A SCALED GRADIENT METHOD FOR DIGITAL TOMOGRAPHIC IMAGE RECONSTRUCTION

JIANJUN ZHANG
Department of Mathematics, Shanghai University
Shanghai 200444, China

YUNYI HU AND JAMES G. NAGY
Department of Mathematics and Computer Science, Emory University
Atlanta, GA 30322, USA

(Communicated by Hao-Min Zhou)

Abstract. Digital tomographic image reconstruction uses multiple x-ray projections obtained along a range of different incident angles to reconstruct a 3D representation of an object. For example, computed tomography (CT) generally refers to the situation when a full set of angles are used (e.g., 360 degrees) while tomosynthesis refers to the case when only a limited (e.g., 30 degrees) angular range is used. In either case, most existing reconstruction algorithms assume that the x-ray source is monoenergetic. This results in a simplified linear forward model, which is easy to solve but can result in artifacts in the reconstructed images. It has been shown that these artifacts can be reduced by using a more accurate polyenergetic assumption for the x-ray source, but the polyenergetic model requires solving a large-scale nonlinear inverse problem. In addition to reducing artifacts, a full polyenergetic model can be used to extract additional information about the materials of the object; that is, to provide a mechanism for quantitative imaging. In this paper, we develop an approach to solve the nonlinear image reconstruction problem by incorporating total variation (TV) regularization. The corresponding optimization problem is then solved by using a scaled gradient descent method. The proposed algorithm is based on KKT conditions and Nesterov’s acceleration strategy. Experimental results on reconstructed polyenergetic image data illustrate the effectiveness of this proposed approach.

1. Introduction. Since its invention by Hounsfield in 1972, x-ray computerized tomography (CT) has been a modern and widely used technique to visualize the inner structure of various bodies, and it has become an important tool in the medical diagnosis of many different diseases. Typically, in x-ray CT imaging, an object is illuminated with a series of x-rays from multiple angles, and the change in the intensity of x-rays along a series of linear paths is measured. Then a relationship between the intensity change and material attenuation coefficients is determined by Beer’s law [17]. Finally, a specialized algorithm is used to reconstruct material attenuation coefficients and accordingly reconstruct the image in question. We will

2010 Mathematics Subject Classification. Primary: 65F22, 65F10; Secondary: 65K10.
Key words and phrases. Tomosynthesis, gradient descent, ill-posed problems, beam hardening artifacts, total variation.

The first author is supported by grant no. 15ZR1416300 from the Shanghai Municipal Natural Science Foundation, the third author is supported by grant no. DMS-1522760 from the US National Science Foundation.

* Corresponding author: James G. Nagy.
use the term full CT to refer to the situation when the x-ray source and detector make a complete rotation around the object, to distinguish from the term tomosynthesis, which refers to limited angle CT, where the rotation angle is 60 degrees or less. While the resolution in the direction orthogonal to the x-ray source of a tomosynthesis reconstructed image is not as good as it would be for full CT, an advantage of tomosynthesis is that it requires significantly less radiation dose to the patient compared to full CT [21, 29, 30].

In a real imaging system (either full CT or tomosynthesis), the x-ray field emitted from the source is polyenergetic, and an accurate polyenergetic model for image reconstruction requires solving a large-scale nonlinear inverse problem to compute the attenuation coefficients. This is mathematically and computationally difficult. Thus, traditional reconstruction techniques use a simplified model that assumes the x-ray source is monoenergetic, which only requires solving a linear inverse problem. Although it is much easier to solve the simplified linear model, ignoring the energy dependence in the mathematical model leads to so-called beam hardening artifacts [4, 9, 14].

In recent years, several approaches to address the polyenergetic tomographic image reconstruction problem have been proposed; for full CT see [3, 15, 16, 26, 27], and for tomosynthesis see [10, 13]. In addition to reducing beam hardening artifacts, the additional information provided by a polyenergetic source model has the potential to allow for quantitative imaging; specifically, to identify material composition of objects being imaged. In 1976, Alvarez and Macovski [1] and Macovski et al. [24] proposed using dual energy CT, where two full sets of CT projection data are obtained and each uses distinct (assumed monoenergetic) x-ray energies. With the monoenergetic assumption, each set of projection data corresponds to a distinct linear inverse problem. This idea has been extended to whole-body CT, including generalizations of using three or more distinct energies; see, for example [2, 20, 31, 32]. This pioneering work on multispectral CT is an important contribution to diagnostic medical imaging. However, it is important to emphasize that these schemes still assume that the x-ray source is monoenergetic, which is not physically accurate. Moreover, because they require collecting two or more full sets of data (i.e., two or more complete scans), there is a significant increase in radiation dose to the patient.

In this paper, we consider both full CT and tomosynthesis problems and we use the nonlinear polyenergetic model of the x-ray source. The potential of conducting quantitative imaging in this situation (for breast imaging) was demonstrated in [10, 13]. The numerical approach used in these previous papers, which applied gradient descent to an un-regularized Poisson negative log-likelihood objective function with no explicit constraints, illustrated clearly that beam hardening artifacts could be significantly reduced with the polyenergetic model. The purpose of this paper is to investigate the use of bound constraints and more sophisticated regularization methods that provide better quantitative imaging capabilities. Specifically, we use total variation (TV) regularization, and solve the corresponding optimization problem with a scaled gradient descent method. The proposed algorithm is based on KKT conditions and Nesterov’s acceleration strategy.

The remainder of the paper is organized as follows. Section 2 reviews the polyenergetic model developed in [10]. Our proposed method is described in Section 3. Numerical results are presented in Section 4, and some conclusions are given in Section 5.
2. Mathematical model. In this section, we review the polyenergetic x-ray computerized tomography image formation process with material parameterization, and the corresponding minimization problem to be considered.

2.1. Polyenergetic x-ray transmission with material parameterization. In this subsection, we describe the polyenergetic x-ray computerized tomography image formation process with multimaterial decomposition. To simplify some of the notation, we will show a 2-D image formation process, but the mathematical model extends naturally to 3-D imaging.

Let \( b_d(\theta) \) be the quantity measured at the \( d \)th bin of a digital x-ray detector, obtained when the x-ray source is at an incident angle \( \theta \). Using Beer’s law [17], these measured values are related to the object’s attenuation (that is, the fraction of the x-ray absorbed or scattered by the object) through the integral equation

\[
\frac{\partial}{\partial \theta} b_d(\theta) = \int_{\varepsilon} s(\varepsilon) e^{-\int_{\Gamma} \mu(\bar{x}(t),\varepsilon) dt} d\varepsilon + \eta_d(\theta), \quad d = 1, 2, \ldots, N_d, \quad \theta = 1, 2, \ldots, N_\theta,
\]

where

- \( \int_{\varepsilon} \) denotes integration along the ray \( \Gamma \);
- \( N_d \) is the number of bins in the digital x-ray detector;
- \( N_\theta \) is the number of projection images obtained as the x-ray source is moved to a new position, which is defined by an angle \( \theta \);
- \( \varepsilon \) represents the energies that are emitted by the source x-ray beam, which range from, for example, 10 keV to 28 keV;
- \( s(\varepsilon) \) is the energy fluence, which is a product of the x-ray energy with the number of incident photons at that energy;
- \( \mu(\bar{x},\varepsilon) \) is the linear attenuation coefficient, which depends on the energy of the x-ray beam and on the material in the object at the position \( \bar{x} \); lower energy will be attenuated more than higher energy, and denser materials such as bone or calcifications will attenuate more than soft tissue;
- \( \eta_d(\theta) \) represents additional contributions (noise) measured at the detector, which can include x-ray scatter and electronic noise.

Discretizing the integrals in (1) leads to the discrete image formation model

\[
\frac{\partial}{\partial \theta} b_d(\theta) = \sum_{i=1}^{N_\varepsilon} s_i \exp \left( -\sum_{k=1}^{K} \sum_{\ell=1}^{L} a_{d,k\ell}(\theta) \mu_{k\ell,i} \right) + \eta_d(\theta), \quad d = 1, 2, \ldots, N_d, \quad \theta = 1, 2, \ldots, N_\theta,
\]

where we use the notation

- \( N_p = K \times L \) is the number of pixels in the discretized 2-D object;
- \( N_\varepsilon \) is the number of discrete energy levels (in general \( N_\varepsilon \ll N_p \));
- \( s_i, i = 1, 2, \ldots, N_\varepsilon \) are discrete values of the energy fluence function, \( s(\varepsilon) \);
- \( a_{d,k\ell}(\theta) \) is the length of the path traveled by the x-ray that passes through pixel \((k, \ell)\), contributing to bin \( d \) of the detector.

We remark that this model easily extends to 3D discretized objects with 2D detectors.

The linear attenuation coefficients \( \mu_{k\ell,i} \), of the composite material making up the object (e.g., the breast), can be approximated as a linear combination of individual materials. That is, suppose that there are \( N_m \) distinct materials making up the
object, we have

\( \mu_{k\ell,i} \approx \sum_{m=1}^{N_m} w_{k\ell,m} c_{m,i} \),

where

- \( c_{m,i} \) are known as linear attenuation coefficients for the \( m \)th material in pixel \((k, \ell)\) at the \( i \)th discrete x-ray energy value; these coefficients, for a variety of specific materials, can be obtained from [7], using the chemical compositions of each material as determined by Hammerstein et al. [18];
- \( w_{k\ell,m} \) are unknown weight fractions (or percentages) of the \( m \)th material in the \((k, \ell)\)th pixel of the object.

We also assume that the weight fractions for each pixel should add to 1 (or percentages should add to 100), which means

\[ \sum_{m=1}^{N_m} w_{k\ell,m} = 1, \quad k = 1, 2, \ldots, K, \quad \ell = 1, 2, \ldots, L, \]

or, equivalently,

\[ w_{k\ell,1} = 1 - \sum_{m=2}^{N_m} w_{k\ell,m}. \]

Substituting relations (3) and (4) into (2), we obtain

\( b^{(\theta)} = \sum_{i=1}^{N_s} s_i \exp \left( - \sum_{k=1}^{K} \sum_{\ell=1}^{L} a_{d,k\ell}^{(\theta)} \left( c_{1,i} + \sum_{m=2}^{N_m} w_{k\ell,m} (c_{m,i} - c_{1,i}) \right) \right) \) + \( \eta^{(\theta)}_d \).

To help simplify some of the mathematical formulations and discussion in the rest of the paper, it is useful to rewrite equation (5) in matrix-vector form. To do this, we use the following notation:

- \( b^{(\theta)} \) is defined to be the vector of length \( N_d \) whose \( d \)th element is \( b_d^{(\theta)} \).
- \( A^{(\theta)} \) is defined to be the \( N_d \times N_p \) matrix whose \((d, k + (\ell - 1)L)\)th element is \( a_{d,k\ell}^{(\theta)} \).
- Let \( w_{k\ell,m+1} \) be denoted as \( x_{k\ell,m} \) for \( k = 1, 2, \ldots, K, \ell = 1, 2, \ldots, L \) and \( m = 1, 2, \ldots, N_m - 1 \), which represent the unknown weight fractions; and let \( x_m \) be the vector of length \( N_p \) containing the unknown weight fractions for the \((m + 1)\)th material \( w_{k\ell,m+1} \) (that is, the vector with entries \( w_{k\ell,m+1} \) or equivalently \( x_{k\ell,m} \)), and define the matrix \( X \) as

\[ X = [x_1 \mid x_2 \mid \cdots \mid x_{N_m-1}]. \]

- Let \( c_m \) be a vector of length \( N_c \) containing the known attenuation coefficients of the \( m \)th material at the various discrete energy levels (that is, the vector with entries \( c_{1,m} \)), and define the matrix \( \hat{C} \) as

\[ \hat{C} = [c_2 - c_1 \mid c_3 - c_1 \mid \cdots \mid c_{N_c} - c_1]. \]

If we now define \( 1 \) be the vector of length \( N_p \) with all elements equal one, then (5) can be written in matrix-vector form as

\[ b^{(\theta)} = \exp(-A^{(\theta)}(1e_1^T + X\hat{C}^T))s + \eta^{(\theta)}, \quad \theta = 1, 2, \ldots, N_\theta, \]

where the exponentiation operation is done element-wise on the matrix \(-A^{(\theta)}(1e_1^T + X\hat{C}^T)\).
Now let

\[
\begin{bmatrix}
  b^{(1)} \\
  b^{(2)} \\
  \vdots \\
  b^{(N)}
\end{bmatrix},
\begin{bmatrix}
  A^{(1)} \\
  A^{(2)} \\
  \vdots \\
  A^{(N)T}
\end{bmatrix}, \quad \text{and} \quad \begin{bmatrix}
  \eta^{(1)} \\
  \eta^{(2)} \\
  \vdots \\
  \eta^{(N)T}
\end{bmatrix}.
\]

Then a discrete model of (5) that comprises all projections can be written in matrix-vector form as

\[
b = \exp(-A(1c^T + X\hat{C}^T))s + \eta.
\]

2.2. Constrained optimization model. Let

\[
K(X) \equiv \exp(-A(1c^T + X\hat{C}^T)).
\]

Then our goal is solving large-scale inverse problems of the form

\[
b = K(X)s + \eta.
\]

In the Poisson statistical model [33], which is commonly used for medical imaging applications, it is assumed that the measured data is a realization of a Poisson random variable, and hence \(b_n \sim \text{Poisson}([K(X)s]_n + \eta_n), \ n = 1, 2, \ldots, N\), where \(N = N_pN_d\), and \(b_n, [K(X)s]_n, \) and \(\eta_n\) are, respectively, the \(n\)th elements of the vectors \(b, K(X)s\), and \(\eta\). The likelihood of observing \(b\) given \(X\) is described by the function

\[
p(b|X) = \prod_{n=1}^{N} \frac{([K(X)s]_n + \eta_n)^{b_n} \exp(-([K(X)s]_n + \eta_n))}{b_n!}.
\]

In the case of a well-posed problem, one can find an approximation of \(X\) by computing the maximum of this likelihood function, or equivalently, by computing the minimum of the negative log-likelihood.

However, in the case of severely ill-posed problems, such as tomosynthesis, it is typically recommended to incorporate regularization, or in the Bayesian framework, to include additional prior information on the unknown \(X\). In particular, suppose that the probability density of \(X\), which we denote as \(p(X)\), is known. The most frequently used priors are of the Gibbs type, which have the form

\[
p(X) = \frac{1}{Z} \exp(-\lambda\Omega(X)),
\]

where \(Z\) is a normalization constant, \(\lambda\) is a positive regularization parameter, and \(\Omega(X)\) is a functional, possibly convex. From Bayes’ formula we obtain the a posteriori probability density of \(X\) for a given value \(b\):

\[
p(X|b) = \frac{p(b|X)p(X)}{p(b)}.
\]

By maximizing the a posteriori estimate (9) of the unknown \(X\), or equivalently, by minimizing the negative log of the posteriori function (9), and notice that the elements of \(X\) lie in \([0, 1]\), (we ignore the constant terms associated with \(b_n!\) and \(p(b)\), and add \(b_n \ln(b_n) - b_n\) to the negative log of the posteriori function (9)), we obtain the following minimization problem

\[
\min_{0 \leq X \leq 1} \Phi(X) = \min_{0 \leq X \leq 1} \left(\mathcal{L}(X) + \lambda\Omega(X)\right),
\]
where
\begin{equation}
L(X) = \sum_{n=1}^{N} ((K(X)s)|n + \eta_n) - b_n - b_n \ln((|K(X)s|n + \eta_n)/\eta_n)),
\end{equation}
which has a close relationship with Kullback-Leibler divergence [22]. Since the publication of the seminal paper [28], there have been extensive studies on using TV regularization for inverse problems in imaging applications; see for example [11, 12]. We have compared with Tikhonov and TV regularization schemes (as well as composite approaches that use a combination of Tikhonov and TV), and find that TV is generally most effective for the tomosynthesis image reconstruction problem. We therefore take \( \Omega \) as the TV functional in (10); see Section 3 for the explicit definition of \( \Omega \).

3. Scaled gradient descent algorithm. In this section we develop an algorithm to compute an approximation of the minimization problem (10), where \( \Omega(X) \) is the TV of \( X \) [28]. More specifically, recalling how the unknown \( X \) is defined in equation (6), \( \Omega(X) \) is then written as follows:
\[ \Omega(X) = \sum_{m=1}^{N_{m-1}} \sum_{k=1}^{K} \sum_{\ell=1}^{L} (|\nabla X|_{k\ell,m}) \]
with the discrete gradient operator \( \nabla = (\nabla_1, \nabla_2) : \mathbb{R}^{K \times L} \to \mathbb{R}^{K \times L} \times \mathbb{R}^{K \times L} \), where \( \nabla_1 : \mathbb{R}^{K \times L} \to \mathbb{R}^{K \times L} \) is defined by
\[ (\nabla_1 X)_{k\ell,m} = \begin{cases} x_{k,\ell+1,m} - x_{k\ell,m}, & 1 \leq \ell < L, \\ x_{k,L,m} - x_{k\ell,m}, & \ell = L, \end{cases} \]
for \( k = 1, 2, \ldots, K, \ell = 1, 2, \ldots, L, \) and \( m = 1, 2, \ldots, N_{m-1} \), and \( \nabla_2 : \mathbb{R}^{K \times L} \to \mathbb{R}^{K \times L} \) is defined by
\[ (\nabla_2 X)_{k\ell,m} = \begin{cases} x_{k+1,\ell,m} - x_{k\ell,m}, & 1 \leq k < K, \\ x_{kK,m} - x_{k\ell,m}, & k = K, \end{cases} \]
for \( k = 1, 2, \ldots, K, \ell = 1, 2, \ldots, L, \) and \( m = 1, 2, \ldots, N_{m-1} \).

The rest of this section is devoted to outlining the algorithm we use to solve the optimization problem (10), including computation of the gradient.

3.1. General form of the algorithm. Let \( \nabla \Phi(X) \) denote the gradient of \( \Phi \) with respect to \( X \), and \( \nabla \Phi_{n,m}(X) \) the derivative of \( \Phi(X) \) with respect to \( X_{n,m} \). The Karush-Kuhn-Tucker (KKT) conditions [8] defining the solution \( X^* \) of (10) satisfies:
\[ \begin{cases} 
\nabla \Phi_{n,m}(X^*) = 0, & 0 < X_{n,m}^* < 1, \\
\nabla \Phi_{n,m}(X^*) \geq 0, & X_{n,m}^* = 0, \quad n = 1, 2, \ldots, N_p; \quad m = 1, 2, \ldots, N_{m-1}, \\
\nabla \Phi_{n,m}(X^*) \leq 0, & X_{n,m}^* = 1.
\end{cases} \]
Accordingly, these conditions lead us to solving the following system of nonlinear equations:
\[ \min(X_{n,m}, 1 - X_{n,m}) \nabla \Phi_{n,m}(X) = 0, \]
which amounts to solving the following system of nonlinear equations:

\[(X_{n,m} \nabla \Phi_{n,m}(X) = 0, \text{ for } 0 \leq X_{n,m} \leq \frac{1}{2},)\]

\[(1 - X_{n,m}) \nabla \Phi_{n,m}(X) = 0, \text{ for } \frac{1}{2} < X_{n,m} \leq 1.\]

Based on the nonlinear equations (13) and (14), we propose a scaled gradient descent method for solving the minimization problem (10). Our proposed method is very similar to the multiplicative iterative method proposed in [11], however, the iteration in our algorithm is different from that of [11]. In principle, in each iteration, our strategy is to update all of the \(X_{n,m}\) terms using (13) for \(0 \leq X_{n,m} \leq \frac{1}{2}\), and then update all of the \(X_{n,m}\) terms using (14) for \(\frac{1}{2} < X_{n,m} \leq 1\). We also combine an acceleration strategy in the proposed algorithm. To describe our method, we first introduce some notations needed for the method. For a constant \(c\), we write \(c = [c]^+ - [c]^-\), where \([c]^+ \geq 0, [c]^- \geq 0\), which will be called \(c\)'s positive and negative components respectively. For example, let \([c]^+ = \max(0, c)\) and \([c]^- = -\min(0, c)\), then \(c = [c]^+ - [c]^-.\) Similarly for a function \(f(X)\), let \(f(X)^+\) and \(f(X)^-\) be its positive and negative components respectively, such that \(f(X) = f(X)^+ - f(X)^-\). Rearranging the terms in (13) such that both sides are nonnegative, we have

\[X_{n,m}[(\nabla \Phi_{n,m}(X))^+] = X_{n,m}[(\nabla \Phi_{n,m}(X))^-].\]

This equation immediately provides an updating scheme for \(X_{n,m}\):

\[X^{k+\frac{1}{2}}_{n,m} = X^k_{n,m} \frac{[(\nabla \Phi_{n,m}(X^k))^-]}{[(\nabla \Phi_{n,m}(X^k))^+]}.\]

Similar to (15), we can rearrange (14) as

\[(1 - X_{n,m})[(\nabla \Phi_{n,m}(X))^+] = (1 - X_{n,m})[(\nabla \Phi_{n,m}(X))^+].\]

This equation immediately provides an updating scheme for \(X_{n,m}\):

\[X^{k+\frac{1}{2}}_{n,m} = 1 - (1 - X^k_{n,m}) \frac{[(\nabla \Phi_{n,m}(X^k))^+]}{[(\nabla \Phi_{n,m}(X^k))^+]}.\]

By writing the updating formulae (16) and (17) into a single formula, we have

\[X^{k+\frac{1}{2}}_{n,m} = \begin{cases} 
X^k_{n,m} \frac{[(\nabla \Phi_{n,m}(X^k))^-]}{[(\nabla \Phi_{n,m}(X^k))^+]}, & 0 \leq X^k_{n,m} \leq \frac{1}{2}, \\
1 - (1 - X^k_{n,m}) \frac{[(\nabla \Phi_{n,m}(X^k))^+]}{[(\nabla \Phi_{n,m}(X^k))^+]}, & \frac{1}{2} < X^k_{n,m} \leq 1,
\end{cases}\]

which leads to the following scaled gradient expression:

\[X^{k+\frac{1}{2}}_{n,m} = X^k_{n,m} - S^k_{n,m} \nabla \Phi_{n,m}(X^k),\]

where

\[S^k_{n,m} = \begin{cases} 
\frac{X^k_{n,m}}{[(\nabla \Phi_{n,m}(X^k))^+]}, & 0 \leq X^k_{n,m} \leq \frac{1}{2}, \\
\frac{(1 - X^k_{n,m})}{[(\nabla \Phi_{n,m}(X^k))^+]}, & \frac{1}{2} < X^k_{n,m} \leq 1.
\end{cases}\]

This update has the disadvantage that, if \(X^k_{n,m} = 0\) or \(X^k_{n,m} = 1\), then \(S^k_{n,m} = 0\) so that \(X^{k+\frac{1}{2}}_{n,m}\) cannot be further improved [11, 23]. We adopt the following strategy used in [11] to treat this situation. We keep \(S^k_{n,m} = 0\), if \(X^k_{n,m} = 0\) and \(\nabla \Phi_{n,m}(X^k) \geq 0\), or if \(X^k_{n,m} = 1\) and \(\nabla \Phi_{n,m}(X^k) \leq 0\),
since they satisfy the KKT conditions. For those \( n, m \) where \( X_{n,m}^k = 0 \) with \( \nabla \Phi_{n,m}(X^k) < 0 \), we set \( S_{n,m}^k = \frac{k_0}{\|
abla \Phi_{n,m}(X^k)\|^2} \), with a nonzero constant \( k_0 \). This \( k_0 \) must also ensure the corresponding \( X_{n,m}^{k+\frac{1}{2}} = \frac{k_0}{\|
abla \Phi_{n,m}(X^k)\|^2} \nabla \Phi_{n,m}(X^k) \leq 1 \), and so we have

\[
S_{n,m}^k = \min \left( \frac{k_0}{\|
abla \Phi_{n,m}(X^k)\|^2}, \frac{1}{\nabla \Phi_{n,m}(X^k)} \right).
\]

Similarly, for those \( n, m \) where \( X_{n,m}^k = 1 \) with \( \nabla \Phi_{n,m}(X^k) > 0 \), we set \( S_{n,m}^k = \frac{1-k_0}{\|
abla \Phi_{n,m}(X^k)\|^2} \). This \( 1-k_0 \) must also ensure the corresponding \( X_{n,m}^{k+\frac{1}{2}} = 1 - \frac{1-k_0}{\|
abla \Phi_{n,m}(X^k)\|^2} \nabla \Phi_{n,m}(X^k) \geq 0 \), and so we have

\[
S_{n,m}^k = \min \left( \frac{k_0}{\|
abla \Phi_{n,m}(X^k)\|^2}, \frac{1}{\nabla \Phi_{n,m}(X^k)} \right).
\]

We choose \( k_0 = 2e - 2 \) in this paper. To keep the algorithm more stable, we also take the following strategies: for the iteration, if \( X_{n,m}^k \frac{\|\nabla \Phi_{n,m}(X^k)\|}{\|\nabla \Phi_{n,m}(X^k)\|^2} > 1 \), we use the iteration (17) to update \( X_{n,m}^k \), and similarly, if \( 1 - (1 - X_{n,m}^k) \frac{\|\nabla \Phi_{n,m}(X^k)\|}{\|\nabla \Phi_{n,m}(X^k)\|^2} < 0 \), we use the iteration (16) to update \( X_{n,m}^k \). Thus the scale matrix is determined in each iteration. By incorporating a line search strategy, we get the following iteration:

\[
X_{n,m}^{k+1} = X_{n,m}^k - \alpha^k S^k \nabla \Phi(X^k),
\]

where the scale matrix \( S^k \) is a diagonal matrix with diagonal entries \( S_{n,m}^k \geq 0 \), and \( \alpha^k \) is a step size determined by the line search.

Generally, a line search forces objective function decrease at every iteration. Here, we take the following Armijo like rule:

\[
\Phi(X^{k+1}) \leq \Phi(X^k) + \mu \alpha^k \langle \nabla \Phi(X^k), D^k \rangle,
\]

where \( \langle \cdot, \cdot \rangle \) denotes inner product, \( D^k = X^{k+1/2} - X^k \), \( 0 < \mu < 1 \) is a fixed parameter such as \( \mu = 10^{-5} \). Concretely, we first set \( \alpha = \alpha^0 \), and if

\[
\Phi(X^k + \alpha D^k) \leq \Phi(X^k) + \mu \alpha \langle \nabla \Phi(X^k), D^k \rangle,
\]

then we update the iteration \( X_{n,m}^{k+1} = X^k + \alpha D^k \) and stop; otherwise, we reset \( \alpha = \rho \alpha \) (using, for example, \( \rho = 0.5 \)) and reevaluate the line search condition. Notice that because \( S^k \) is non-negative definite, iteration (23) with \( \alpha^k = 1 \) (which is actually \( X^{k+\frac{1}{2}} \)) cannot move along the uphill direction. Thus if \( \Phi(X^{k+1}) > \Phi(X^k) \), it indicates the necessity of reducing the step size. Therefore, we use the initial step length \( \alpha^0 = 1 \) in the line search.

3.2. Acceleration and stabilization. Gradient based methods have the disadvantage of slow convergence rate. To accelerate the speed of an algorithm, we may generate new iteration steps utilizing the two previous iterates, and this strategy often improves the performance of gradient based methods. For example, it is well known for linear systems that if \( X^k \) and \( X^{k-1} \) are two iterates of the Gauss-Seidel method, and \( \omega = (1 + \nu) > 1 \) is a relaxation parameter (over relaxation), then the new iterate \( Z^k = \omega X^k + (1 - \omega)X^{k-1} = X^k + \nu(X^k - X^{k-1}) \) generates the SOR iterate. This extrapolation strategy can also be used to design fast gradient based methods to solve convex optimization problems. Specifically, suppose
\(X^k\) and \(X^{k-1}\) are two iterates of a gradient based method for solving convex optimization problems, then it is proposed in [5, 6, 25] to generate a new iterate \(Z^k = X^k + \nu^k(X^k - X^{k-1})\), where \(\nu^k\) is adaptively updated by
\[
(25) \quad \nu^k = \frac{(t^{k-1} - 1)}{t^k},
\]
with
\[
(26) \quad t^k = \frac{1 + \sqrt{1 + 4(t^{k-1})^2}}{2}, \text{ and } t^0 = 1.
\]
It has been shown that this strategy to automatically update \(\nu\) can significantly improve convergence speed for convex optimization problems. However, since our problem is nonconvex, the iterates may oscillate seriously and even diverge. To overcome this drawback, we adopt the idea suggested in [6] that incorporates a monotone version of the fast proximal gradient method. Specifically, we generate a new iterate
\[
Z^k = \arg\min \left(\Phi(X^k + \nu^k(X^k - X^{k-1})), \Phi(X^k)\right),
\]
which helps to stabilize the behavior of our algorithm. We now state our algorithm explicitly in the following.

**Algorithm 1.** Scaled gradient descent algorithm for polyenergetic image reconstruction.

Let \(k\) be the iteration number and \(\Phi(X)\) be the objective function defined in (10). Initialization: Set \(k = -1\). Choose an initial value \(0 \leq X^0 \leq 1\), and take \(Y^0 = X^0\). Set \(t^0 = 1\). While not convergent, do

1. **Iteration \(k = k + 1\):**
   1. **Step 1:** Compute \(\nabla \Phi_{n,m}(Y^k)\), and set \(\delta_{n,m}^{k,1} = [\nabla \Phi_{n,m}(Y^k)]^-\), \(\delta_{n,m}^{k,2} = [\nabla \Phi_{n,m}(Y^k)]^+\).
   2. **Step 2:** For \(n,m\), where \(0 \leq Y_{n,m}^k \leq \frac{1}{2}\) or \(\frac{1}{2} < Y_{n,m}^k \leq 1\) with \(1 - (1 - Y_{n,m}^k)\frac{\delta_{n,m}^{k,2}}{\delta_{n,m}^{k,1}} < 0\),
      compute \(S_{n,m}^k = \frac{Y_{n,m}^k}{\delta_{n,m}^{k,1}}\);
      for \(n,m\), where \(\frac{1}{2} < Y_{n,m}^k \leq 1\), or \(0 \leq Y_{n,m}^k \leq \frac{1}{2}\) with \(Y_{n,m}^k \frac{\delta_{n,m}^{k,1}}{\delta_{n,m}^{k,2}} > 1\),
      compute \(S_{n,m}^k = \frac{(1 - Y_{n,m}^k)}{\delta_{n,m}^{k,1}}\).
   3. **Step 3:** Compute \(X^{k+\frac{1}{2}} = Y^k - S^k \nabla \Phi(Y^k)\).
   4. **Step 4:** Line search:
      If \(\Phi(X^{k+\frac{1}{2}}) < \Phi(X^k)\), then set \(Z^{k+1} = X^{k+\frac{1}{2}}\);
      otherwise \(Z^{k+1} = Y^k - \alpha^k S^k \nabla \Phi(Y^k)\), where \(\alpha^k \in (0, 1)\) is the step size satisfies
      \(\Phi(Z^{k+1}) \leq \Phi(Y^k) + \mu \alpha^k (\nabla \Phi(X^k), D^k)\),
      with \(D^k = X^{k+\frac{1}{2}} - Y^k\).
   5. **Step 5:** Compute \(X^{k+1} = \arg\min \left(\Phi(Z^{k+1}), \Phi(X^k)\right)\);
   6. **Step 6:** Compute
      \[
      t^k = \frac{1 + \sqrt{1 + 4(t^{k-1})^2}}{2}, \text{ and } t^0 = 1.
      \]
\[ Y^{k+1} = \text{Prj}\{X^{k+1} + \frac{t^k}{t^{k+1}} (Z^{k+1} - X^{k+1}) + \frac{t^k - 1}{t^{k+1}} (X^{k+1} - X^k)\}, \]

where \( \text{Prj} \) is a projection operator. That is, let \( Y = \text{Prj}\{X\} \), then

\[ Y_{n,m} = \begin{cases} X_{n,m}, & \text{if } X_{n,m} \in [0,1], \\ 0, & \text{if } X_{n,m} < 0, \\ 1, & \text{if } X_{n,m} > 1. \end{cases} \]

Proposition 1. For the scaled gradient algorithm 1, if \( Y_{n,m}^k \in [0,1] \), then \( Z_{n,m}^{k+1} \in [0,1] \) for \( n = 1,2,\ldots, n_v, m = 1,2,\ldots, N_m - 1 \). Accordingly, \( X_{n,m}^{k+1} \in [0,1] \) for \( n = 1,2,\ldots, N_v, m = 1,2,\ldots, N_m - 1 \).

Proof. First consider \( 0 < Y_{n,m}^k \leq \frac{1}{2} \). According to (18), for those \( Y_{n,m}^k, 0 < Y_{n,m}^k \leq \frac{1}{2} \), that satisfy \( Y_{n,m}^k \frac{\delta_{n,m}^{k+2}}{\delta_{n,m}^{k+1}} \leq 1 \), the corresponding \( X_{n,m}^{k+2} \in [0,1] \); for those \( Y_{n,m}^k, 0 < Y_{n,m}^k \leq \frac{1}{2} \), that satisfy \( Y_{n,m}^k \frac{\delta_{n,m}^{k+2}}{\delta_{n,m}^{k+1}} > 1 \), \( X_{n,m}^{k+2} \leq 1 \) from (18), and \( X_{n,m}^{k+2} \geq 0 \) can be obtained from the fact that

\[ X_{n,m}^{k+2} = \left( 1 - \frac{\delta_{n,m}^{k+2}}{\delta_{n,m}^{k+1}} \right) + Y_{n,m}^k \frac{\delta_{n,m}^{k+2}}{\delta_{n,m}^{k+1}}, \]

and the fact that \( \frac{\delta_{n,m}^{k+2}}{\delta_{n,m}^{k+1}} < 1 \). We then consider the case \( \frac{1}{2} < Y_{n,m}^k < 1 \). According to (18), for those \( Y_{n,m}^k, \frac{1}{2} < Y_{n,m}^k < 1 \), that satisfy \( 1 - (1 - Y_{n,m}^k) \frac{\delta_{n,m}^{k+2}}{\delta_{n,m}^{k+1}} \geq 0 \), then \( X_{n,m}^{k+2} = 1 - (1 - Y_{n,m}^k) \frac{\delta_{n,m}^{k+2}}{\delta_{n,m}^{k+1}} \geq 0 \); and \( 1 - X_{n,m}^{k+2} = (1 - Y_{n,m}^k) \frac{\delta_{n,m}^{k+2}}{\delta_{n,m}^{k+1}} \geq 0 \), so \( X_{n,m}^{k+2} \in [0,1] \); for those \( Y_{n,m}^k, \frac{1}{2} < Y_{n,m}^k < 1 \), that satisfy \( 1 - (1 - Y_{n,m}^k) \frac{\delta_{n,m}^{k+2}}{\delta_{n,m}^{k+1}} < 0 \), we have, \( \frac{\delta_{n,m}^{k+2}}{\delta_{n,m}^{k+1}} < (1 - Y_{n,m}^k) < 1 \), so \( X_{n,m}^{k+2} = Y_{n,m}^k \frac{\delta_{n,m}^{k+2}}{\delta_{n,m}^{k+1}} \in [0,1] \); For \( Y_{n,m}^k = 0 \) or \( Y_{n,m}^k = 1 \), we select the corresponding \( S_{n,m}^{k+1} \) ascertains that \( X_{n,m}^{k+2} \in [0,1] \). Now from \( X_{n,m}^{k+2} \in [0,1] \), we have \( Z_{n,m}^{k+1} \in [0,1] \) since \( Z_{n,m}^{k+1} = (1 - \alpha^k) Y_{n,m}^k + \alpha^k X_{n,m}^{k+2} \) and \( 0 < \alpha^k < 1 \), and accordingly, \( X_{n,m}^{k+1} \in [0,1] \) for \( n = 1,2,\ldots, N_v, m = 1,2,\ldots, N_m - 1 \).

3.3. Gradient of the objection function. In our algorithm, we need to evaluate the gradient of the objection function \( \Phi(X) \). So we need to compute \( \nabla \mathcal{L}(X) \), and \( \nabla \Omega(X) \). To compute the gradient of the TV term, we consider the slightly modified approximation of the TV function

\[ \Omega(X) = \sum_{m=1}^{N_v-1} \sum_{k=1}^{K} \sum_{\ell=1}^{L} \sqrt{(\nabla_1 X)_{k\ell,m}^2 + (\nabla_2 X)_{k\ell,m}^2} + \epsilon \]

where \( \epsilon \) is a small positive number that is used to avoid numerical difficulties when computing the gradient. This approximation is typical when working with TV; see, for example, [33]. The performance is not sensitive to the selection of this parameter; in our experiments we choose \( \epsilon \) to be on the order of machine precision.

If we use the notation \( |\nabla(X)_{k\ell,m}|_2 = \sqrt{(\nabla_1 X)_{k\ell,m}^2 + (\nabla_2 X)_{k\ell,m}^2} + \epsilon \), then we have

\[ \nabla \Omega(X) = \sum_{m=1}^{N_v-1} \sum_{k=1}^{K} \sum_{\ell=1}^{L} \nabla_1 ((\nabla_1 X)_{k\ell,m}) + \nabla_2 (|\nabla(X)_{k\ell,m}|_2), \]
where $\nabla^T: \mathcal{R}^{K \times L} \rightarrow \mathcal{R}^{K \times L}$ is defined by

$$\nabla^T X_{k\ell,m} = \begin{cases} x_{k-1,\ell,m} - x_{k\ell,m}, & 1 < \ell < L, \\ x_{k,\ell-1,m} - x_{k\ell,m}, & \ell = 1, \\ -x_{k\ell,m}, & \ell = L, \\ x_{k,\ell-1,m}, & \ell = L, \end{cases}$$

for $k = 1, 2, \ldots, K$, $\ell = 1, 2, \ldots, L$, $m = 1, 2, \ldots, N_m - 1$, and $\nabla^T_2: \mathcal{R}^{K \times L} \rightarrow \mathcal{R}^{K \times L}$ is defined by

$$\nabla^T_2 X_{k\ell,m} = \begin{cases} x_{k-1,\ell,m} - x_{k\ell,m}, & 1 < k < K, \\ x_{k,\ell-1,m} - x_{k\ell,m}, & k = 1, \\ -x_{k\ell,m}, & k = 1, \\ x_{k-1,\ell,m}, & k = K, \end{cases}$$

for $k = 1, 2, \ldots, K$, $\ell = 1, 2, \ldots, L$, and $m = 1, 2, \ldots, N_m - 1$.

For the gradient of $\mathcal{L}(X)$, let $e_k^T A$ denote the $k$th row of matrix $A$, where $e_k$ is the $k$th standard unit vector (that is, the $k$th column of the identity matrix). Similarly, we denote the $k$th row of matrix $\hat{C}$ by $e_k^T \hat{C}$. Then we have

$$\mathcal{L}(X) = \sum_{n=1}^N (\langle [K(X)s]_n + \eta_n \rangle - b_n - b_n \ln(\langle [K(X)s]_n + \eta_n \rangle/b_n))$$

$$= \sum_{n=1}^N \left( \sum_{i=1}^{N_x} \exp(-e_n^T A(1c_{1,i} + X\hat{C}^T e_i))s_i + \eta_n - b_n \right)$$

$$- \sum_{n=1}^N b_n \ln \left( \sum_{i=1}^{N_x} \left( \exp(-e_n^T A(1c_{1,i} + X\hat{C}^T e_i))s_i + \eta_n \right)/b_n \right).$$

So we have

$$\nabla \mathcal{L}(X) = \sum_{n=1}^N \left( -A^T e_n \sum_{i=1}^{N_x} \exp(-e_n^T A(1c_{1,i} + X\hat{C}^T e_i))e_k^T \hat{C} s_i \right)$$

$$- \sum_{n=1}^N A^T e_n b_n \sum_{i=1}^{N_x} \left( \exp(-e_n^T A(1c_{1,i} + X\hat{C}^T e_i))s_i + \eta_n \right)$$

$$= A^T V(X),$$

where

$$V(X) = [v_1(X) \mid v_2(X) \mid \cdots \mid v_{N_m-1}(X)]$$

and

$$v_j(X) = \langle b \odot (K(X)s + \eta) - 1 \rangle \odot (K(X)(s \odot (\hat{C}e_j))), \ j = 1, 2, \ldots, N_m - 1.$$
and \( \max \{ \cdot \} \) and \( \min \{ \cdot \} \) are componentwise taken, where

\[
\begin{align*}
\{V(X)\}^+ &= [v_1(X)]^+ | [v_2(X)]^+ | \cdots | [v_{N_m-1}(X)]^+], \\
\{V(X)\}^- &= [v_1(X)]^- | [v_2(X)]^- | \cdots | [v_{N_m-1}(X)]^-], \\
[v_j(X)]^+ &= b \odot (K(X)s + \eta) \odot (K(X)(s \odot (\hat{C}e_j))), \quad j = 1, 2, \ldots, N_m - 1, \\
[v_j(X)]^- &= K(X)(s \odot (\hat{C}e_j)), \quad j = 1, 2, \ldots, N_m - 1.
\end{align*}
\]

4. Experimental results. In this section, we illustrate the feasibility of the proposed algorithm for both the polyenergetic tomosynthesis and x-ray CT problems. Specifically, we consider \( 128 \times 128 \) images that simulate breast tissue with two materials, glandular and adipose. In one test image (see top of Figure 1) we assume the material is made up of distinct materials in various regions. That is, each region is either 100% adipose and 0% glandular, or 0% adipose and 100% glandular. In the second test problem, we consider the situation where regions are as in the first test image, but also contain regions with a mix of 50% adipose and 50%
glandular materials (see the top of Figure 2). The test images were generated using the phantoms provided in the MATLAB package AIRtools [19]. All computations in our experiments were performed in double precision using MATLAB R2014a on an Intel(R) Core(TM) i5-4200 CPU @1.6 GHz, 4.00 GB RAM. We stop our algorithm when the approximation obtained at iteration $k$ satisfies the inequality

$$\|X^k - X^{k-1}\| \leq 10^{-4},$$

or when the maximum number of iterations exceeds 20000.

To provide a measure of quantitative performance, we use the relative error (RErr):

$$\text{RErr} = \frac{\|X - \tilde{X}\|}{\|X\|},$$

where $\tilde{X}$ and $X$ denote the approximate (reconstructed) solution and the original true solution, respectively.

For each experiment we use AIRtools [19] to construct a fan beam projection matrix $A$, assuming 70 cm from source to detector, 2.5 cm air gap between object and detector and an object size of 50 cm (these quantities are typical for breast tomosynthesis). In addition, for an initial guess we always use a random set of weights $X_0$, with $0 \leq X_0 \leq 1$.

**Experiment 1.** For the first experiment, we test the algorithm for the full CT case. We use AIRtools to construct a fan beam projection matrix $A$ with the angular range of 0 to 359 degrees. Then we use (7) to obtain $\hat{b}$,

$$\hat{b} = \exp(-A(1c^T + CX^T))s + \eta.$$

Poisson noise was incorporated so that the relative difference between the noise free data and noisy data was 1e-3.

We take $k_0 = 10^{-3}$, $\mu = 10^{-4}$, $\rho = 0.5$, $\epsilon = \text{eps}$, the machine epsilon (for double precision, this is approximately $2.2204 \cdot 10^{-16}$). The regularization parameter is chosen as $\lambda = 10^{-6}$. The obtained results are shown in Figures 1 and 2, where in each case the top images show the two true materials (left is adipose, right is glandular), and the middle images show the corresponding reconstructions. The bottom of these figures display the sinogram data on the left, and algorithm convergence history (i.e., the RErr plots) on the right. From the plots of the relative error and the quality of the restored image, we see that our algorithm is very effective.

**Experiment 2.** For the second experiment, we test the limited angle (tomosynthesis) case with an angular range of -30 to 30 degrees. We take $k_0 = 0.02$, $\mu = 10^{-5}$, $\lambda = 10^{-7}$, $\rho = 0.5$, and $\epsilon = \text{eps}$, the machine epsilon (for double precision, this is approximately $2.2204 \cdot 10^{-16}$). Moreover, we corrupt the data with Poisson noise so that the relative difference between the noise and noise free data is $10^{-3}$. The obtained results for this case are shown in Figures 3 and 4. As with the previous experiment, the top rows of these figures show the true (left) and reconstructed (right) weights for the first material, and the second rows show analogous results for the second material. In the bottom rows we display the data (i.e., sinogram) on the left, and the RErr plots on the right.

As previously mentioned, we collect the projections within a very limited angle range for the tomosynthesis problem, so we do not expect to obtain high-quality resolution in all directions. In particular, if the source at 0 degrees is at the top of the object, then we would expect the resolution to be much worse for slices in the vertical direction than those in the horizontal direction. To confirm this observation,
we compare the sum of all pixels along each column then along each row, and the relative errors of coarse samplings (in both horizontal and vertical directions) of the true and reconstructed weights. These results, which are shown in Figures 5 and 6, show that we indeed obtain better results with coarse vertical (and fine horizontal) resolution than that with coarse horizontal (and fine vertical) resolution.

We also notice that, in Figure 3, the relative error decrease first and then increase as the iteration number increases, this phenomenon is called semi-convergence. Some more elegant strategy for the regularization parameter selection may reduce this phenomenon.

**Experiment 3.** In the final experiment, we begin with an object that has finer details in the horizontal direction than in the vertical direction. Specifically, we assume the same geometric dimensions of the object as in the previous experiments, but here we use a coarse discretization of $16 \times 128$ (that is, we have thick, uniform horizontal slices). We form the measured data,

$$\hat{b} = \exp(-A(1c_1^T + X\hat{C}^T))s + \eta.$$
and again incorporate Poisson noise so that the relative difference between the noise free data and the noisy data is 0.001.

The obtained results are in Figures 7 and 8. As with the previous two experiments, the top rows of these figures show the true (left) and reconstructed (right) weights for the first material, and the second rows show analogous results for the second material. In the bottom rows we display the data (i.e., sinogram) on the left, and the RErr plots on the right. From the figures, we see that the obtained results are acceptable.

5. **Conclusions.** Based on KKT conditions, and extending idea from fast gradient methods for solving convex optimization problems, we develop a scaled gradient method for polyenergetic digital tomographic image reconstruction. We show that by incorporating TV regularization and bound constraints that the method can produce results that allow for quantitative imaging.
Acknowledgments. We would like to thank the anonymous referees for their constructive comments and suggestions which have greatly improved the development and presentation of this paper.

REFERENCES

[1] R. E. Alvarez and A. Macovski, Energy-selective reconstructions in x-ray computerized tomography, Phys. Med. Biol., 21 (1976), 733–744.
[2] R. E. Alvarez, J. A. Seibert and S. K. Thompson, Comparison of dual energy detector system performance, Med. Phys., 31 (2004), 556–565.
[3] R. F. Barber, E. Y. Sidky, T. G. Schmidt and X.-C Pan, An algorithm for constrained one-step inversion of spectral CT data, Phys. Med. Biol., 61 (2016), 3784–3818.
[4] M. Bazalova, J. Carrier, L. Beaulieu and F. Verhaegen, Dual-energy CT-based material extraction for tissue segmentation in Monte Carlo dose calculations, Phys. Med. Biol., 53 (2008), 2439–2456.
Scaled Gradient Method for Digital Tomographic Image

Figure 5. From left to right and upper to bottom: sum along each row for first material, sum along each column for first material, sum along each row for second material, sum along each column for second material, relative error of reduced resolution image along rows and columns for first material, relative error of reduced resolution image along rows and columns for second material.

[5] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imag. Sci., 2 (2009), 183–202.
[6] A. Beck and M. Teboulle, Gradient-based algorithms with applications to signal recovery problems, in Convex Optimization in Signal Processing and Communications (eds. Y. Eldar and D. Palomar), Cambridge University Press, (2010), 42–88.
[7] M. J. Berger, J. H. Bubbell, S. M. Seltzer, J. Chang, J. S. Coursey, R. Sukumar, D. S. Zucker and K. Olsen, XCOM: Photon Cross Sections Database, National Institutes of Standards, 2009. http://www.nist.gov/pml/data/xcom/index.cfm.
[8] D. P. Bertsekas, Nonlinear Programming, 2nd edition, Athena Scientific, Belmont, Mass., 1999.
[9] R. Brooks and G. Di Chiro, Beam hardening in x-ray reconstructive tomography, Phys. Med. Biol., 21 (1976), 390–398.
Figure 6. From left to right and upper to bottom: sum along each row for first material, sum along each column for first material, sum along each row for second material, sum along each column for second material, relative error of reduced resolution image along rows and columns for first material, relative error of reduced resolution image along rows and columns for second material.

[10] V. M. Bustamante, J. G. Nagy, S. S. J. Feng and I. Sechopoulos, Iterative breast tomosynthesis image reconstruction, SIAM J. Sci. Comput., 35 (2013), S192–S208.
[11] R. Chan and J. Ma, A multiplicative iterative algorithm for box-constrained penalized likelihood image restoration, IEEE Trans. Image Process., 21 (2012), 3168–3181.
[12] T. F. Chan and J. Shen, Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods SIAM, Philadelphia, 2005.
[13] J. Chung, J. G. Nagy and I. Sechopoulos, Numerical algorithms for polyenergetic digital breast tomosynthesis reconstruction, SIAM J. Imaging Sci., 3 (2010), 133–152.
[14] B. De Man, J. Nuyts, P. Dupont, G. Marchal and P. Suetens, An iterative maximumlikelihood polychromatic algorithm for CT, IEEE Trans. Med. Imaging, 20 (2001), 999–1008.
[15] I. A. Elbakri and J. A. Fessler, Statistical image reconstruction for polyenergetic x-ray computed tomography, IEEE Trans. Med. Imaging, 21 (2002), 89–99.
Figure 7. From left to right and upper to bottom: original first material, reconstructed first material, original second material, reconstructed second material, sinogram image and RErr with iteration.

[16] I. A. Elbakri and J. A. Fessler, Segmentation-free statistical image reconstruction for polyenergetic x-ray computed tomography with experimental validation, *Phys. Med. Biol.*, 48 (2003), 2453–2477.

[17] C. L. Epstein, *Introduction to the Mathematics of Medical Imaging*, 2nd edition, SIAM, Philadelphia, 2008.

[18] G. R. Hammerstein, D. W. Miller, D. R. White, M. E. Masterson, H. Q. Woodard and J. S. Laughlin, Absorbed radiation dose in mammography, *Radiology*, 130 (1979), 485–491.

[19] P. C. Hansen and M. Saxild-Hansen, AIR tools – A MATLAB package of algebraic iterative reconstruction methods, *J. Comput. Appl. Math.*, 236 (2012), 2167–2178.

[20] B. Heismann and M. Balda, Quantitative image-based spectral reconstruction for computed tomography, *Med. Phys.*, 36 (2009), 4471–4485.

[21] A. Karellas, J. Lo and C. Orton, Point/counterpoint: Cone beam x-ray CT will be superior to digital x-ray tomosynthesis in imaging the breast and delineating cancer, *Med. Phys.*, 35 (2008), 409–411.

[22] S. Kullback and R. A. Leibler, On information and sufficiency, *Annals of Math. Statis.*, 22 (1951), 79–86.
Figure 8. From left to right and upper to bottom: original first material, reconstructed first material, original second material, reconstructed second material, sinogram image and RErr with iteration.

[23] D. D. Lee and H. S. Seung, Algorithms for non-negative matrix factorization, in NIPS, (2000), 556–562.

[24] A. Macovski, R. E. Alvarez, J. L.-H. Chan, J. P. Stonestrom and L. M. Zatz, Energy dependent reconstruction in x-ray computerized tomography, Comput. Biol. Med., 6 (1976), 335–336.

[25] Y. E. Nesterov, A method for solving the convex programming problem with convergence rate O(1/k^2), Dokl. Akad. Nauk SSSR, 269 (1983), 543–547.

[26] J. Noh, J. A. Fessler and P. E. Kinahan, Statistical sinogram restoration in dual-energy CT for PET attenuation correction, IEEE Trans. Med. Imaging, 28 (2009), 1688–1702.

[27] J. A. O’Sullivan and J. Benac, Alternating minimization algorithms for transmission tomography, IEEE Trans. Med. Imaging, 26 (2007), 283–297.

[28] L. I. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D, 60 (1992), 259–268.

[29] I. Sechopoulos, A review of breast tomosynthesis. Part I. The image acquisition process, Med. Phys., 40 (2013), 014301, 12pp.

[30] I. Sechopoulos, A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications, Med. Phys., 40 (2013), 014302, 17pp.
[31] P. Stenner, T. Berkus and M. Kachelriess, Empirical dual energy calibration (EDEC) for cone-beam computed tomography, *Med. Phys.*, **34** (2007), 3630–3641.

[32] P. Sukovic and N. H. Clinthorne, Penalized weighted least-squares image reconstruction for dual energy x-ray transmission tomography, *IEEE Trans. Med. Imaging*, **19** (2000), 1075–1081.

[33] C. R. Vogel, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002.

Received March 2017; 1st revision June 2017; 2nd revision July 2017.

*E-mail address*: jjzhang@staff.shu.edu.cn

*E-mail address*: yunyi.hu@emory.edu

*E-mail address*: nagy@mathcs.emory.edu