate the elemental unit spectrum as follows. Based on Sherman's equation [27], a first-order approximation (i.e., no chain effect) of the net-line intensity corresponding to the characteristic fluorescence energy E_e emitted from element e is

$$I_{e,l,s} = I_0 c_e \omega_{e,l} \left(1 - \frac{1}{r_{e,s}} \right) \mu_e^E,$$

where c_e is the total concentration of element e in the sample, $\omega_{e,l}$ is the fluorescence yield of e for the spectral line l, and $r_{e,s}$ is the probability that a shell s electron (rather than other shell electrons) will be ejected.

Sources for values of $\omega_{e,l}$, μ_e^E, and $r_{e,s}$ include the tabulated data [28] and the online

database xraylib [25]. In our calculations, we used the XRF cross sections from xraylib to obtain the quantity $\omega_{e,l}\left(1 - \frac{1}{r_{e,s}}\right)\mu_e^E$.

For an energy-dispersive detector used for detecting fluorescence, let $\mathcal{I}$ denote the set of the complete collection of its $|\mathcal{I}|$ energy channel indices and let $\mathbf{x}$ be the $|\mathcal{I}|$-dimensional vector denoting its energy channels. Let $\mathbf{1}_{E_e}$ be the $|\mathcal{I}|$-dimensional unit indicator vector with its component defined as

$$
[\mathbf{1}_{E_e}^{\mathbf{x}}]_i := \begin{cases} 1 & \text{if } |x_i - E_e| = \min_j(|x_j - E_e|) \quad \text{and} \quad x_i \neq 2E_e - x_{i-1} \\ 0 & \text{otherwise} \end{cases} , \ i = 1, \ldots, |\mathcal{I}|,
$$

where we choose to break ties between neighboring channels lexicographically. We then obtain the perfect spectral line $\mathbf{I}^{\mathbf{x}} = I_{e,l,s}\mathbf{1}_{E_e}^{\mathbf{x}}$. However, the fluctuations in the number of excitations and ionizations in the detector material mean that one usually observes a Gaussian-like peak rather than an ideal, delta-function peak for a monoenergetic particle beam. This forms the basis of the detector response principle [28]. The width of this Gaussian-like peak determines the capability to distinguish particles with different energies. The energy resolution ΔE is given by the full-width-at-half-maximum (FWHM) of the single peak. For a Gaussian distribution with standard deviation σ, we have FWHM $\approx 2.35\sigma$. Therefore, the final unit spectrum obtained by convolving the perfect spectral line $\mathbf{I}^{\mathbf{x}}$ with a Gaussian distribution is

$$
\mathbf{M}_{e,l,s} = \mathcal{F}^{-1}\left(\mathcal{F}(\mathbf{I}^{\mathbf{x}}) * \mathcal{F}\left(\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{\frac{-\mathbf{x}^2}{2\sigma^2}\right\}\right)\right),
$$

where $*$ denotes convolution and $\mathcal{F}$ $(\mathcal{F}^{-1})$ is the (inverse) Fourier transform. In our simulation, we consider only $K_\alpha, K_\beta, L_\alpha, L_\beta$ lines (see [28]), which result in the final unit spectrum of element e given by $\mathbf{M}_e = \sum_{l,s} \mathbf{M}_{e,l,s}$.

For given elemental unit spectra, we model the total fluorescence spectrum of a sample with multiple elements by considering the attenuation of the beam energy and self-absorption of the fluorescence energy. We let $A_v^{E,\theta,\tau}$ represent the attenuation experienced by beamlet (θ,τ) (at incident beam energy E) as it travels toward voxel $v \in \mathcal{V}$:

$$
A_v^{E,\theta,\tau}(\mathbf{W}) \quad = \exp\left\{-\sum_{v'}\mu_{v'}^E L_{v'}^{\theta,\tau}\mathbb{I}_{v'\in\mathcal{U}_v^{\theta,\tau}}\right\} = \exp\left\{-\sum_{v'}\sum_{e'}W_{v',e'}\mu_{e'}^E L_{v'}^{\theta,\tau}\mathbb{I}_{v'\in\mathcal{U}_v^{\theta,\tau}}\right\},
$$

where $\mathbb{I}_X$ is the indicator (Dirac delta) function for the event X and $\mathcal{U}_v^{\theta,\tau} \subset \mathcal{V}$ is the set of voxels that are intersected by beamlet (θ,τ) before it enters voxel v.

By $F_{v,e}^{\theta,\tau}(\mathbf{W})$, we denote the attenuation of fluorescence energy emitted for beamlet (θ,τ) from element $e \in \mathcal{E}$ at voxel $v \in \mathcal{V}$ before it reaches the detector. Note that the energy-dispersive detectors typically used to detect fluorescence spectra are nonimaging detectors that record only the count of photons within specific energy levels but do not record the initial spatial information of the detected photons. In order to provide the spatial restriction necessary for tomographic reconstruction, a pencil beam illuminating only a single line is used to scan through the object while the object is rotating. Therefore, we need to track the history of the emitted photons only along the corresponding beam line rather than the whole
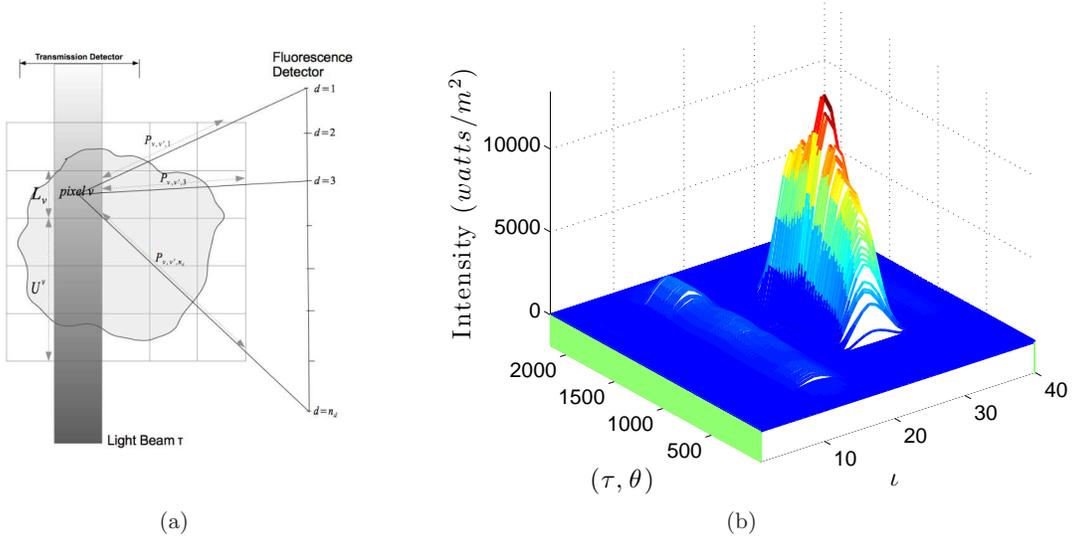
Figure 2: (a) Geometry and notation of XRF setup; (b) Corresponding structure of XRF data, with each projection (θ, τ) yielding a spectrum indexed by $\iota \in \{1, \ldots, |\mathcal{I}|\}$.

sample space. Also, instead of considering all possible traveling paths of the emitted photons isotropically, we restrict their traveling paths to the solid angle area between the emitting source and the fluorescence detector. By numerically subdividing the fluorescence detector into n_d detectorlets, we approximate $F_{v,e}^{\theta,\tau}(\mathbf{W})$ as

$$F_{v,e}^{\theta,\tau}(\mathbf{W}) = \frac{1}{n_d} \sum_{d=1}^{n_d} \exp\left\{ -\sum_{v'} \sum_{e'} W_{v',e'} \mu_{e'}^{E_e} P_{v,v',d}^{\theta,\tau} \right\},$$

where $\mathbf{P} = [P_{v,v',d}^{\theta,\tau}]$ denotes a tensor of intersection lengths of fluorescence detectorlet path d (originated from voxel v and beamlet (θ, τ)) with the voxel v' and $\mu_{e'}^{E_e}$ is the linear attenuation coefficient of element e' at the fluorescence energy E_e of element e. Accordingly, the fluorescence spectrum $\boldsymbol{F}_{\theta,\tau}^R$ of the object resulting from beamlet (θ, τ) is the $|\mathcal{I}|$-dimensional vector

$$
\begin{aligned}
\boldsymbol{F}_{\theta,\tau}^R(\mathbf{W}) \ &= \sum_e \left(\sum_v L_v^{\theta,\tau} A_v^{E,\theta,\tau}(\mathbf{W}) F_{v,e}^{\theta,\tau}(\mathbf{W}) W_{v,e} \right) \boldsymbol{M}_e \\
&= \sum_{v,e,d} \frac{L_v^{\theta,\tau} W_{v,e} \boldsymbol{M}_e}{n_d} \exp\left\{ -\sum_{v',e'} W_{v',e'} \left(\mu_{e'}^E L_{v'}^{\theta,\tau} \mathbb{I}_{v' \in \mathcal{U}_v^{\theta,\tau}} + \mu_{e'}^{E_e} P_{v,v',d}^{\theta,\tau} \right) \right\}.
\end{aligned}
$$

(2.2)

6

One can show that the first-order derivative with respect to $W_{v,e}$ is given by

(2.3)

$$
\begin{aligned}
&\frac{\partial}{\partial W_{v,e}} \boldsymbol{F}^R_{\theta,\tau}(\mathbf{W}) \\
&= \sum_{e'} \left(\sum_{v'} L^{\theta,\tau}_{v'} W_{v',e'} \left(\frac{\partial A^{E,\theta,\tau}_{v'}}{\partial W_{v,e}} F^{\theta,\tau}_{v',e'} + A^{E,\theta,\tau}_{v'} \frac{\partial F^{\theta,\tau}_{v',e'}}{\partial W_{v,e}} \right) \right) \boldsymbol{M}_{e'} + L^{\theta,\tau}_v A^{E,\theta,\tau}_v F^{\theta,\tau}_{v,e} \boldsymbol{M}_e \\
&= -\sum_{e'} \left(\sum_{v'} L^{\theta,\tau}_{v'} W_{v',e'} A^{E,\theta,\tau}_{v'} F^{\theta,\tau}_{v',e'} \left(\mu^E_e L^{\theta,\tau}_v \mathbb{I}_{v\in\mathcal{U}^{\theta,\tau}_{v'}} \right) \right) \boldsymbol{M}_{e'} \\
&\quad -\sum_{e'} \left(\sum_{v'} L^{\theta,\tau}_{v'} W_{v',e'} A^{E,\theta,\tau}_{v'} \left(\frac{1}{n_d} \sum_{d=1}^{n_d} \exp\left\{ -\sum_{v''}\sum_{e''} W_{v'',e''} \mu^{E_{e'}}_{e''} P^{\theta,\tau}_{v',v'',d} \right\} \mu^{E_{e'}}_e P^{\theta,\tau}_{v',v,d} \right) \right) \boldsymbol{M}_{e'} \\
&\quad + L^{\theta,\tau}_v A^{E,\theta,\tau}_v F^{\theta,\tau}_{v,e} \boldsymbol{M}_e .
\end{aligned}
$$

3. Optimization-Based Reconstruction Formulations and Algorithms.

We now describe the optimization-based approach for solving inverse problems involving the models $F^T_{\theta,\tau}(\mathbf{W})$ and $\boldsymbol{F}^R_{\theta,\tau}(\mathbf{W})$. We assume that for these models the respective data are given by

- $\boldsymbol{D}^R_{\theta,\tau} \in \mathbb{R}^{n_E}$, the measurement data of XRF detected at angle θ from light beam τ, and
- $D^T_{\theta,\tau} \in \mathbb{R}$, the measurement data of XRT detected at angle θ from light beam τ

for a set of (θ,τ) values.

Each of the approaches follows a constrained least-squares formulation [7, 33],

$$
\min_{\mathbf{x}\in\mathcal{X}} \phi(\mathbf{x}) = \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|^2 := \frac{1}{2} \mathbf{r}(\mathbf{x})^\top \mathbf{r}(\mathbf{x}),
$$

where $\mathbf{x}$ denotes the $n \le m$ parameters, $\mathbf{r} : \mathbb{R}^n \mapsto \mathbb{R}^m$ denotes the residual mapping, and $\mathcal{X} \in \mathbb{R}^n$ denotes a feasible region for the parameters.

3.1. Joint Reconstruction Technique.

In this section, we describe our formulation of multimodal imaging as a joint inversion problem. The inverse problem we solve for the mass tensor $\mathbf{W}$ is

(3.1)
$$
\begin{aligned}
\min_{\mathbf{W}\ge\mathbf{0}} \quad \phi(\mathbf{W}) \quad &= \phi^R(\mathbf{W}) + \beta \phi^T(\mathbf{W}) \\
&= \frac{1}{2} \left\| \mathbf{F}^R(\mathbf{W}) - \mathbf{D}^R \right\|^2 + \frac{\beta}{2} \left\| \mathbf{F}^T(\mathbf{W}) - \mathbf{D}^T \right\|^2 \\
&= \frac{1}{2} \sum_{\theta,\tau} \left(\left\| \boldsymbol{F}^R_{\theta,\tau}(\mathbf{W}) - \boldsymbol{D}^R_{\theta,\tau} \right\|^2 + \beta \left\| F^T_{\theta,\tau}(\mathbf{W}) - D^T_{\theta,\tau} \right\|^2 \right),
\end{aligned}
$$

where the constraint $\mathbf{W} \ge \mathbf{0}$ is due to the physical nature of mass and $\beta > 0$ is a scaling between the two modalities. Therefore, the first-order derivative of the objective function in (3.1) with respect to $\mathbf{W}$ is

$$
\nabla\phi(\mathbf{W}) \quad = \boldsymbol{J}^R(\mathbf{W})^\top \left(\mathbf{F}^R(\mathbf{W}) - \mathbf{D}^R \right) + \beta \boldsymbol{J}^T(\mathbf{W})^\top \left(\mathbf{F}^T(\mathbf{W}) - \mathbf{D}^T \right).
$$

For $|\Theta|$ angular and $|\mathcal{T}|$ translation scans, $\boldsymbol{J}^T$ is the $|\Theta||\mathcal{T}| \times |\mathcal{V}||\mathcal{E}|$ Jacobian matrix of ϕ^T whose components are given by $\left[\boldsymbol{J}^T(\mathbf{W}) \right]_{i,j} := \frac{\partial F^T_i(\mathbf{W})}{\partial W_j}$, where i and j index the vectorizations

of the indices (θ, τ) and (v, e), respectively. Similarly, we let $\iota \in \mathcal{I}$ denote the energy channel index and $\boldsymbol{J}^R$ denote the $|\Theta||\mathcal{T}||\mathcal{I}| \times |\mathcal{V}||\mathcal{E}|$ Jacobian matrix of ϕ^R whose components are given by $\left[\boldsymbol{J}^R(\mathbf{W})\right]_{i,j} := \frac{\partial \boldsymbol{F}_i^R(\mathbf{W})}{\partial W_j}$, where i and j index the vectorizations of the indices (θ, τ, ι) and (v, e), respectively. Consequently, the master Jacobian matrix and master residual vector of (3.1) are $\mathbf{J}(\mathbf{W}) = \begin{bmatrix} \boldsymbol{J}^R(\mathbf{W}) \\ \beta \boldsymbol{J}^T(\mathbf{W}) \end{bmatrix}$ and $\mathbf{r}(\mathbf{W}) = \begin{bmatrix} \boldsymbol{F}^R(\mathbf{W}) - \boldsymbol{D}^R \\ \beta(\boldsymbol{F}^T(\mathbf{W}) - \boldsymbol{D}^T) \end{bmatrix}$, respectively.

3.2. Optimality Conditions.
We now consider the optimality conditions corresponding to problem (3.1). We define the set of binding constraints by $B(\mathbf{W}) = \{i : W_i = 0, \frac{\partial}{\partial W_i}\phi(\mathbf{W}) \geq 0\}$ and the strongly binding set by $B_s(\mathbf{W}) = B(\mathbf{W}) \cap \{i : \frac{\partial}{\partial W_i}\phi(\mathbf{W}) > 0\}$.

Therefore, the first-order necessary condition for $\mathbf{W}^* \geq \mathbf{0}$ to be a local minimizer of (3.1) is

$$\frac{\partial \phi(\mathbf{W}^*)}{\partial W_i} = \left[\mathbf{J}(\mathbf{W}^*)^\top \mathbf{r}(\mathbf{W}^*)\right]_i = 0 \qquad \forall i \notin B(\mathbf{W}^*).$$

The second-order sufficient condition additionally requires that

$$\mathbf{z}^\top \nabla^2 \phi(\mathbf{W}^*) \mathbf{z} > 0$$

for all vectors $\mathbf{z} \neq 0$ with $z_i = 0, i \in B_s(\mathbf{W}^*)$.

3.3. Motivation.
In practical experimental settings, the projection data are known to be incomplete and are typically corrupted by noise. Hence, small perturbations of the projection data will result in the solution being unstable. For this reason, the generalized tomography problem is ill-conditioned [4]. The most commonly used remedy for this ill-posed nature is to add a regularization term (e.g., Tikhonov regularization [1, 30, 31]). An appropriately chosen regularization can facilitate direct numerical solution; however, choosing a good regularizer is difficult [10, 19]. In this section, we examine the Jacobian of the nonlinear least-square formulations involving $\mathbf{F}^R$ and $\mathbf{F}^T$, and we provide an example motivating how joint inversion can be used to improve the solvability of such inverse problems.

Where an exact solution exists, the residual vector $\mathbf{r}(\mathbf{W}^*)$ vanishes, and the Hessian is therefore given by $\nabla^2 \phi(\mathbf{W}^*) = \mathbf{J}(\mathbf{W}^*)^\top \mathbf{J}(\mathbf{W}^*)$. For the second-order sufficient condition to hold strictly, a reduction of this matrix must be positive definite. In particular, this condition requires $\mathbf{J}(\mathbf{W}^*)_\mathcal{N}$ to be full column rank, where $\mathcal{N} = \{i \notin B_s(\mathbf{W}^*)\}$ is used to denote the columns of $\mathbf{J}(\mathbf{W}^*)$ not belonging to the strongly binding set. Analogously, we can say that the bound-constrained ($\mathbf{W} \geq \mathbf{0}$) system of nonlinear equations is well-posed; that is, the solution manifold consists of an isolated local minimum. Otherwise, if $\mathbf{J}(\mathbf{W}^*)_\mathcal{N}$ is rank-deficient, the system is ill-posed and can have infinitely many solutions. Even when no exact solution exists, a rank-deficient reduced Jacobian can indicate the existence of many local minima that are not isolated [6]. This case arises often in traditional tomographic reconstructions because the data is often insufficient and thus linear dependent.

The following example identifies a simple joint inversion case with an improved Jacobian (in terms of its rank) relative to the Jacobian from either of the single-modality formulations.

Example 3.1. *Consider a 2D (3×3 discretized) sample composed of the single element Ca. It is imaged by using $|\mathcal{I}| = 2$ energy channels, one angular ($\theta = 0^o$), and three translation*

scans, for which $W_{v,e}^* = 1$, $v \in [1, \ldots, 9]$, $e = 1$. The Jacobian

$$\boldsymbol{J} = \left[\frac{\boldsymbol{J}^R}{\beta \boldsymbol{J}^T} \right]$$

at $\mathbf{W}^*$ is given by

$$
\left[
\begin{array}{ccccccccc}
173 & -16 & -17 & 171 & -23 & -23 & 179 & -16 & -17 \\
2730 & -252 & -268 & 2706 & -361 & -361 & 2825 & -260 & -270 \\
173 & 177 & -31 & 171 & 170 & -53 & 179 & 173 & -36 \\
2730 & 2806 & -497 & 2706 & 2696 & -834 & 2825 & 2728 & -575 \\
173 & 177 & 179 & 171 & 170 & 156 & 179 & 173 & 175 \\
2730 & 2806 & 2830 & 2706 & 2696 & 2466 & 2825 & 2728 & 2770 \\
\hline
0.12\beta & 0 & 0 & 0.12\beta & 0 & 0 & 0.12\beta & 0 & 0 \\
0 & 0.12\beta & 0 & 0 & 0.12\beta & 0 & 0 & 0.12\beta & 0 \\
0 & 0 & 0.12\beta & 0 & 0 & 0.12\beta & 0 & 0 & 0.12\beta
\end{array}
\right].
$$

We note that
- $B_s(\mathbf{W}^*) = \emptyset$;
- $\texttt{rank}(\boldsymbol{J}^R) = 6$;
- $\texttt{rank}(\boldsymbol{J}^T) = 3$; and
- $\texttt{rank}(\boldsymbol{J}) = 9$, which returns a full column rank, well-determined system of equations.

Hence, $\mathbf{W}^*$ is an isolated minimizer of the joint inversion problem for all $\beta > 0$. The single modality inversions XRT and XRF consist of many connected local minima.

Similarly, if we increase the number of angular scans to two (i.e., $|\Theta| = 2$), then XRF returns in an overdetermined system with the corresponding $\boldsymbol{J}^R$ of size 12×9 and $\boldsymbol{J}^T$ of size 6×9. However, since $\texttt{rank}(\boldsymbol{J}^R) = 6$ and $\texttt{rank}(\boldsymbol{J}^T) = 6$, the XRF system is rank-deficient. This example shows that it is not always sufficient to capture more measurements, and the overall number of measurements can be reduced by joint reconstruction technique (JRT).

Regarding the overdetermined case, Figure 3 illustrates the local behavior of three different objectives in the neighborhood of the local minimum $\mathbf{W}^*$ given $\beta = 10^8$, where Figure 3a corresponds to perturbations in the null space of the XRT system and Figure 3b corresponds to perturbations in the null space of the XRF system. Taking both scenarios into account, W^* is the isolated solution of JRT, while there are other, connected solutions for XRF and XRT.

We mention that in practice JRT may not result in a full-rank Jacobian, but that the rank is often improved. Therefore, the area of the solution manifold of JRT is often smaller than the one from single modality inversion; as we will demonstrate in Section 5, this translates into better convergence performance of JRT.

The benefit of joint inversion can also be seen from a physical point of view. The XRF and XRT modalities considered here are based on different physical phenomena, which generate complementary morphological, structural, and chemical information. Consequently, their combination can naturally provide richer information about an imaged sample than possible with only one of these modalities. Furthermore, hardware technology has dramatically improved, so that simultaneous detection of correlative multimodality data is now available [5,29].

4. Optimization Complexity and Computational Expense. Throughout the remainder of this paper, for ease of visualization, we consider only 2D samples; and we fix the number
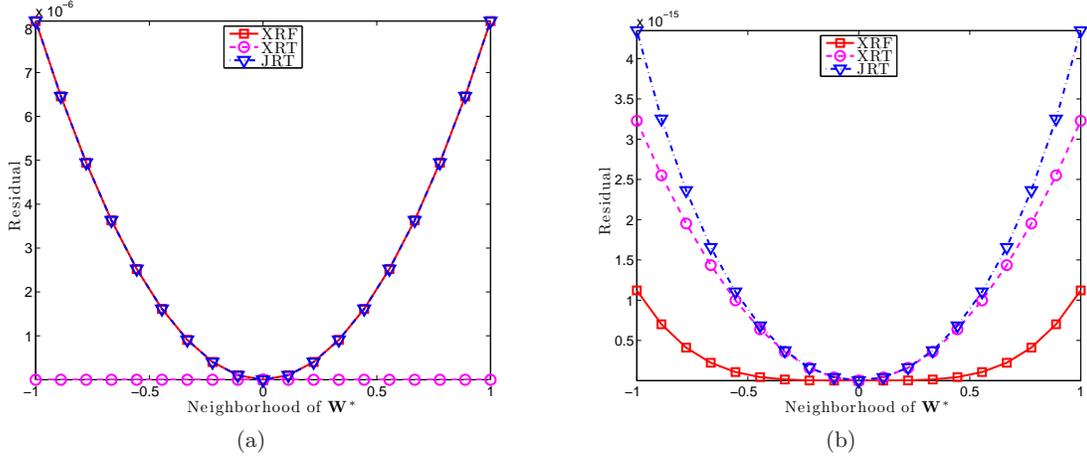
Figure 3: Residual objectives in a neighborhood of $\mathbf{W}^*$ for the overdetermined system in Example 3.1: Local behavior of different objectives in (a) the null space of the XRT system and (b) the null space of the XRF system.

of energy channels to $|\mathcal{I}| = 2000$, the number of detectorlets to $n_d = 5$, and the number of beamlets to $|\mathcal{T}| = \sqrt{2|\mathcal{V}|}$, where $|\mathcal{V}|$ denotes the number of voxels to be recovered.

If computed elementwise, the XRF forward model (2.2) requires about $|\mathcal{V}|^2|\mathcal{E}|^2|\mathcal{I}|n_d$ flops; accordingly, its elementwise derivative (2.3) requires about $|\mathcal{V}|^3|\mathcal{E}|^3|\mathcal{I}|n_d$ flops; hence, the overall cost of elementwise function and derivative evaluations for XRF is $\mathcal{O}(|\mathcal{V}|^3)$ flops. Similarly, the overall cost of function and gradient evaluations for XRT is $\mathcal{O}(|\mathcal{V}|^2)$ flops. Therefore, the XRF term dominates the computational cost of the joint inversion objective function (3.1).

In our implementation, however, we utilize the precalculated terms $L_v^{\theta,\tau}\mu_e^E$ and $\mu'^{E_e}_e P_{v,v',d}^{\theta,\tau}$ for every possible combination in order to avoid duplicated calculation. We also code in a tensor-product fashion for the summations in (2.2) and (2.3). Based on these strategies, Figure 4a shows the computational time of one $(\phi^R, \nabla\phi^R)$ evaluation for increased numbers of angular projections. In particular, with a model fit to the time shown, we observe that the actual time complexity is on the order of

$$(4.1) \qquad\qquad\qquad\qquad |\mathcal{V}|^{1.5}|\Theta|.$$

This shows that our approach is more feasible for larger-scale problems than the elementwise complexity analysis suggests.

We also compare the efficiency of different ways to calculate the gradient $\nabla\phi^R$. We use the Matlab-based `AdiMat` tool [2] to perform automatic differentiation-based (AD) derivative calculations. If $\mathbf{W}$ is a 2D matrix (i.e., $|\mathcal{E}| = 1$), then AD is performed in reverse mode; otherwise, if $\mathbf{W}$ is a 3D tensor with $|\mathcal{E}| > 1$, then AD is performed in forward mode. Figure 4b shows a comparison between the AD codes and our hand-coded derivatives. We observe that using the reverse mode, the cost of AD is approximately 6 times greater than that of our implementation of the tensor-based analytical derivative; furthermore, with the forward mode, the
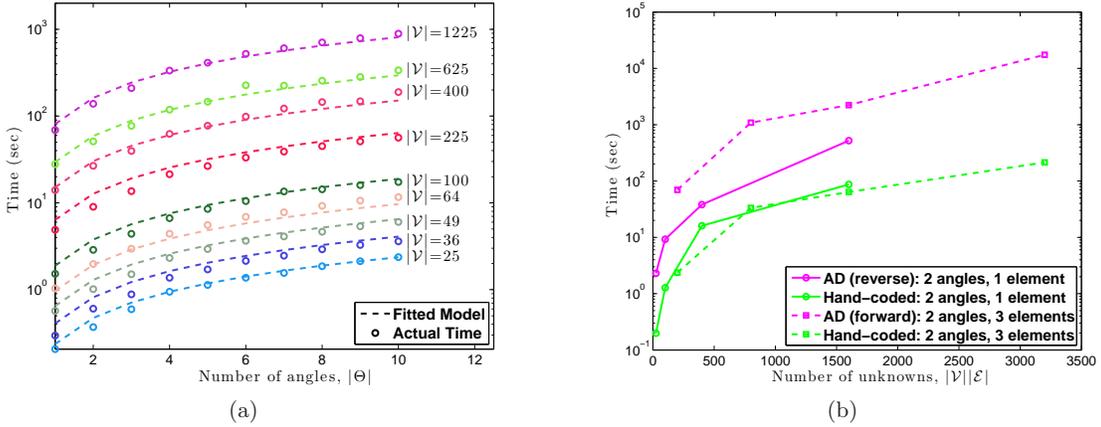
Figure 4: Time required for a single function and gradient evaluation $(\phi^R, \nabla\phi^R)$ for different input sizes (number of angular projections and voxels): (a) Actual computational time and the model in Equation (4.1) for fixed numbers of voxels; (b) Comparison of computational costs of a function and gradient evaluation $(\phi^R, \nabla\phi^R)$ using our hand-coded, tensor-based derivative and the AD-based derivatives.

AD derivative is roughly 30 times more expensive than our tensor-based analytical derivative. Therefore, we use our tensor-based analytical derivative in all the numerical experiments.

Because we are targeting large problems (with at least $|\mathcal{V}|$ nonzeros), having a fast and memory-efficient algorithm to solve (3.1) is highly desirable. Therefore, in all the numerical experiments, we use an inexact truncated-Newton (TN) method [18] with preconditioned projected conjugate gradient (PPCG) providing the search direction.

We also estimate the complexity of TN. Each outer iteration of the inexact TN performs the following computations (to "vectors" of dimension $|\mathcal{V}||\mathcal{E}|$):

- 1 infinity-norm calculation, 1 vector addition, and 2 (function, gradient) evaluations.
- a number of PPCG iterations, with cost per inner iteration given by
  - 1 (function, gradient) evaluation,
  - 4 inner products, and
  - 5 "vector+constant·vector" operations.

For the synthetic problems examined in the next section, 5 PPCG iterations are required on average per outer TN iteration. Based on the above estimate, one outer TN iteration requires 1 infinity norm, 20 inner products, 25 "vector+constant·vector", 1 vector add calculation, and 7 (function, gradient) evaluations. The overall cost of one TN iteration amounts to $|\mathcal{V}||\mathcal{E}|(1 + 20 + 25 + 1) = 47|\mathcal{V}||\mathcal{E}|$ flops, plus 7 (function, gradient) evaluations, whose time empirically grows like $|\mathcal{V}|^{1.5}$ (see (4.1)).

5. Numerical Results. In this section, we examine the performance of the joint reconstruction algorithm on two synthetic samples. All numerical experiments are performed on a platform with 32 GB of RAM and two Intel E5430 Xeon CPUs.
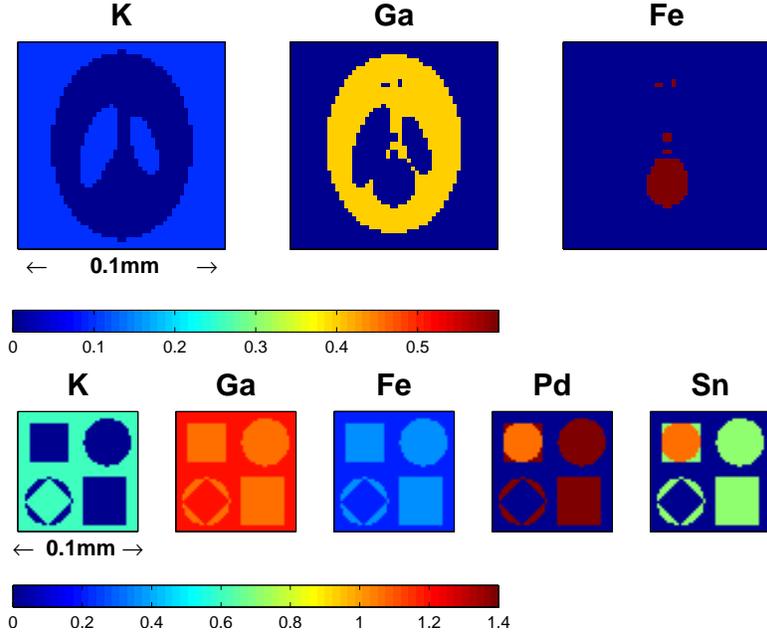
Figure 5: Chemical composition of the Shepp Logan phantom (top) and phase (bottom) synthetic samples; color indicates the mass of the existing elements.

The primary goal of our tests with the first sample is to measure the performance of JRT with respect to different problem sizes and contamination of the measurement data.

For the second sample, we test the independence of JRT performance with respect to the given sample and the benefit of our XRF forward model in correcting the attenuation coefficient by including the self-absorption effect.

5.1. Synthetic Samples. We present results for two synthetic samples: the "phantom" sample and the "phase" sample. Both samples are shown in Figure 5, together with their chemical composition. For the phantom sample, the chemical elements are potassium (K), gallium (Ga), and iron (Fe), each of which results in a significant amount of fluorescent radiation. The phase sample is typically used (see, e.g., [8, 15]) to explore the self-absorption effect in fluorescence imaging and additionally contains palladium (Pd) and tin (Sn).

We denote these samples by $\mathbf{W}^*$ and run each through our forward simulation to obtain fluorescence and transmission output. For all forward simulations in these tests, the simulated incident beam is monochromatic, with 20 keV energy and $0.2 \times 0.2\,\mu$m transverse size. A 2.4 mm-diameter transmission detector is placed at 90° to the incident beam at a distance 1.6cm from the sample. A fluorescence detector is placed parallel to the incident beam at a distance 1.6 cm from the sample. Given a numerical resolution of the sample image, we vary the translation and angular scan steps to get different-sized measurement data, which allows us to study the dependence of the reconstruction algorithms on the measurement data size.
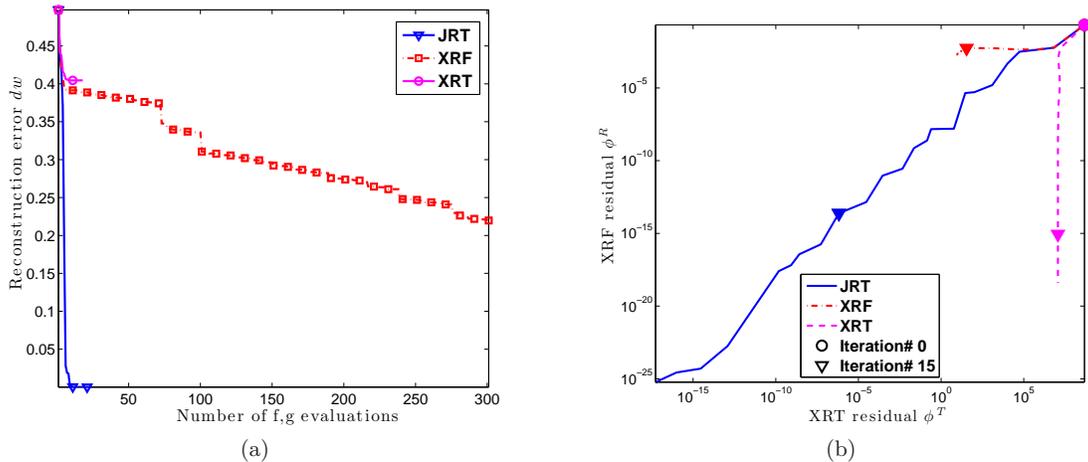
Figure 6: Optimization starting from $\mathbf{W}^0_{\text{bad}}$ for the 3×3 phantom sample: (a) Reduction of reconstruction error with respect to the number of (function, gradient) evaluations; (b) Trajectories of the XRF and XRT residuals for the three reconstruction approaches.

5.2. Performance Metrics. The known chemical composition of the synthetic samples facilitates our quantitative evaluation of the reconstruction algorithms. We compare JRT with the existing single XRT and XRF reconstruction approaches based on two metrics.

First, the conventional way to compare the performance of different reconstruction methods in the tomography community is to compare the recovered samples directly with the synthetic sample. In other words, we measure the "reconstruction error"

$$dw(\mathbf{W}) = \|\mathbf{W} - \mathbf{W}^*\|_F, \tag{5.1}$$

where $\mathbf{W}^*$ is the synthetic sample.

The second metric that we use is based on the forward residuals. We examine both the fluorescence residual ϕ^R and the transmission residual ϕ^T. Since the three approaches we consider are based on minimizing the residuals ϕ^R, ϕ^T, and $\phi^R + \beta\phi^T$, one might expect that each approach would perform best for its respective residual. By construction in the noise-free tests, we have that $\phi^R(\mathbf{W}^*) = \phi^T(\mathbf{W}^*) = 0$, and hence $\mathbf{W}^*$ is a global minimizer of all three residuals.

5.3. Phantom Sample Results. First, we compare the performance of single modality reconstruction (i.e., XRF and XRT), with the joint version JRT for problems of different resolution. To show the insensitivity of JRT to the initial guess, we consider both $\mathbf{W}^0_{\text{good}}$, generated by adding to $\mathbf{W}^*$ uniform noise with support $[0, 0.1]$, and $\mathbf{W}^0_{\text{bad}}$, generated independent of $\mathbf{W}^*$ by uniform noise with support $[0, 0.1]$.

Given $\mathbf{W}^0_{\text{bad}}$, Figure 6 illustrates the performance of different algorithms in reconstructing a 3×3 phantom sample with only 3 translation and 4 angular scans. To be more specific, Figure 6a shows the reduction of reconstruction error as defined in Equation (5.1) from three

13
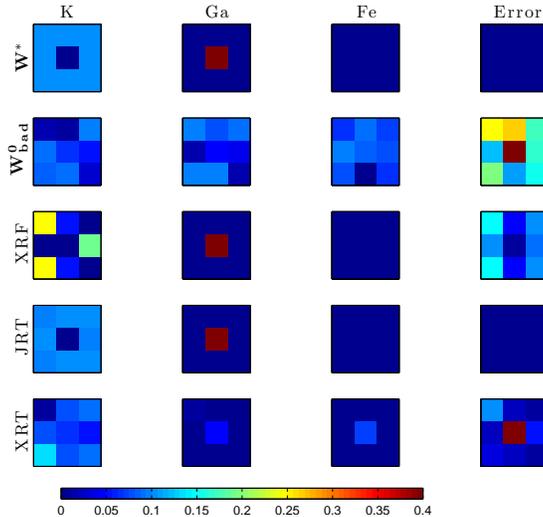
Figure 7: Comparison of reconstruction results for optimization starting from $\mathbf{W}^0_{\text{bad}}$ for the 3×3 phantom.

different algorithms with respect to the number of (function, gradient) evaluations. We observe that JRT produces an error that is an order of magnitude smaller than XRT or XRF.

Moreover, owing to the ill-posed nature of both XRF and XRT, given an initial guess, the solution manifold of XRF does not agree with that of XRT in general; that is, minimizing ϕ^R and ϕ^T separately is unlikely to return a common solution. Since global solutions are unlikely to be obtained, this situation can have severe effects in practice. For example, Figure 6b illustrates that the single-modality inversion approaches get stuck at poor local minima of their respective objectives. In contrast, joint inversion not only converges much faster than the two single-modality inversions, the local minimum found also achieves a much better reconstruction result in terms of both the XRT and XRF residual objectives. The corresponding reconstructed elemental maps, along with the resulting summed reconstruction errors, are shown in Figure 7. Since the performance of XRT is poor for fluorescence tomography, for the remainder of our experiments we report only the reconstruction qualities of JRT and XRF, the approach mostly used in practice.

Next, given $\mathbf{W}^0_{\text{good}}$, we increase the resolution of the phantom sample to 10×10 pixels with $|\Theta| = 6$ angular scans. Figure 8 shows the convergence results for different algorithms with respect to solution error and function value residual. Correspondingly, Figure 9 illustrates the reconstructed elemental maps. With an even finer resolution of 20×20 pixels with 10 angular scans, given $\mathbf{W}^0_{\text{bad}}$, the corresponding convergence performance and reconstructed elemental maps are shown in Figures 10 and 11, respectively. From these tests, we can see that, with roughly the same computational cost, JRT consistently returns results with better accuracy than does XRF. In particular, even for the highly fluorescing element iron, JRT returns a reconstruction error with an 83% improvement compared with XRF only.
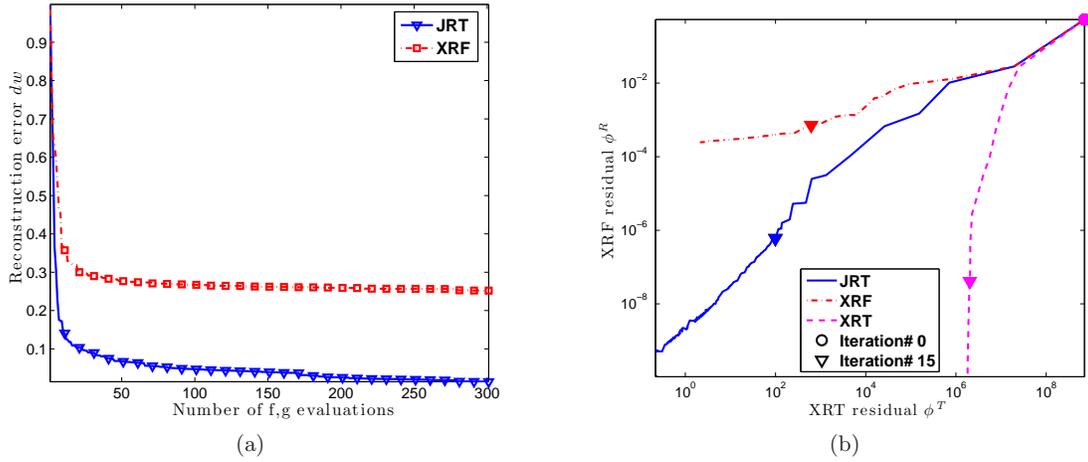
Figure 8: Optimization starting from $\mathbf{W}^0_{\text{good}}$ for the 10×10 phantom sample: (a) Reduction of reconstruction error with respect to the number of (function, gradient) evaluations; (b) Trajectories of the XRF and XRT residuals for the three reconstruction approaches.
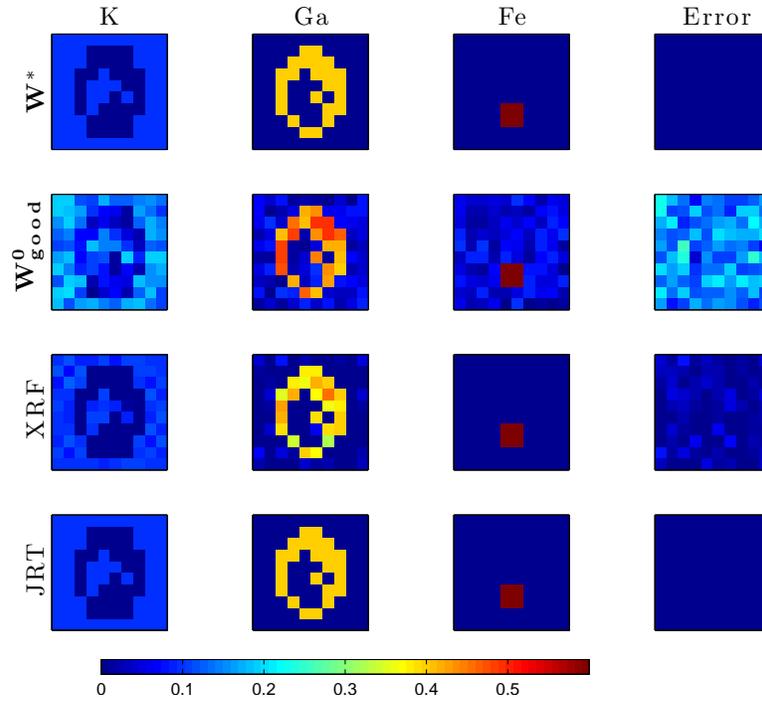
Figure 9: Comparison of reconstruction results for optimization starting from $\mathbf{W}^0_{\text{good}}$ for the 10×10 phantom.
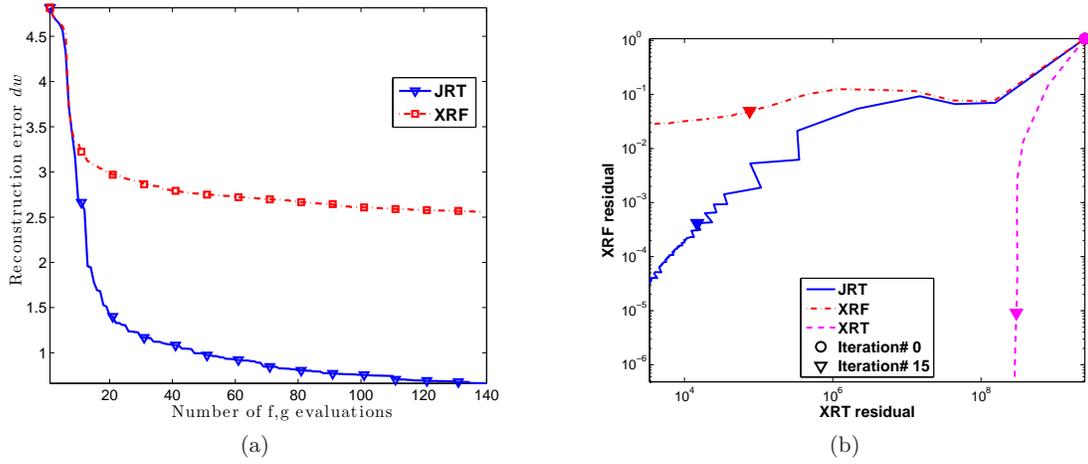
Figure 10: Optimization starting from $\mathbf{W}_{\text{bad}}^0$ for the 20×20 phantom sample: (a) Reduction of reconstruction error with respect to the number of (function, gradient) evaluations; (b) Trajectories of the XRF and XRT residuals for the three reconstruction approaches.
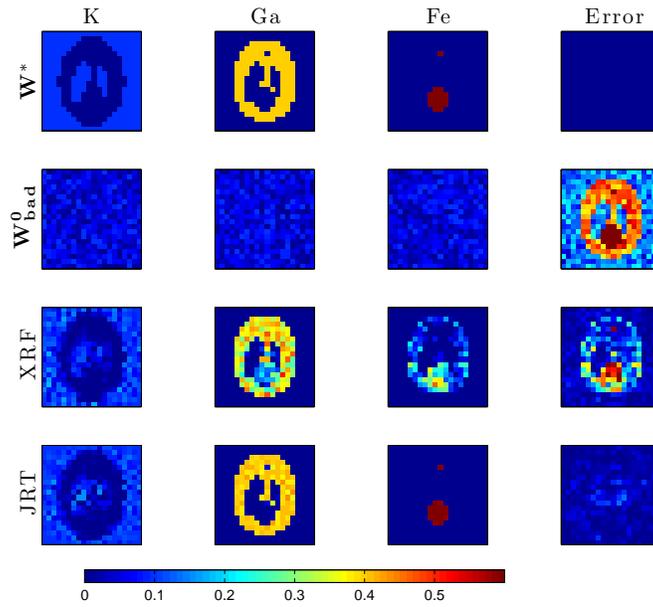
Figure 11: Comparison of reconstruction results for optimization starting from $\mathbf{W}_{\text{bad}}^0$ for the 20×20 phantom.
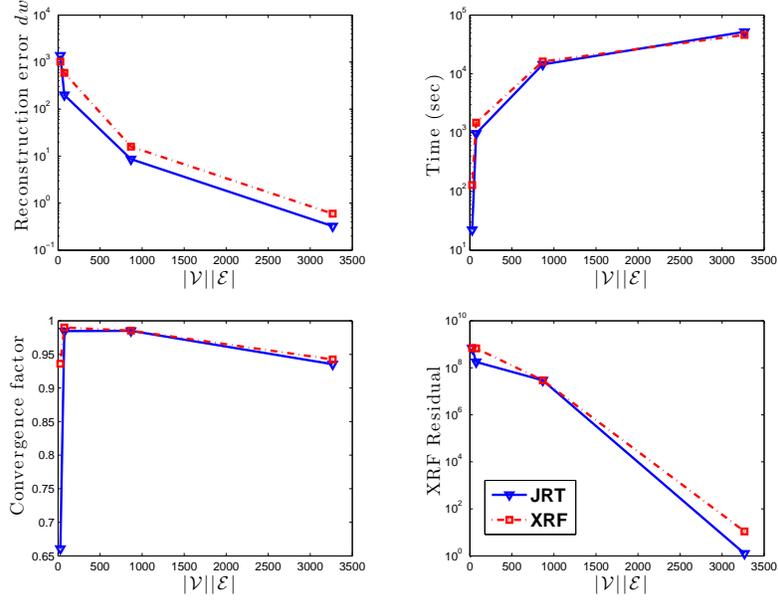
Figure 12: Performance of different reconstruction algorithms for increased resolution size and a fixed number of angular projections.

In the case of limited angular scans (i.e., $|\Theta| = 2$), we also compare the convergence factors of JRT and XRF (given the same stopping criteria) for an increasing size of sample resolution. The convergence factor here is

$$
C_k = \left(\frac{\phi(\mathbf{W}^{(k+1)}) - \phi(\mathbf{W}^*)}{\phi(\mathbf{W}^{(1)}) - \phi(\mathbf{W}^*)} \right)^{\frac{1}{k+1}},
$$

where $\mathbf{W}^*$ is the synthetic solution at the highest resolution down interpolated to the resolution of $\mathbf{W}^{(k)}$. Figure 12 shows the corresponding results. As we increase the problem resolution (i.e., the number of voxels $|\mathcal{V}|$), the convergence factor and the XRF objective residual are always better with JRT than with XRF. We thus observe that, in the case of the limited data, JRT is obtaining better reconstruction results than XRF and doing so with a performance that does not depend more strongly on problem size than does XRF.

Given $\mathbf{W}^0_{\text{good}}$ as the initial guess, we also compare the reconstruction ability of XRF and JRT when we consider contaminated data by adding standard Gaussian noise to the simulated measurement data. Figure 13 demonstrates that JRT converges faster than single XRF and XRT, in addition to providing improved residual and error. The corresponding reconstructed elemental maps are given in Figure 14.

5.4. Phase Sample Results. Now we examine the performance of JRT on a sample normally used to explore the self-absorption effect in fluorescence imaging; see Figure 5. Notably, we address not only the independent performance of JRT with respect to different samples
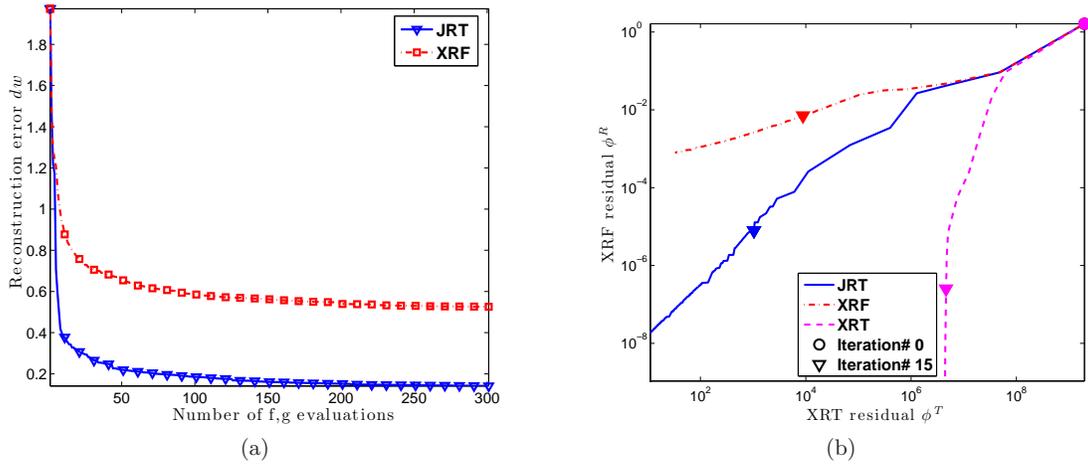
Figure 13: Optimization starting from $\mathbf{W}^0_{\mathrm{good}}$ for the 20×20 phantom sample with noisy data: (a) Reduction of reconstruction error with respect to the number of (function, gradient) evaluations; (b) Trajectories of the XRF and XRT residuals for the three reconstruction approaches.
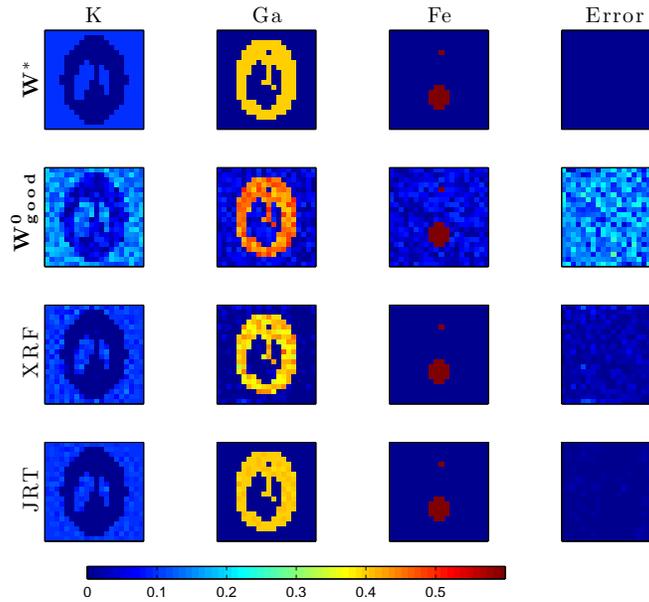
Figure 14: Comparison of reconstruction results for optimization starting from $\mathbf{W}^0_{\mathrm{good}}$ for the 20×20 phantom with noisy data.

(a) Error Reduction

(b) Residual Reduction
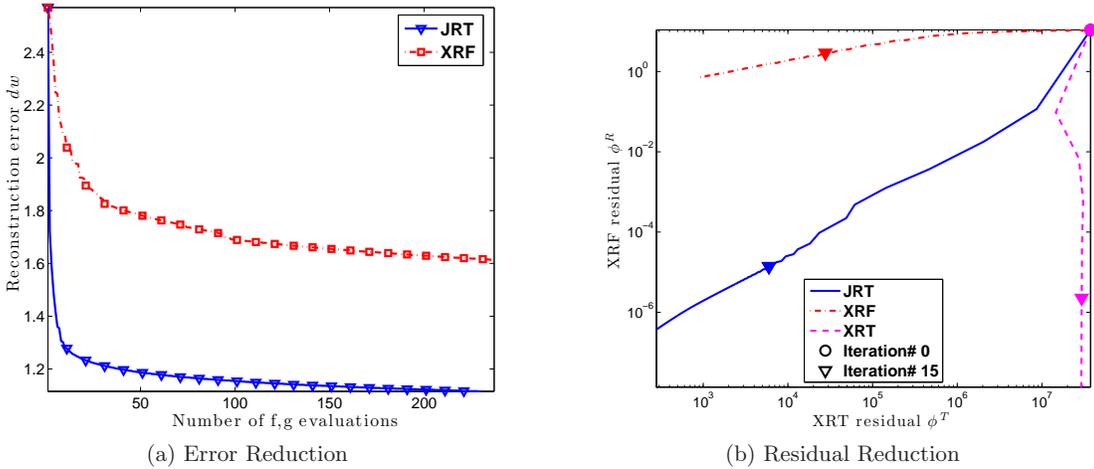
Figure 15: A 20×20 phase sample from $\mathbf{W}^0_{\text{bad}}$. (a) Reduction of reconstruction error with respect to the number of function gradient evaluations; (b) Trajectories of the value ϕ^R with respect to the value ϕ^T for each iteration.

but also the advantage of our XRF forward model by including the self-absorption effect.

First, analogous to previous experiments, we compare the performance of JRT with single XRF and XRT on a 20×20 phase sample with 20 translation and 12 angular scans. Again, JRT beats XRF and XRT in reducing both the objective function residuals and reconstruction error as shown in Figure 15. The resulting elemental maps are shown in Figure 16.

In Figure 17, we can see that with self-absorption correction, the reconstructed result is much more accurate. In particular, without self-absorption correction, the mass in the inner region of the shapes is underestimated because of the self-absorption effect.

6. Comments and Conclusions. This paper addresses a joint inversion framework to solve tomography problems by using multimodality data to improve the ill-conditioned nature of such problems. In detail, we model the X-ray fluorescence and X-ray transmission phenomena separately as nonlinear equations in terms of chemical element composition of a given sample. We also formulate reconstruction of the composition from given XRF and XRT tomography data as an optimization-based joint inversion problem by utilizing multimodality data. To solve the optimization problem, we use the truncated-Newton method for minimizing the sum of errors between measured and synthetic data generated by our forward models.

The experimental results presented in the paper show that significant improvements are achieved by performing joint inversion for limited data sources in terms of both convergence speed and accuracy during the reconstruction process.
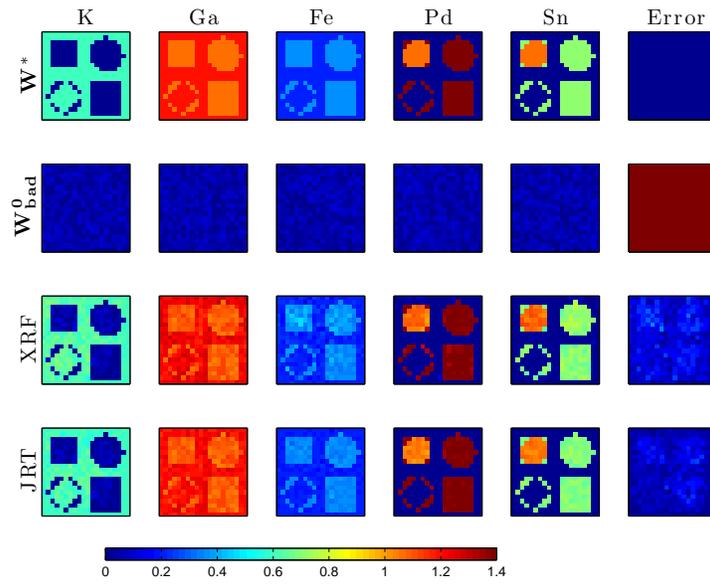
Figure 16: Reconstruction results of the 20×20 phase sample reconstruction.
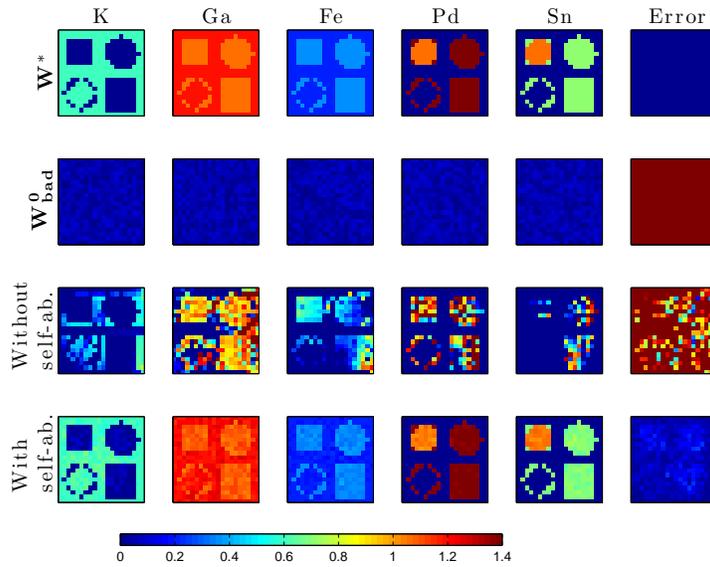
Figure 17: Solution of the phase sample reconstruction with and without self-absorption correction using JRT.

REFERENCES

[1] Mario Bertero and Patrizia Boccacci, *Introduction to inverse problems in imaging*, CRC Press, 1998.

[2] Christian H. Bischof, H. Martin Bücker, Bruno Lang, A. Rasch, and Andre Vehreschild, *Combining source transformation and operator overloading techniques to compute derivatives for MATLAB programs*, in Proceedings of the Second IEEE International Workshop on Source Code Analysis and Manipulation (SCAM 2002), Los Alamitos, CA, USA, 2002, IEEE Computer Society, pp. 65–72.

[3] Roberto Cesareo and Sergio Mascarenhas, *A new tomographic device based on the detection of fluorescent X-rays*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 277 (1989), pp. 669–672.

[4] Mark E. Davison, *The ill-conditioned nature of the limited angle tomography problem*, SIAM Journal on Applied Mathematics, 43 (1983), pp. 428–448.

[5] Junjing Deng, David J. Vine, Si Chen, Youssef S. G. Nashed, Qiaoling Jin, Nicholas W. Phillips, Tom Peterka, Rob Ross, Stefan Vogt, and Chris J. Jacobsen, *Simultaneous cryo X-ray ptychographic and fluorescence microscopy of green algae*, in Proceedings of the National Academy of Sciences, 112 (2015), pp. 2314–2319.

[6] Heinz W. Engl and Philipp Kügler, *Nonlinear inverse problems: theoretical aspects and some industrial applications*, in Multidisciplinary Methods for Analysis Optimization and Control of Complex Systems, Springer, 2005, pp. 3–47.

[7] Jeffrey A. Fessler, *Penalized weighted least-squares image reconstruction for positron emission tomography*, IEEE Transactions on Medical Imaging, 13 (1994), pp. 290–300.

[8] Bruno Golosio, Alexandre Simionovici, Andrea Somogyi, Laurence Lemelle, Marina Chukalina, and Antonio Brunetti, *Internal elemental microanalysis combining X-ray fluorescence, Compton and transmission tomography*, Journal of Applied Physics, 94 (2003), pp. 145–156.

[9] Doğa Gürsoy, Tekin Biçer, Antonio Lanzirotti, Matthew G Newville, and Francesco De Carlo, *Hyperspectral image reconstruction for x-ray fluorescence tomography*, Optics Express, 23 (2015), pp. 9014–9023.

[10] Per Christian Hansen and Dianne Prost O'Leary, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM Journal on Scientific Computing, 14 (1993), pp. 1487–1503.

[11] John P. Hogan, Robert A. Gonsalves, and Allen S. Krieger, *Fluorescent computer tomography: A model for correction of X-ray absorption*, IEEE Transactions on Nuclear Science, 38 (1991), pp. 1721–1727.

[12] Avinash C. Kak and Malcolm Slaney, *Principles of Computerized Tomographic Imaging*, IEEE Press, 1988.

[13] Patrick La Rivière, *Accurate analytic reconstruction in X-ray fluorescence computed tomography*, in the 2002 IEEE International Symposium on Biomedical Imaging, Proceeding, 2002, pp. 637–640.

[14] Patrick La Rivière and David M. Billmire, *Penalized-likelihood image reconstruction for X-ray fluorescence computed tomography with unknown fluorescence attenuation maps*, International Society for Optics and Photonics, 2004, pp. 243–252.

[15] Patrick La Rivière and Phillip A. Vargas, *Monotonic penalized-likelihood image reconstruction for X-ray fluorescence computed tomography*, IEEE Transactions on Medical Imaging, 25 (2006), pp. 1117–1129.

[16] Eduardo X. Miqueles and Alvaro R. De Pierro, *Iterative reconstruction in X-ray fluorescence tomography based on Radon inversion*, IEEE Transactions on Medical Imaging, 30 (2011), pp. 438–450.

[17] Arrate Munoz-Barrutia, Carlos Pardo-Martin, Thomas Pengo, and Carlos Ortiz-de Solorzano, *Sparse algebraic reconstruction for fluorescence mediated tomography*, International Society for Optics and Photonics, 2009, pp. 744604–744604.

[18] Stephen G. Nash, *A survey of truncated-Newton methods*, Journal of Computational and Applied Mathematics, 124 (2000), pp. 45–59.

[19] Arnold Neumaier, *Solving ill-conditioned and singular linear systems: A tutorial on regularization*, SIAM Review, 40 (1998), pp. 636–666.

[20] Tatjana Paunesku, Stefan Vogt, Jörg Maser, Barry Lai, and Gayle Woloschak, *X-ray fluorescence microprobe imaging in biology and medicine*, Journal of Cellular Biochemistry, 99 (2006), pp. 1489–1502.

[21] Johann Radon, *On the determination of functions from their integral values along certain manifolds*, IEEE Transactions on Medical Imaging, 5 (1986), pp. 170–176.

[22] David Rendell and David J. Mowthorpe, *Fluorescence and phosphorescence spectroscopy*, Published on behalf of ACOL, London by Wiley, 1987.

[23] Georg-Friedemann Rust and Jörg Weigelt, *X-ray fluorescent computer tomography with synchrotron radiation*, IEEE Transactions on Nuclear Science, 45 (1998), pp. 75–88.

[24] Alex Sawatzky, Qiaofeng Xu, Carsten Schirra, and Mark Anastasio, *Proximal ADMM for multi-channel image reconstruction in spectral X-ray CT*, IEEE Transactions on Medical Imaging, 33 (2014), pp. 1657–1668.

[25] Tom Schoonjans, Antonio Brunetti, Bruno Golosio, Manuel Sanchez del Rio, Vicente Armando Solé, Claudio Ferrero, and Laszlo Vincze, *The xraylib library for X-ray–matter interactions. Recent developments*, Spectrochimica Acta Part B: Atomic Spectroscopy, 66 (2011), pp. 776–784.

[26] Lawrence A. Shepp and Benjamin F. Logan, *The Fourier reconstruction of a head section*, IEEE Transactions on Nuclear Science, 21 (1974), pp. 21–43.

[27] Jacob Sherman, *The theoretical derivation of fluorescent X-ray intensities from mixtures*, Spectrochimica Acta, 7 (1955), pp. 283–306.

[28] Albert Thompson et al., eds., *X-ray Data Booklet*, Lawrence Berkeley National Laboratory, University of California, 2001.

[29] D. J. Vine, D. Pelliccia, C. Holzner, S. B. Baines, A. Berry, I. McNulty, S. Vogt, A. G. Peele, and K. A. Nugent, *Simultaneous X-ray fluorescence and ptychographic microscopy of cyclotella meneghiniana*, Optics Express, 20 (2012), pp. 18287–18296.

[30] Yilun Wang, Junfeng Yang, Wotao Yin, and Yin Zhang, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM Journal on Imaging Sciences, 1 (2008), pp. 248–272.

[31] John Paul Ward, Minji Lee, Jong Chul Ye, and Michael Unser, *Interior tomography using 1d generalized total variation. part i: Mathematical foundation*, SIAM Journal on Imaging Sciences, 8 (2015), pp. 226–247.

[32] Qun Yang, Biao Deng, Weiwei Lv, Fei Shen, Rongchang Chen, Yudan Wang, Guohao Du, Fuhua Yan, Tiqiao Xiao, and Hongjie Xu, *Fast and accurate X-ray fluorescence computed tomography imaging with the ordered-subsets expectation maximization algorithm*, Journal of Synchrotron Radiation, (2012), pp. 210–215.

[33] Tetsuya Yuasa, Masahiro Akiba, Tohoru Takeda, Masahiro Kazama, Atsunori Hoshino, Yuki Watanabe, Kazuyuki Hyodo, F. Avraham Dilmanian, Takao Akatsuka, and Yuji Itai, *Reconstruction method for fluorescent X-ray computed tomography by least-squares method using singular value decomposition*, IEEE Transactions on Nuclear Science, 44 (1997), pp. 54–62.

[34] Tetsuya Yuasa, Tohoru Takeda, F. Avraham Dilmanian, Masahiro Akiba, Kazuyuki Hyodo, Takao Akatsuka, and Yuji Itai, *Fast algorithm for fluorescent X-ray CT reconstruction*, in Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1997, vol. 2, 1997, pp. 863–866.