

ARGONNE NATIONAL LABORATORY  
9700 South Cass Avenue  
Argonne, IL 60439

**Optimization-Based Approach for Joint X-ray  
Fluorescence and Transmission Tomographic Inversion**

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

Mathematics and Computer Science Division

Preprint ANL/MCS-P5327-0415

May 2015 (*Revised, September 2015*)

# Optimization-Based Approach for Joint X-ray Fluorescence and Transmission Tomographic Inversion

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 X-ray transmission tomography can be used for reconstructing the spatial distribution of the absorption coefficient inside a sample. In this work, we integrate both X-ray fluorescence and X-ray transmission data modalities and formulate a nonlinear optimization-based approach for reconstruction of the elemental composition of a given object. 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 (i.e., the X-ray transmission or fluorescence alone), the joint inversion provides a better-posed problem, which implies a better recovery. Therefore, the challenges in X-ray fluorescence tomography arising mainly from the effects of self absorption 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 physics, chemistry, and biology to astronomy. The technique requires an accurate image reconstruction, however, and the resulting reconstruction problem typically is ill-posed and does not have a unique solution because of insufficient measurements. Different modalities of tomograms have been derived by using different physical phenomena, such as computed tomography (CT), positron emission tomography (PET), and magnetic resonance imaging (MRI). 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 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 [20, 33, 38]. However, the reconstruction problem for XRF tomography is much more difficult than it is for XRT tomography, because of the absorption of the photons along the excitation and detection paths.

The first reconstruction algorithm used for XRF imaging was filtered backprojection (FBP) [4]. Alternatively, in order to incorporate more accurate imaging models, iterative

---

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

methods are often preferred for XRF reconstruction since they generally reduce sensitivity to noise and can reconstruct a high-quality image in the case of incomplete data. In particular, the algebraic reconstruction technique (ART) [15,29,38,49] and the expectation-maximization algorithm [16,25,26,47] have been widely applied for XRF reconstruction.

One phenomena that limits XRF reconstruction is the self-absorption effect, which refers to the fact that emission at a particular point in space can be reabsorbed as it travels through space. As a consequence, another research effort has focused on achieving a more accurate approach with a correction for attenuation. This effort involves simultaneously estimating emission and attenuation maps in conventional emission tomography when no XRT scan is performed. The correction of fluorescent attenuation was first introduced by Hogan et al. [20]. In [25,26], La Rivière et al. propose a penalized-likelihood reconstruction for XRF with an approximate expression for the XRF attenuation map as a linear combination of known quantities and an element’s unknown distribution. In [27], Miqueles and De Pierro develop an alternating method for retrieving simultaneously the XRF density and the attenuation coefficients for XRF reconstruction.

A common characteristic of these XRF reconstruction approaches is that the linear attenuation coefficients of the sample either are assumed to be known or are approximated from a single imaging modality. Recent advances in simultaneous acquisition of different modalities, however, allow one to better incorporate many types of a priori information via a joint inversion. Joint inversion of two different data sets was formalized by Haber and Oldenburg [18] in 1997. This concept has been widely applied to the geophysical sciences [11,13,14,17,21,23,32] where different ways of coupling priors have been discussed. Subsequently, Ehrhardt et al. [8] introduced a joint inversion tomographic framework to simultaneously reconstruct samples from PET and MRI data.

In this work, we develop a new nonlinear model to correct the self-absorption effect in the XRF reconstruction, 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 approach has the potential to provide the distribution of the heavy elements that emit detectable fluorescent lines, even in the presence of very light 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 source 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 (see Figure 1a). 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 (see Figure 2a) in order to minimize contamination by Compton scattering photons from the incident beam. The

detector outputs a spectrum comprising multiple peaks [35]. These peaks correspond to the principal emission lines of the various elements present in the sample, since the number of detected XRF 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 [34], 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 particular, we use  $\Theta$  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 (in cm) of beamlet  $(\theta, \tau)$  with the voxel  $v \in \mathcal{V}$ . We use  $\mathcal{E}$  to denote the collection of  $|\mathcal{E}|$  possible element indices and  $\mu_e^E$  to denote the mass attenuation coefficient (in  $\text{cm}^2\text{g}^{-1}$ ) of element  $e$  at beam incident energy  $E$ .

Our goal is to recover the tensor  $\mathbf{W} = [W_{v,e}]$  denoting the density (in  $\text{g cm}^{-3}$ ) 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 for the 2D case. A traditional way (see, e.g., [22]) to model the XRT projection of a sample  $\tilde{F}_{\theta,\tau}^T$  (in units of counts per second) from beamlet  $(\theta, \tau)$  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 incident X-ray intensity (in units of counts per second) and  $\boldsymbol{\mu} = [\mu_v^E]$  is the linear attenuation coefficient (in  $\text{cm}^{-1}$ ) 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

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

Equation 2.1 plays an essential role in our joint inversion method. It not only links the XRF and XRT with the common variable  $\mathbf{W}$ , so that the correlation of these two modalities can be better explored, but also allows a better reconstruction of  $\boldsymbol{\mu}$  given a better reconstructed  $\mathbf{W}$ , which is the goal of joint inversion. Therefore, in principle, joint inversion should also provide a better reconstructed attenuation coefficients over XRT reconstruction alone.

**2.2. Discrete XRF Imaging Model.** The discrete model we use to simulate XRF is built on the one proposed in [26], and the corresponding geometry is illustrated in Figure 2. Two main features distinguish our model from the one in [24, 26]. First, instead of modeling the

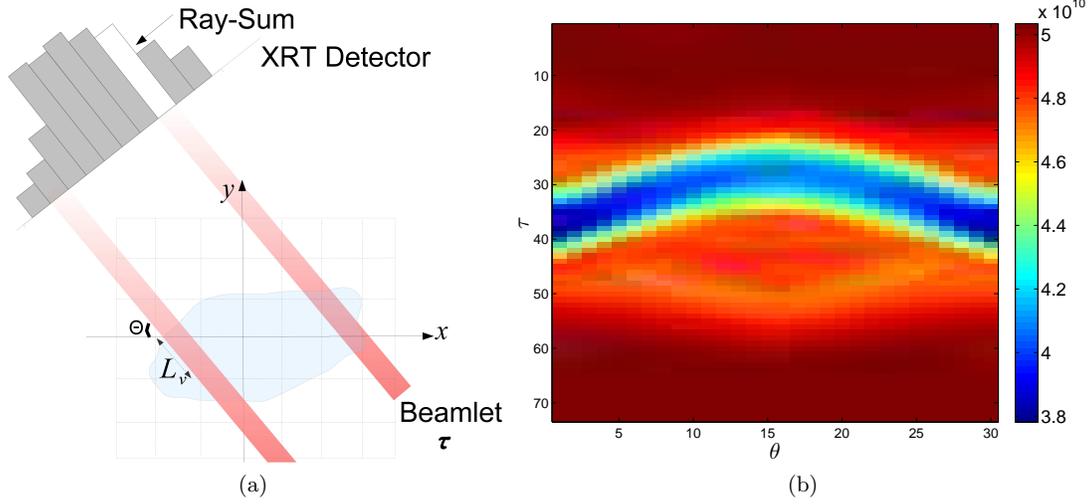


Figure 1: (a) Illustration of the discrete XRT projection geometry: the sample resides in space (discretized in 2D for simplicity of illustration), whereas the beam is parameterized by its angular and translation scan (indexed by  $\theta$  and  $\tau$ , respectively); (b) Corresponding structure of XRT data for all  $|\Theta||\mathcal{T}|$  projections.

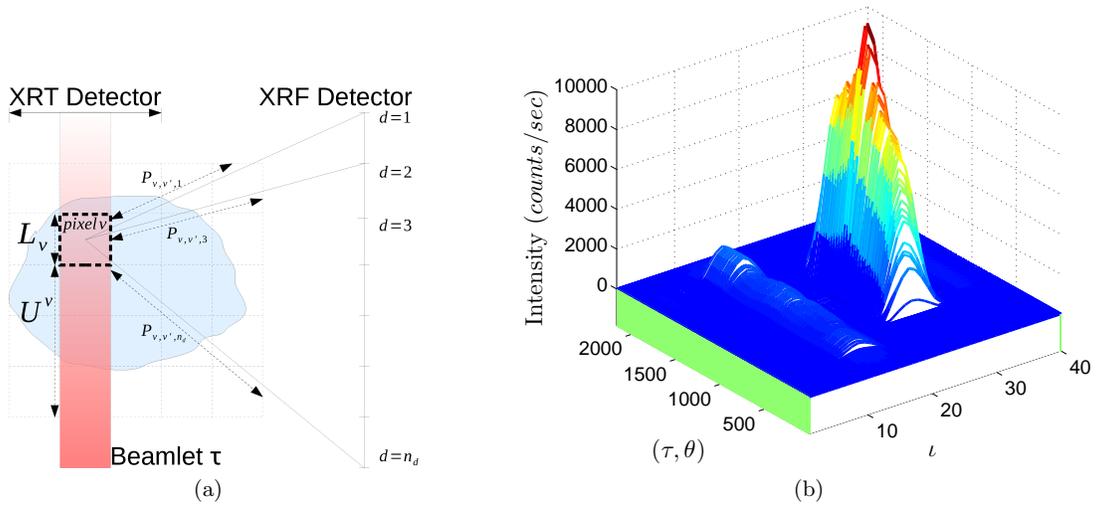


Figure 2: (a) Geometry and notation of XRF setup; (b) Corresponding structure of XRF data, with each projection  $(\theta, \tau)$  yielding a spectrum indexed by  $\iota \in \{1, \dots, |\mathcal{I}|\}$ .

XRF process in terms of both the elemental distribution and the attenuation map of XRF energy, we model the XRF only in terms of the elemental distribution. Second, rather than modeling the detected XRF photons directly, we follow an elemental approach, whereby we model the XRF emitted from an element by its corresponding elemental unit spectrum. Then, the total XRF spectrum detected from the given sample is modeled as a weighted sum of the unit spectra of the elements being recovered.

To approximate the elemental emission spectrum, we adapt Sherman's equation [40]. Sherman's equation calculates the net X-ray fluorescence intensities emitted by each element in a specimen of a known composition when it is irradiated by a polychromatic X-ray beam. Sherman's equation is the most fundamental analytical technique to link elemental composition and intensity in quantitative XRF analysis. Unfortunately, this fundamental equation has limitations. For example, it does not account for the X-ray tube current, the reflectivity of the analyzing crystal, and the detector efficiency; it also ignores the contributions due to the scattering of both the incident and the fluorescence X-ray radiations inside the object. However, the model performs well in practice given an appropriate calibration procedure [37]. Based on Sherman's equation, a first-order approximation (i.e., neglecting effects such as Rayleigh and Compton scattering) of the net XRF intensity corresponding to the characteristic XRF 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 XRF 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  $\omega_{e,l}$ ,  $\mu_e^E$ , and  $r_{e,s}$  include the tabulated data [41] and the online database xraylib [39], which provides library functions that can be easily included in software applications with a focus on quantitative X-ray fluorescence applications. The database is a compilation of data sets from different published works including the NIST and EPDL databases. In our calculations, we use 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 XRF, let  $\mathcal{I}$  denote the set of the complete collection of its  $|\mathcal{I}|$  energy channel indices, let  $\iota \in \mathcal{I}$  denote the index of the energy channel, and let  $\mathbf{x} = [x_\iota]$  be the  $|\mathcal{I}|$ -dimensional vector denoting its energy channels. Let  $\mathbf{1}_{E_e}^{\mathbf{x}}$  be the  $|\mathcal{I}|$ -dimensional unit indicator vector with its  $i$ th component defined by

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

where we choose to break ties between neighboring channels  $(\iota, \iota - 1)$  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 [41]. The width of this Gaussian-like peak determines the capability to distinguish particles with different energies. The energy resolution

$\Delta E$  is given by the full width at half maximum (FWHM) of the single peak. For a Gaussian distribution with standard deviation  $\sigma$ , we have  $\text{FWHM} \approx 2.35\sigma$ . Therefore, the final unit spectrum obtained by convolving the perfect spectral line  $\mathbf{I}^x$  with a Gaussian distribution can be obtained via the convolution theorem as

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

where  $*$  denotes pointwise multiplication (Hadamard product) and  $\mathcal{F}$  ( $\mathcal{F}^{-1}$ ) is the (inverse) Fourier transform. In our simulation, we assume  $\Delta E$  to be constant for a given detector. In the case of unequally distributed energy channels,  $\Delta E$  and  $\sigma$  can be modeled separately as vectors with the same size as the energy channels. Furthermore, we consider only  $K_\alpha, K_\beta, L_\alpha$ , and  $L_\beta$  lines (see [41]), 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 XRF spectrum of a sample with multiple elements by considering the attenuation of the beam energy and self-absorption effect of the XRF energy. We let  $A_v^{E,\theta,\tau}$  represent the attenuation experienced by beamlet  $(\theta, \tau)$  (at incident beam energy  $E$ ) as it travels toward voxel  $v \in \mathcal{V}$ :

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

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  $(\theta, \tau)$  before it enters voxel  $v$ .

We model the self-absorption effect, i.e., the attenuation of XRF energy emitted for beamlet  $(\theta, \tau)$  from element  $e \in \mathcal{E}$  at voxel  $v \in \mathcal{V}$  before it reaches the detector, and denote it by  $F_{v,e}^{\theta,\tau}(\mathbf{W})$ . Note that the energy-dispersive detectors typically used to detect XRF 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 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 XRF detector. By numerically subdividing the fluorescence detector into  $n_d$  detectorlets, we approximate the self-absorbed XRF energy by

$$(2.2) \quad F_{v,e}^{\theta,\tau}(\mathbf{W}) = \frac{1}{n_d} \sum_{d=1}^{n_d} \exp \left\{ - \sum_{v'} \sum_{e'} W_{v',e'} \mu_{e'}^{E_e} P_{v,v',d}^{\theta,\tau} \right\},$$

where  $\mathbf{P} = [P_{v,v',d}^{\theta,\tau}]$  denotes a tensor containing the intersection length (in cm) of the XRF detectorlet path  $d$  (originating from voxel  $v$  and beamlet  $(\theta, \tau)$ ) with the voxel  $v'$  and  $\mu_{e'}^{E_e}$  is the linear attenuation coefficient of element  $e'$  at the XRF energy  $E_e$  of element  $e$ . In the case of no self-absorption correction, we set

$$(2.3) \quad F_{v,e}^{\theta,\tau}(\mathbf{W}) = 1.$$

Accordingly, the fluorescence spectrum  $\mathbf{F}_{\theta,\tau}^R$  (in units of counts per second) of the object resulting from beamlet  $(\theta, \tau)$  is the  $|\mathcal{Z}|$ -dimensional vector

$$\begin{aligned}
\mathbf{F}_{\theta,\tau}^R(\mathbf{W}) &= \sum_e \left( \sum_v L_v^{\theta,\tau} A_v^{E,\theta,\tau}(\mathbf{W}) F_{v,e}^{\theta,\tau}(\mathbf{W}) W_{v,e} \right) \mathbf{M}_e \\
(2.4) \quad &= \sum_{v,e,d} \frac{L_v^{\theta,\tau} W_{v,e} \mathbf{M}_e}{n_d} \exp \left\{ - \sum_{v',e'} W_{v',e'} \left( \mu_{e'}^E L_{v'}^{\theta,\tau} \mathbb{1}_{v' \in \mathcal{U}_v^{\theta,\tau}} + \mu_{e'}^{E_e} P_{v,v',d}^{\theta,\tau} \right) \right\}.
\end{aligned}$$

**3. Optimization-Based Reconstruction Formulations and Algorithms.** We now describe the optimization-based approach for solving inverse problems involving the models  $F_{\theta,\tau}^T(\mathbf{W})$  and  $\mathbf{F}_{\theta,\tau}^R(\mathbf{W})$ . We assume that, for a set of  $(\theta, \tau)$  values, the respective data for these models are given by

- $D_{\theta,\tau}^T \in \mathbb{R}$ , the measurement data of XRT detected at angle  $\theta$  from light beam  $\tau$ , and
- $\mathbf{D}_{\theta,\tau}^R \in \mathbb{R}^{n_E}$ , the measurement data of XRF detected at angle  $\theta$  from light beam  $\tau$ .

Each of the approaches follows a constrained least-squares formulation [10, 48],

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

where  $\mathbf{x}$  denotes the  $n \leq m$  parameters,  $\mathbf{r} : \mathbb{R}^n \mapsto \mathbb{R}^m$  denotes the residual mapping, and  $\mathcal{X} \subseteq \mathbb{R}^n$  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

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

where the constraint  $\mathbf{W} \geq \mathbf{0}$  is due to the physical nature of mass and  $\beta > 0$  is the scaling between the two modalities that plays an important role in the inversion process because it balances the experimental errors of the two modalities. In this work, in the context of XRF and XRT reconstruction,  $\beta$  is chosen from the perspective of physics so that the magnitudes of the two data modalities are comparable; from a maximum-likelihood-estimate point-of-view, this is equivalent to balancing the modeling and measuring variability of each modality. In general, and especially when the iterates are far from optimal, the selection of  $\beta$  can be accomplished by techniques summarized in [44], such as the generalized cross-validation and the L-curve method. Here we focus on techniques for solving (3.1) for a given value of  $\beta$  that can be naturally extended to provide efficient solutions for a variety of other  $\beta$  values.

We note that nonlinear transformations of the residuals is also possible. For example, one could consider a log-transformation of  $\phi^T$  to obtain a linear term that might induce a problem

that is easier to solve. However, in order to make the joint inversion formula consistent between the modalities and natural in terms of physics, here we keep  $\phi^T$  in its original form.

The first-order derivative of the objective function in (3.1) with respect to  $\mathbf{W}$  is

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

The first-order derivatives of the XRT and XRF forward models are provided in Appendix A. For  $|\Theta|$  angular and  $|\mathcal{T}|$  translation scans,  $\mathbf{J}^T$  is the  $|\Theta||\mathcal{T}| \times |\mathcal{V}||\mathcal{E}|$  Jacobian matrix of  $\phi^T$  whose components are given by  $[\mathbf{J}^T(\mathbf{W})]_{i,j} = \frac{\partial F_i^T(\mathbf{W})}{\partial W_j}$ , where  $i$  and  $j$  index the vectorizations of the indices  $(\theta, \tau)$  and  $(v, e)$ , respectively. Similarly, we let  $\mathbf{J}^R$  denote the  $|\Theta||\mathcal{T}||\mathcal{I}| \times |\mathcal{V}||\mathcal{E}|$  Jacobian matrix of  $\phi^R$  whose components are given by  $[\mathbf{J}^R(\mathbf{W})]_{i,j} = \frac{\partial F_i^R(\mathbf{W})}{\partial W_j}$ , where  $i$  and  $j$  index the vectorizations of the indices  $(\theta, \tau, \iota)$  and  $(v, e)$ , respectively. Consequently, the master Jacobian matrix and master residual vector of (3.1) are  $\mathbf{J}(\mathbf{W}) = \begin{bmatrix} \mathbf{J}^R(\mathbf{W}) \\ \beta\mathbf{J}^T(\mathbf{W}) \end{bmatrix}$  and  $\mathbf{r}(\mathbf{W}) = \begin{bmatrix} \mathbf{F}^R(\mathbf{W}) - \mathbf{D}^R \\ \beta(\mathbf{F}^T(\mathbf{W}) - \mathbf{D}^T) \end{bmatrix}$ , 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

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

The second-order sufficient condition additionally requires that

$$(3.2) \quad \mathbf{z}^\top \nabla^2\phi(\mathbf{W}^*)\mathbf{z} > 0$$

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

**3.3. A Motivating Example.** 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 [5]. The most commonly used remedy for this ill-posed nature is to add a regularization term (e.g., Tikhonov regularization [1, 45, 46]). An appropriately chosen regularization can facilitate direct numerical solution; however, choosing a good regularizer is difficult [19, 31]. In this section, we examine the Jacobian of the nonlinear least-squares formulations involving  $\mathbf{F}^R$  and  $\mathbf{F}^T$ , and we provide an example motivating the use of joint inversion 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 (3.2) 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 [9]. This case arises often in traditional tomographic reconstructions because the data is often insufficient and noisy.

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 ( $3 \times 3$  discretized) sample composed of the single element Ca. It is imaged by using  $|\mathcal{I}| = 2$  energy channels, one angular ( $\theta = 0^\circ$ ), and three translation scans, for which  $W_{v,e}^* = 1$ ,  $v \in [1, \dots, 9]$ , and  $e = 1$ . The Jacobian at  $\mathbf{W}^*$ ,

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}^R \\ \beta \mathbf{J}^T \end{bmatrix},$$

decomposed into parts corresponding to  $\phi^R$  and  $\beta\phi^T$ , is given by

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

We note that

- $B_s(\mathbf{W}^*) = \emptyset$ ;
- $\text{rank}(\mathbf{J}^R) = 6$ ;
- $\text{rank}(\mathbf{J}^T) = 3$ ; and
- $\text{rank}(\mathbf{J}) = 9$ , 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  $\beta > 0$ . On the other hand, the single modality inversions XRT and XRF are underdetermined and have many connected local minima.

Similarly, if we increase the number of angular scans to two (i.e.,  $|\Theta| = 2$ ), then XRF results in an overdetermined system with the corresponding  $\mathbf{J}^R$  of size  $12 \times 9$  and  $\mathbf{J}^T$  of size  $6 \times 9$ . However, since  $\text{rank}(\mathbf{J}^R) = 6$  and  $\text{rank}(\mathbf{J}^T) = 6$ , the XRF system is still 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 a 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  $\beta = 10^8$ . 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,  $\mathbf{W}^*$  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 (numerically) full-rank Jacobian but that the (numerical) rank is often improved. Therefore, the area of the solution manifold of JRT is often smaller than the one from a single-modality inversion; as we will demonstrate in Section 5, this translates into better convergence performance of JRT.

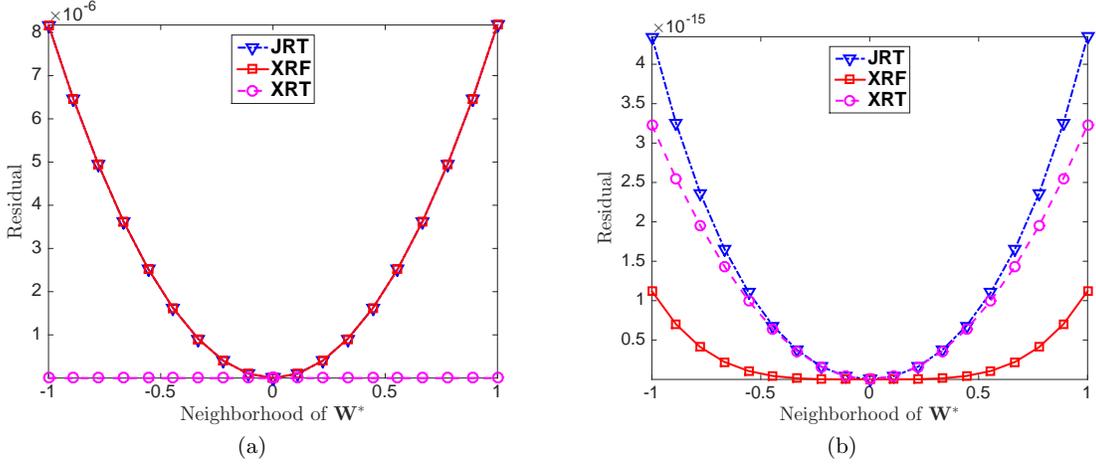


Figure 3: Residual objectives in a neighborhood of  $\mathbf{W}^*$  for the overdetermined system in Example 3.1: Local behavior of different objectives in (a) the null space of the XRT system and (b) the null space of the XRF system.

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 is possible with only one of these modalities. Furthermore, hardware technology has dramatically improved, so that simultaneous detection of correlative multimodality data is now experimentally possible [7, 43].

**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 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 (2.4) requires about  $|\mathcal{V}|^2|\mathcal{E}|^2|\mathcal{I}|n_d$  flops; accordingly, its elementwise derivative (A.1) 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 (2.4) and (A.1). Based on these strategies, Figure 4 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

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

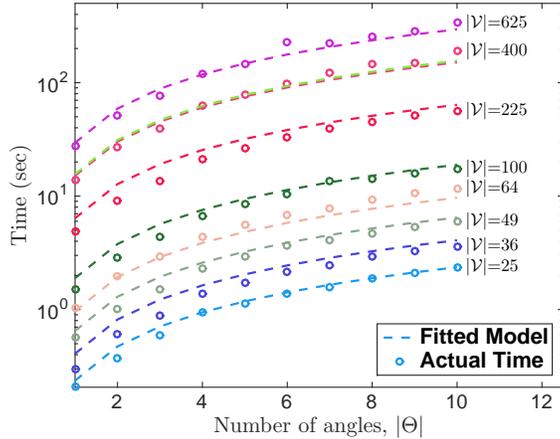


Figure 4: Time required for a single function and gradient evaluation ( $\phi^R, \nabla\phi^R$ ) for different input sizes (number of angular projections and voxels): Actual computational time and the model in Equation (4.1).

This shows that our approach is more feasible for larger-scale problems than the elementwise complexity analysis suggests. In Appendix B we examine the performance of our calculation of the gradient  $\nabla\phi^R$  relative to an automatic differentiation-based calculation.

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 with preconditioned projected conjugate gradient (PPCG) providing the search direction [30].

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; and
- a number of PPCG iterations, with cost per PPCG iteration given by: 1 (function, gradient) evaluation, 1 diagonal matrix-vector product, 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, 5 diagonal matrix-vector products, 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  $|\mathcal{V}||\mathcal{E}|(1 + 5 + 20 + 25 + 1) = 52|\mathcal{V}||\mathcal{E}|$  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.

The primary goal of our tests with the first sample is to measure the performance of

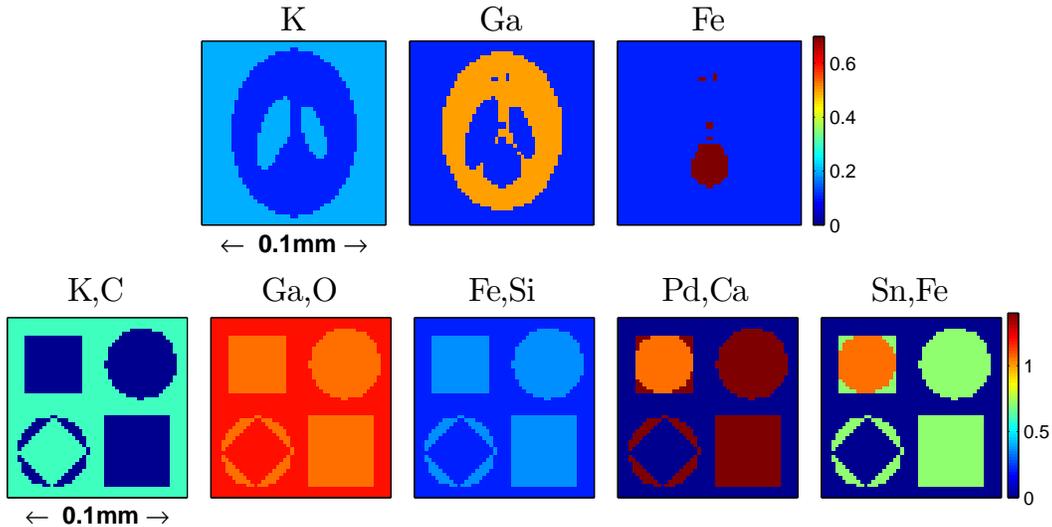


Figure 5: Chemical composition of the (*top*) Shepp Logan phantom and (*bottom*) phase  $50 \times 50$  synthetic samples; color indicates the mass of the elements present.

JRT with respect to different problem sizes and levels of noise in the measurement data. By testing the performance of different approaches for different size problems, we also show that the performance of our joint inversion method is independent of the mathematical numerical resolution; furthermore, it allows us to approximate the convergence factor of the underlying optimization solver.

For the second sample, we test the independence of JRT’s performance with respect to the given sample, and we demonstrate the benefit of our XRF forward model in correcting the attenuation coefficient by including the self-absorption effect (modeled by Equation (2.2)).

**5.1. Synthetic Samples.** We present results for two synthetic samples: a “phantom” sample and a “phase” sample. The chemical compositions of both samples are shown in Figure 5. 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 adapted from [15, 26] and is typically used to explore the self-absorption effect in XRF imaging. First, we demonstrate the experiment on the phase sample containing additionally palladium (Pd) and tin (Sn). Then, we demonstrate the joint inversion using a phase sample comprising lighter elements including carbon (C), oxygen (O), silicon (Si), calcium (Ca), and iron (Fe).

We denote these samples by  $\mathbf{W}^*$  and run each through our forward simulation to obtain XRF and XRT output. For all forward simulations in these tests, the simulated incident beam is monochromatic, with 20 keV energy and  $0.2 \times 0.2 \mu\text{m}$  transverse size. A 2.4 mm-diameter XRT detector is placed at  $90^\circ$  to the incident beam at a distance 1.6 cm from the sample. An XRF detector is placed parallel to the incident beam at a distance of 1.6 cm from the sample.

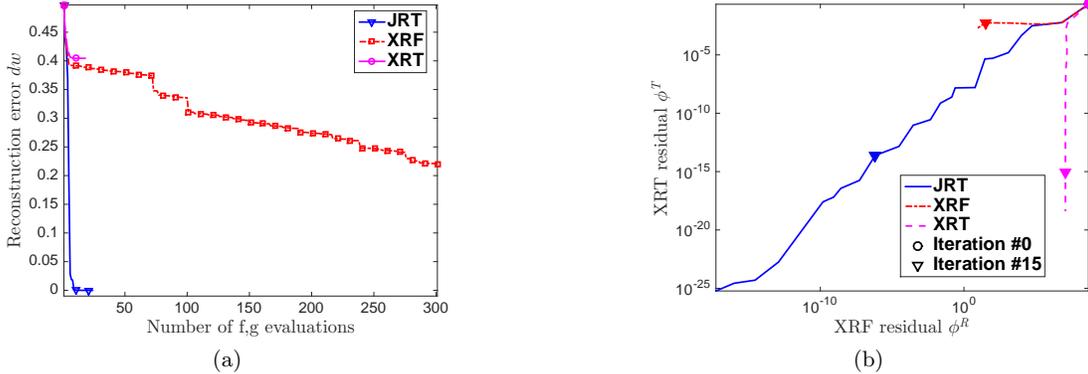


Figure 6: Optimization starting from  $\mathbf{W}_{\text{bad}}^0$  for the  $3 \times 3$  phantom sample: (a) Reconstruction error with respect to the number of (function, gradient) evaluations; (b) XRF and XRT residuals for the three reconstruction approaches.

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.

**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”

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

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

The second metric that we use is based on the forward residuals. We examine both the XRF residual  $\phi^R$  and the XRT residual  $\phi^T$ . Since the three approaches we consider are based on minimizing the residuals  $\phi^R$ ,  $\phi^T$ , and  $\phi^R + \beta\phi^T$ , one might expect that each approach would perform best for its respective residual. In a noise-free setting, we have that, by construction,  $\phi^R(\mathbf{W}^*) = \phi^T(\mathbf{W}^*) = 0$ , 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., only XRF or XRT) with JRT for problems of different resolution. To show the insensitivity of JRT to the starting point, we consider both  $\mathbf{W}_{\text{good}}^0$ , generated by adding to  $\mathbf{W}^*$  uniform noise with support  $[0, 0.1]^{|\mathcal{V}| \times |\mathcal{E}|}$ , and  $\mathbf{W}_{\text{bad}}^0$ , generated uniformly in  $[0, 0.1]^{|\mathcal{V}| \times |\mathcal{E}|}$  (and therefore independently of  $\mathbf{W}^*$ ).

Given  $\mathbf{W}_{\text{bad}}^0$ , Figure 6 illustrates the performance of different algorithms in reconstructing a  $3 \times 3$  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

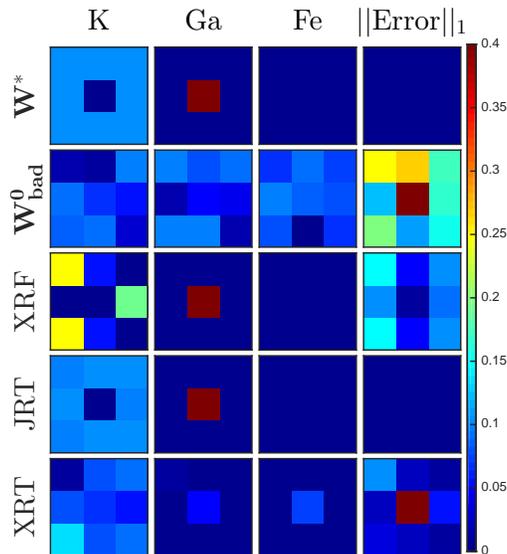


Figure 7: Reconstruction results starting from  $\mathbf{W}_{\text{bad}}^0$  for the  $3 \times 3$  phantom sample.

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, the solution manifold of XRF does not agree with that of XRT in general; that is, minimizing  $\phi^R$  and  $\phi^T$  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, not only does joint inversion converge much faster than the two single-modality inversions, but the local minimum found also achieves a much better reconstruction result in terms of both the XRT and XRF residual objectives. Therefore, JRT improves not only the XRF reconstruction but also the XRT reconstruction. 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 XRF tomography, for the remainder of our experiments we report only the reconstruction qualities of JRT and XRF, the approach mostly used in practice.

We now examine the performance of JRT on phantom in the case of limited angular scans (i.e.,  $|\Theta| = 2$ ). As we increase the size of sample resolution, the performance of JRT compared with XRF is measured, respectively, in terms of convergence factor, reconstruction error/residual, and time elapsed. For  $k \geq 1$ , the convergence factor (given the same stopping criteria) is

$$C_k = \left( \frac{\phi(\mathbf{W}^k) - \phi(\mathbf{W}^*)}{\phi(\mathbf{W}^0) - \phi(\mathbf{W}^*)} \right)^{\frac{1}{k}},$$

where, given a synthetic solution  $\mathbf{W}^*$ ,  $\mathbf{W}^k$  is the corresponding TN-reconstructed solution at

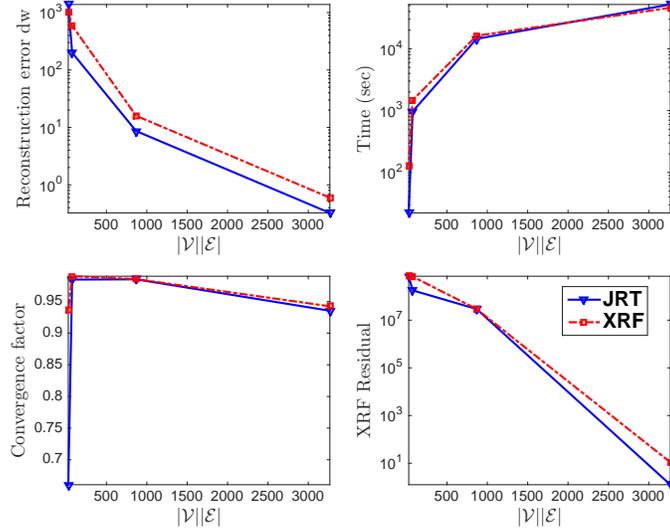


Figure 8: Performance of JRT and XRF starting from  $\mathbf{W}_{\text{good}}^0$  on the phantom sample (with  $|\mathcal{E}| = 3$  elements and  $|\Theta| = 2$  angular projections) as a function of the resolution size.

the  $k$ th iteration. Figure 8 assembles the corresponding results. As we increase the problem resolution (i.e., increase the number of voxels  $|\mathcal{V}|$  following the pattern  $[3 \times 3, 5 \times 5, 9 \times 9, 17 \times 17, 33 \times 33]$ ), the convergence factor and the XRF objective residual are always better with JRT than with XRF. We thus observe that, in the case of limited data, JRT is obtaining better reconstruction results than is XRF and doing so with computational time that is comparable to XRF.

Next, starting from  $\mathbf{W}_{\text{good}}^0$ , we increase the resolution of the phantom sample to  $20 \times 20$  pixels with  $|\Theta| = 10$  angular scans and consider data contaminated by adding 0.1% standard Gaussian noise to the simulated measurements. Figure 9 demonstrates that JRT converges faster than single XRF and XRT, in addition to providing an improved residual and error. The corresponding reconstructed elemental maps are given in Figure 10. As with the other, noise-free tests, we see that, with roughly the same computational cost, JRT consistently returns results with better accuracy than does XRF.

**5.4. Phase Sample Results.** Now we examine the performance of JRT on a sample (see Figure 5) typically used to explore the self-absorption effect in XRF imaging. Tests on this sample further illustrate the performance of JRT and highlight the advantage of our XRF forward model by including the self-absorption effect approximated by Equation (2.2). In cases where no self-absorption correction is computed in the inversion, we use (2.3) in problem (3.1).

First, analogous to the previous experiments, we compare the performance of JRT with XRF- and XRT-only reconstruction on a  $20 \times 20$  phase sample with 20 translation and 12 angular scans. Again, JRT bests XRF and XRT in reducing both the objective function residuals and reconstruction error as shown in Figure 11. The resulting elemental maps are

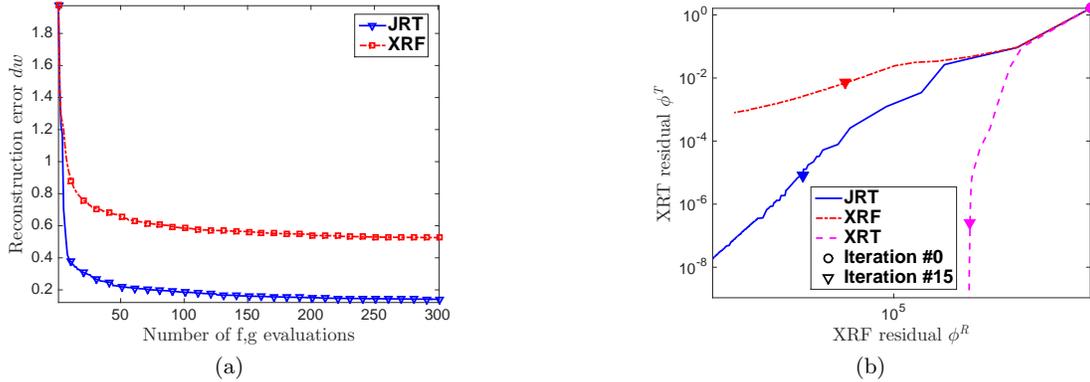


Figure 9: Optimization starting from  $\mathbf{W}_{\text{good}}^0$  for the  $20 \times 20$  phantom sample with noisy data: (a) Reduction of reconstruction error with respect to the number of (function, gradient) evaluations; (b) Trajectories of the XRF and XRT residuals for the three reconstruction approaches.

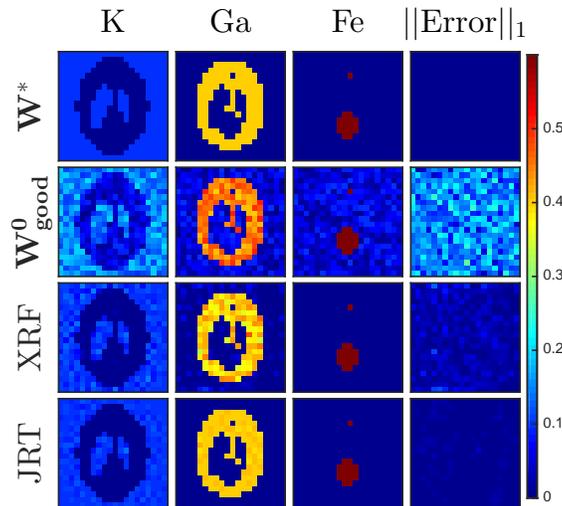


Figure 10: Reconstruction results starting from  $\mathbf{W}_{\text{good}}^0$  for the  $20 \times 20$  phantom sample with noisy data.

shown in Figure 12.

In Figure 12, we can see that with the self-absorption correction, the reconstructed JRT result is much more accurate. In particular, without the self-absorption correction, the mass in the inner region of the shapes is underestimated because of the self-absorption effect.

We also experiment on material containing the relatively light elements carbon and oxygen. For each element, Figures 13 and 15 plot the cumulative distribution of the reconstruction

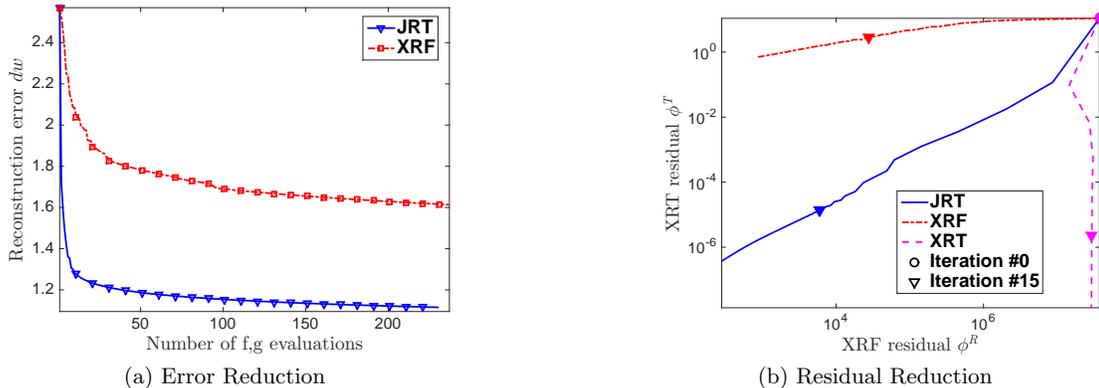


Figure 11: Optimization starting from  $\mathbf{W}_{\text{bad}}^0$  for the  $20 \times 20$  phase sample: (a) Reduction of reconstruction error with respect to the number of (function, gradient) evaluations; (b) Trajectories of the value  $\phi^R$  with respect to the value  $\phi^T$  for each iteration.

error across all voxels for discretization sizes  $20 \times 20$  and  $50 \times 50$ , respectively. For example, for a given error shown on the x-axis, the corresponding y-value is the fraction of the voxels whose reconstruction error is strictly smaller than this given error. Therefore, the uniformly distributed initial error in  $[0, 0.1]^{|\mathcal{V}| \times |\mathcal{E}|}$  is shown as a diagonal line. We see that, for the relatively high fluorescing elements Si, Ca, and Fe, JRT dramatically reduces the error since the majority of the points is scrunched into a small fraction of the plots on the far left. The light elements C and O are not recovered as well, which is not surprising. A limitation of absorption-contrast imaging is that XRT usually cannot be used for observing low X-ray absorption materials (e.g., thin organic materials) [28]. Hard X-ray fluorescence is typically used to quantify heavy elements with atomic numbers greater than 14 [6], and thus we also cannot obtain strong enough characteristic emission data of light elements from XRF. Therefore, we do not expect any of the tested approaches to successfully recover C and O in this test; instead, we view the presence of C and O as adding systematic noise to the reconstruction, and prove again the benefit of JRT to recover the elements of interest in the presence of lighter elements.

Second, we test our XRF forward model with and without the self-absorption correction on the  $20 \times 20$  phase sample. The corresponding result is shown in Figure 14. In particular, without the self-absorption correction, the highly fluorescing elements Ca and Fe are underestimated in the interior region of the sample.

**6. Summary and Discussion.** We briefly summarize our results and discuss future extensions of our work.

**6.1. Summary.** This paper addresses a joint inversion framework to solve tomography problems by using data from multiple imaging modalities in order to improve the ill-conditioned nature of such problems. We model the XRF and XRT phenomena separately as nonlinear equations in terms of the chemical element composition of a given sample. We also formulate an optimization-based joint inversion problem that utilizes data from XRF and XRT tomog-

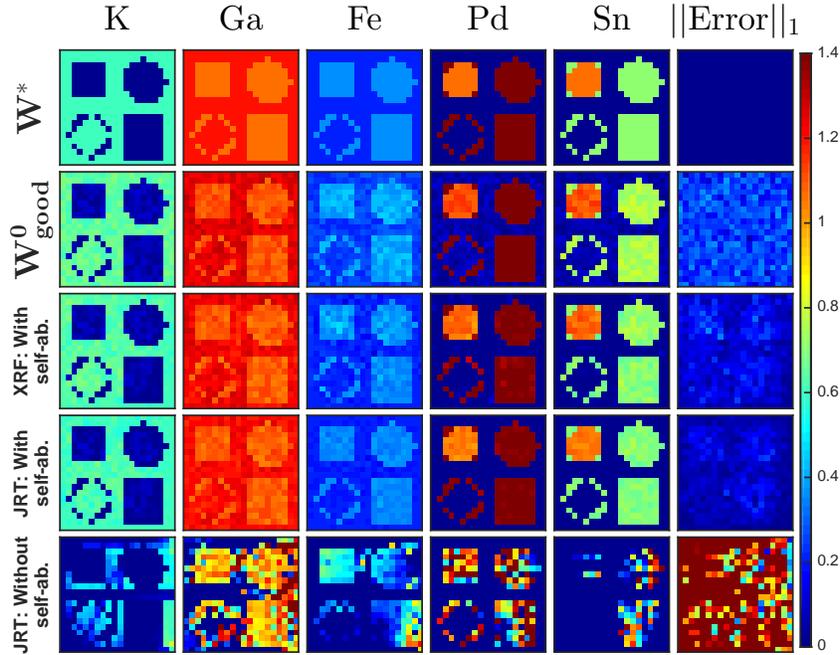


Figure 12: Reconstruction results starting from  $\mathbf{W}_{\text{good}}^0$  for the  $20 \times 20$  phase sample (with and without self-absorption correction for JRT).

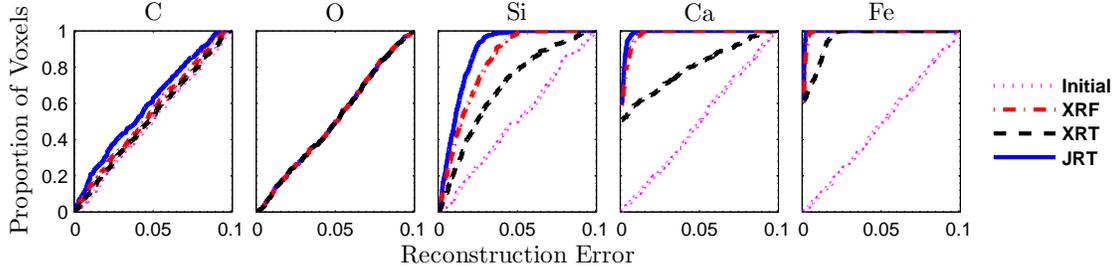


Figure 13: Reconstruction error starting from  $\mathbf{W}_{\text{good}}^0$  for the  $20 \times 20$  phase sample with light elements. Each line corresponds to the cumulative distribution (across all 400 voxels) of the error obtained by a particular algorithm.

raphy modalities. The resulting optimization problem requires us to minimize the errors between the measured data and our forward models, which we achieve with the truncated-Newton method.

The numerical results presented in the paper show that, when limited scans are available, significant improvements are achieved by performing joint inversion. Given a properly chosen weighting factor  $\beta$ , these improvements are in terms of both convergence speed and accuracy

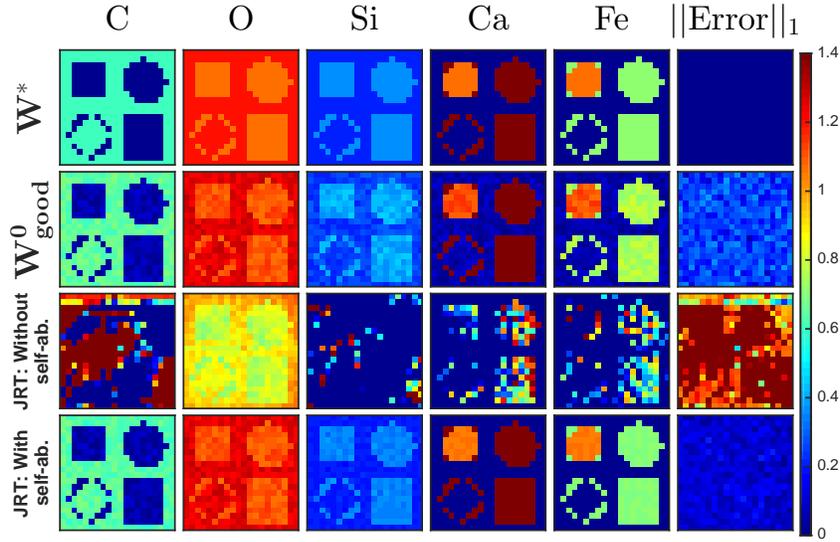


Figure 14: JRT reconstruction result starting from  $\mathbf{W}_{\text{good}}^0$  for the  $20 \times 20$  phase sample with light elements: with and without self-absorption correction.

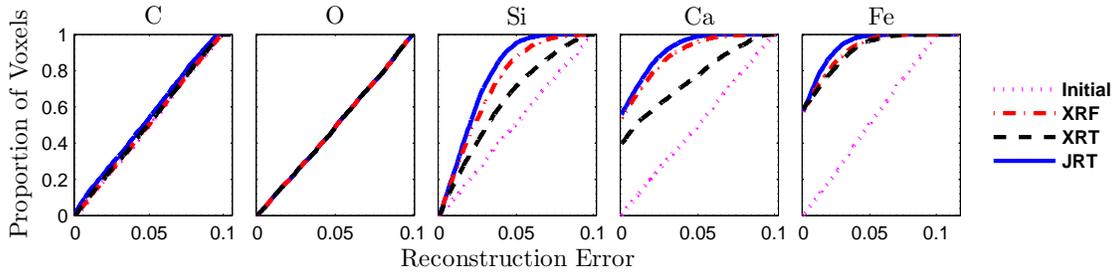


Figure 15: Reconstruction error starting from  $\mathbf{W}_{\text{good}}^0$  for the  $50 \times 50$  phase sample with light elements. Each line corresponds to the cumulative distribution (across all 2,500 voxels) of the error obtained by a particular algorithm.

of the reconstruction.

**6.2. Discussion.** The joint inversion method we present in this work can be naturally extended to 3D by approximating the emission and attenuation of X-ray fluorescence isotropically. Experimental enhancements that facilitate such 3D extensions include collimating the XRF detector so that it will count only the X-ray fluorescence emitted from a particular slice [48]. Another approach is to adopt simpler forward models, such as the model proposed by La Rivière [24].

A limitation of our work is the expensive computational complexity associated with the self-absorption correction. To obtain run time improvements beyond what a compiled code

could provide, we would need to consider a parallel implementation. Such an implementation could benefit from both the parallelized version of TN and parallelizing the individual beamlet calculations, for example, by using so-called *in-slice parallelization* [2]. A multilevel framework [42] could also provide computational acceleration.

In practice, we could also extend the applicability of our forward model. In this work, we ignore the modification of measured intensities caused by the scattering of both the incident and the fluorescence X-ray radiations inside the specimen. One solution to compensate for this effect is a calibration procedure [36]. Furthermore, to expand the usage of joint inversion to soft tissue with light elements (i.e., atomic numbers smaller than 10), we can employ phase contrast imaging [12, 28] to increase the sensitivity of the signal to the light elements.

Overall, our framework is general since it does not depend on specific experiments or samples, but instead exploits complementary information from different imaging modalities. Joint inversion is more appealing theoretically, and it can incorporate modalities, such as ptychographic data [43], beyond those considered here. As long as the different modalities of data are acquired from the same sample and complement each other, joint inversion has the potential to provide better reconstruction results.

## Appendices

**A. Derivatives of the XRT and XRF Forward Models.** For  $e \in \mathcal{E}$  and  $v \in \mathcal{V}$ , the first-order derivative of (2.1) with respect to  $W_{v,e}$  is

$$\frac{\partial}{\partial W_{v,e}} F_{\theta,\tau}^T(\mathbf{W}) = -I_0 \exp \left\{ - \sum_{v',e'} L_{v'}^{\theta,\tau} \mu_{e'}^E W_{v',e'} \right\} L_v^{\theta,\tau} \mu_e^E.$$

Similarly, one can show that the first-order derivative of (2.4) with respect to  $W_{v,e}$  is given by

$$\begin{aligned} & \text{(A.1)} \\ & \frac{\partial}{\partial W_{v,e}} \mathbf{F}_{\theta,\tau}^R(\mathbf{W}) \\ &= \sum_{e'} \left( \sum_{v'} L_{v'}^{\theta,\tau} W_{v',e'} \left( \frac{\partial A_{v'}^{E,\theta,\tau}}{\partial W_{v,e}} F_{v',e'}^{\theta,\tau} + A_{v'}^{E,\theta,\tau} \frac{\partial F_{v',e'}^{\theta,\tau}}{\partial W_{v,e}} \right) \right) \mathbf{M}_{e'} + L_v^{\theta,\tau} A_v^{E,\theta,\tau} F_{v,e}^{\theta,\tau} \mathbf{M}_e \\ &= - \sum_{e'} \left( \sum_{v'} L_{v'}^{\theta,\tau} W_{v',e'} A_{v'}^{E,\theta,\tau} F_{v',e'}^{\theta,\tau} \left( \mu_e^E L_v^{\theta,\tau} \mathbb{I}_{v \in \mathcal{U}_{v'}^{\theta,\tau}} \right) \right) \mathbf{M}_{e'} \\ & \quad - \sum_{\substack{e' \\ v'}} \left( \sum_{v'} L_{v'}^{\theta,\tau} W_{v',e'} A_{v'}^{E,\theta,\tau} \left( \frac{1}{n_d} \sum_{d=1}^{n_d} \exp \left\{ - \sum_{v''} \sum_{e''} W_{v'',e''} \mu_{e''}^E P_{v',v'',d}^{\theta,\tau} \right\} \mu_{e'}^E P_{v',v,d}^{\theta,\tau} \right) \right) \mathbf{M}_{e'} \\ & \quad + L_v^{\theta,\tau} A_v^{E,\theta,\tau} F_{v,e}^{\theta,\tau} \mathbf{M}_e. \end{aligned}$$

**B. Efficiency of the Different Derivative Calculations.** We use the `AdiMat` tool [3], which is a software package for automatic differentiation (AD) of programs written in MATLAB, to perform 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

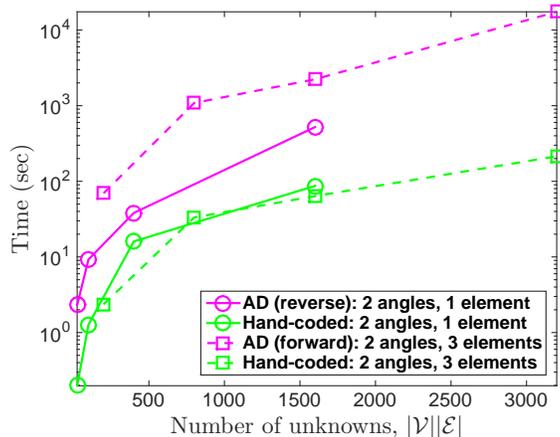


Figure 16: Time required for a single function and gradient evaluation ( $\phi^R, \nabla\phi^R$ ) for different input sizes: comparison of computational costs using our hand-coded, tensor-based derivative and the AD-based derivatives.

mode. Figure 16 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 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.

**Acknowledgments.** This material is based upon work supported by the U.S. Department of Energy, Office of Science, under Contract DE-AC02-06CH11357. We are grateful to Chris Jacobsen for introducing us to this problem and for valuable discussions in the preparation of this paper. We are grateful to three anonymous referees whose feedback improved the presentation.

## REFERENCES

- [1] M. BERTERO AND P. BOCCACCI, *Introduction to inverse problems in imaging*, CRC Press, 1998.
- [2] T. BICER, D. GURSOY, R. KETTIMUTHU, F. DE CARLO, G. AGRAWAL, AND I. T. FOSTER, *Rapid tomographic image reconstruction via large-scale parallelization*, in *Euro-Par 2015: Parallel Processing*, Springer, 2015, pp. 289–302.
- [3] C. H. BISCHOF, H. M. BÜCKER, B. LANG, A. RASCH, AND A. VEHRSCCHILD, *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.
- [4] R. CESAREO AND S. 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.
- [5] M. E. DAVISON, *The ill-conditioned nature of the limited angle tomography problem*, *SIAM Journal on Applied Mathematics*, 43 (1983), pp. 428–448.

- [6] M. D. DE JONGE AND S. VOGT, *Hard X-ray fluorescence tomography: An emerging tool for structural visualization*, Current Opinion in Structural Biology, 20 (2010), pp. 606–614.
- [7] J. DENG, D. J. VINE, S. CHEN, Y. S. G. NASHED, Q. JIN, N. W. PHILLIPS, T. PETERKA, R. ROSS, S. VOGT, AND C. J. JACOBSEN, *Simultaneous cryo X-ray ptychographic and fluorescence microscopy of green algae*, Proceedings of the National Academy of Sciences, 112 (2015), pp. 2314–2319.
- [8] M. J. EHRHARDT, K. THIELEMANS, L. PIZARRO, D. ATKINSON, S. OURSELIN, B. F. HUTTON, AND S. R. ARRIDGE, *Joint reconstruction of PET-MRI by exploiting structural similarity*, Inverse Problems, 31 (2015), p. 015001.
- [9] H. W. ENGL AND P. 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.
- [10] J. A. FESSLER, *Penalized weighted least-squares image reconstruction for positron emission tomography*, IEEE Transactions on Medical Imaging, 13 (1994), pp. 290–300.
- [11] S. FINSTERLE AND M. B. KOWALSKY, *Joint hydrological-geophysical inversion for soil structure identification*, Vadose Zone Journal, 7 (2008), pp. 287–293.
- [12] R. FITZGERALD, *Phase-sensitive X-ray imaging*, Physics Today, 53 (2007), pp. 23–26.
- [13] L. A. GALLARDO AND M. A. MEJU, *Joint two-dimensional DC resistivity and seismic travel time inversion with cross-gradients constraints*, Journal of Geophysical Research: Solid Earth, 109 (2004).
- [14] ———, *Structure-coupled multiphysics imaging in geophysical sciences*, Reviews of Geophysics, 49 (2011).
- [15] B. GOLOSIO, A. SIMIONOVICI, A. SOMOGYI, L. LEMELLE, M. CHUKALINA, AND A. BRUNETTI, *Internal elemental microanalysis combining X-ray fluorescence, Compton and transmission tomography*, Journal of Applied Physics, 94 (2003), pp. 145–156.
- [16] D. GÜRSOY, T. BIÇER, A. LANZIROTTI, M. G. NEWVILLE, AND F. DE CARLO, *Hyperspectral image reconstruction for X-ray fluorescence tomography*, Optics Express, 23 (2015), pp. 9014–9023.
- [17] E. HABER AND M. H. GAZIT, *Model fusion and joint inversion*, Surveys in Geophysics, 34 (2013), pp. 675–695.
- [18] E. HABER AND D. OLDENBURG, *Joint inversion: A structural approach*, Inverse Problems, 13 (1997), p. 63.
- [19] P. C. HANSEN AND D. P. 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.
- [20] J. P. HOGAN, R. A. GONSALVES, AND A. S. KRIEGER, *Fluorescent computer tomography: A model for correction of X-ray absorption*, IEEE Transactions on Nuclear Science, 38 (1991), pp. 1721–1727.
- [21] W. HU, A. ABUBAKAR, AND T. HABASHY, *Joint electromagnetic and seismic inversion using structural constraints*, Geophysics, 74 (2009), pp. R99–R109.
- [22] A. C. KAK AND M. SLANEY, *Principles of Computerized Tomographic Imaging*, IEEE Press, 1988.
- [23] M. B. KOWALSKY, S. FINSTERLE, J. PETERSON, S. HUBBARD, Y. RUBIN, E. MAJER, A. WARD, AND G. GEE, *Estimation of field-scale soil hydraulic and dielectric parameters through joint inversion of GPR and hydrological data*, Water Resources Research, 41 (2005).
- [24] P. LA RIVIÈRE, *Approximate analytic reconstruction in X-ray fluorescence computed tomography*, Physics in Medicine and Biology, 49 (2004), p. 2391.
- [25] P. LA RIVIÈRE AND D. 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.
- [26] P. LA RIVIÈRE AND P. A. VARGAS, *Monotonic penalized-likelihood image reconstruction for X-ray fluorescence computed tomography*, IEEE Transactions on Medical Imaging, 25 (2006), pp. 1117–1129.
- [27] E. X. MIQUELES AND A. R. DE PIERRO, *Iterative reconstruction in X-ray fluorescence tomography based on Radon inversion*, IEEE Transactions on Medical Imaging, 30 (2011), pp. 438–450.
- [28] T. MUKAIDE, A. IIDA, M. WATANABE, K. TAKADA, AND T. NOMA, *Quantitative mass density image reconstructed from the complex X-ray refractive index*, PloS One, 10 (2015), p. e0131401.
- [29] A. MUNOZ-BARRUTIA, C. PARDO-MARTIN, T. PENGO, AND C. ORTIZ-DE SOLORZANO, *Sparse algebraic reconstruction for fluorescence mediated tomography*, International Society for Optics and Photonics, 2009, pp. 744604–744604.
- [30] S. G. NASH, *A survey of truncated-Newton methods*, Journal of Computational and Applied Mathematics, 124 (2000), pp. 45–59.

- [31] A. NEUMAIER, *Solving ill-conditioned and singular linear systems: A tutorial on regularization*, SIAM Review, 40 (1998), pp. 636–666.
- [32] O. OCHOA, A. VELASCO, AND C. SERVIN, *Towards model fusion in geophysics: How to estimate accuracy of different models*, Departmental Technical Reports (CS), (2013).
- [33] T. PAUNESCU, S. VOGT, J. MASER, B. LAI, AND G. WOLOSCHAK, *X-ray fluorescence microprobe imaging in biology and medicine*, Journal of Cellular Biochemistry, 99 (2006), pp. 1489–1502.
- [34] J. RADON, *On the determination of functions from their integral values along certain manifolds*, IEEE Transactions on Medical Imaging, 5 (1986), pp. 170–176.
- [35] D. RENDELL AND D. J. MOWTHORPE, *Fluorescence and phosphorescence spectroscopy*, Published on behalf of ACOL, London by Wiley, 1987.
- [36] R. M. ROUSSEAU, *Fundamental algorithm between concentration and intensity in XRF analysis 1–theory*, X-Ray Spectrometry, 13 (1984), pp. 115–120.
- [37] R. M. ROUSSEAU AND J. A. BOIVIN, *The fundamental algorithm: A natural extension of the Sherman equation, part I: Theory*, Rigaku J, 15 (1998), pp. 13–5.
- [38] G.-F. RUST AND J. WEIGELT, *X-ray fluorescent computer tomography with synchrotron radiation*, IEEE Transactions on Nuclear Science, 45 (1998), pp. 75–88.
- [39] T. SCHOONJANS, A. BRUNETTI, B. GOLOSIO, M. S. DEL RIO, V. A. SOLÉ, C. FERRERO, AND L. VINCZE, *The xraylib library for X-ray–matter interactions. Recent developments*, Spectrochimica Acta Part B: Atomic Spectroscopy, 66 (2011), pp. 776–784.
- [40] J. SHERMAN, *The theoretical derivation of fluorescent X-ray intensities from mixtures*, Spectrochimica Acta, 7 (1955), pp. 283–306.
- [41] A. THOMPSON ET AL., eds., *X-ray Data Booklet*, Lawrence Berkeley National Laboratory, University of California, 2001.
- [42] U. TROTTEBERG, C. W. OOSTERLEE, AND A. SCHULLER, *Multigrid*, Academic Press, 2000.
- [43] 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.
- [44] C. R. VOGEL, *Computational methods for inverse problems*, vol. 23, SIAM, 2002.
- [45] Y. WANG, J. YANG, W. YIN, AND Y. ZHANG, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM Journal on Imaging Sciences, 1 (2008), pp. 248–272.
- [46] J. P. WARD, M. LEE, J. C. YE, AND M. UNSER, *Interior tomography using 1D generalized total variation. part I: Mathematical foundation*, SIAM Journal on Imaging Sciences, 8 (2015), pp. 226–247.
- [47] Q. YANG, B. DENG, W. LV, F. SHEN, R. CHEN, Y. WANG, G. DU, F. YAN, T. XIAO, AND H. 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.
- [48] T. YUASA, M. AKIBA, T. TAKEDA, M. KAZAMA, A. HOSHINO, Y. WATANABE, K. HYODO, F. A. DILMANIAN, T. AKATSUKA, AND Y. 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.
- [49] T. YUASA, T. TAKEDA, F. A. DILMANIAN, M. AKIBA, K. HYODO, T. AKATSUKA, AND Y. 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.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne”). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up, nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.