

GENERAL CONVERGENT EXPECTATION MAXIMIZATION (EM)-TYPE ALGORITHMS FOR IMAGE RECONSTRUCTION

MING YAN

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA 90095, USA

ALEX A.T. BUI

Department of Radiological Sciences
University of California, Los Angeles
Los Angeles, CA 90095, USA

JASON CONG

Department of Computer Sciences
University of California, Los Angeles
Los Angeles, CA 90095, USA

LUMINITA A. VESE

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA 90095, USA

(Communicated by the associate editor name)

ABSTRACT. Obtaining high quality images is very important in many areas of applied sciences, such as medical imaging, optical microscopy, and astronomy. Image reconstruction can be considered as solving the ill-posed and inverse problem $y = Ax + n$, where x is the image to be reconstructed and n is the unknown noise. In this paper, we propose general robust expectation maximization (EM)-Type algorithms for image reconstruction. Both Poisson noise and Gaussian noise types are considered. The EM-Type algorithms are performed using iteratively EM (or SART for weighted Gaussian noise) and regularization in the image domain. The convergence of these algorithms is proved in several ways: EM with *a priori* information and alternating minimization methods. To show the efficiency of EM-Type algorithms, the application in computerized tomography reconstruction is chosen.

1. Introduction. Obtaining high quality images is very important in many areas of applied science, such as medical imaging, optical microscopy, and astronomy. For some applications such as positron-emission-tomography (PET) and computed tomography (CT), analytical methods for image reconstruction are available. For instance, filtered back projection (FBP) is the most commonly used method for image reconstruction from CT by manufacturers of commercial imaging equipments [41].

2010 *Mathematics Subject Classification.* Primary: 90C30, 94A08; Secondary: 90C26, 92C55.

Key words and phrases. Expectation maximization (EM), Richardson-Lucy, simultaneous algebraic reconstruction technique (SART), image reconstruction, total variation (TV), computerized tomography.

This work is supported by the Center for Domain-Specific Computing (CDSC) funded by the NSF Expedition in Computing Award CCF-0926127.

However, it is sensitive to noise and suffers from streak artifacts (star artifacts). An alternative to this analytical reconstruction is the use of the iterative reconstruction technique, which is quite different from FBP. The main advantages of the iterative reconstruction technique over FBP are insensitivity to noise and flexibility [28]. The data can be collected over any set of lines, the projections do not have to be distributed uniformly in angle, and the projections can be even incomplete (limited angle). With the help of parallel computing and graphics processing units (GPUs), even iterative methods can be solved very fast. Therefore, iterative methods become more and more important, and we will focus on the iterative reconstruction technique only.

The degradation model can be formulated as a linear inverse and ill-posed problem:

$$(1) \quad y = Ax + b + n.$$

Here, y is the measured data (vector in \mathbf{R}^M in the discrete case). A is a compact operator (matrix in $\mathbf{R}^{M \times N}$ in the discrete case). For all the applications we will consider, the entries of A are nonnegative and A does not have full column rank. x is the desired exact image (vector in \mathbf{R}^N in the discrete case). b is the background emission and n is the noise (both are vectors in \mathbf{R}^M in the discrete case). We will consider the case without background emission ($b = 0$) in this paper. Since the matrix A does not have full column rank, the computation of x directly by finding the inverse of A is not reasonable because (1) is ill-posed and n is unknown. Even for the case without noise ($n = 0$), there are many solutions because A does not have full column rank. When there is noise in the measured data ($n \neq 0$), finding x is more difficult because of the unknown n . Therefore regularization techniques are needed for solving these problems efficiently.

One powerful technique for applying regularization is the Bayesian model, and a general Bayesian model for image reconstruction was proposed by Geman and Geman [15], and Grenander [18]. The idea is to use *a priori* information about the image x to be reconstructed. In the Bayesian approach, we assume that measured data y is a realization of a multi-valued random variable, denoted by Y and the image x is also considered as a realization of another multi-valued random variable, denoted by X . Therefore, Bayes' theorem gives us

$$(2) \quad p_X(x|y) = \frac{p_Y(y|x)p_X(x)}{p_Y(y)}.$$

This is a conditional probability of having $X = x$ given that y is the measured data. After inserting the detected value y , we obtain *a posteriori* probability distribution of X . Then we can find x^* such that $p_X(x|y)$ is maximized, as maximum a posteriori (MAP) likelihood estimation.

In general, X is assigned a Gibbs random field, which is a random variable with the following probability distribution

$$(3) \quad p_X(x) \sim e^{-\beta J(x)},$$

where $J(x)$ is a given energy function ($J(x)$ can be non-convex), and β is a positive parameter. There are many different choices for $J(x)$ depending on the applications. Some examples are, for instance, quadratic penalization $J(x) = \|x\|_2^2/2$ [12, 34], quadratic gradient $J(x) = \|\nabla x\|_2^2/2$ [54], total variation $J(x) = \|\nabla x\|_1$, and modified total variation [38, 40, 14, 6, 49, 51]. We also mention Good's roughness penalization $J(x) = \|\nabla x\|^2/x\|_1$ [26] for a slightly different regularization. We

assume here that $\|\cdot\|_1$ and $\|\cdot\|_2$ are the ℓ_1 and ℓ_2 norms respectively, and that ∇x denotes the discrete gradient operator of an image x .

For the choices of probability densities $p_Y(y|x)$, we can choose

$$(4) \quad p_Y(y|x) \sim e^{-\|Ax-y\|_2^2/(2\sigma^2)}$$

in the case of additive Gaussian noise, and the minimization of the negative log-likelihood function gives us the famous Tikhonov regularization method [46]

$$(5) \quad \underset{x}{\text{minimize}} \quad \frac{1}{2}\|Ax-y\|_2^2 + \beta J(x).$$

If the random variable Y of the detected values y follows a Poisson distribution [31, 42] with an expectation value provided by Ax instead of Gaussian distribution, we have

$$(6) \quad y_i \sim \text{Poisson}\{(Ax)_i\}, \quad \text{i.e.,} \quad p_Y(y|x) \sim \prod_i \frac{(Ax)_i^{y_i}}{y_i!} e^{-(Ax)_i}.$$

By minimizing the negative log-likelihood function, we obtain the following optimization problem

$$(7) \quad \underset{x \geq 0}{\text{minimize}} \quad \sum_i ((Ax)_i - y_i \log(Ax)_i) + \beta J(x).$$

In this paper, we will focus on solving (5) and (7). It is easy to see that the objective functions in (5) and (7) are convex if $J(x)$ is a convex function. Additionally, with suitably chosen regularization $J(x)$, the objective functions are strictly convex, and the solutions to these problems are unique. If $J(x) = \|\nabla x\|_1$, i.e., the regularization is the total variation, the well-posedness of the regularization problems is shown in [1] and [6] for Gaussian and Poisson noise respectively. In this paper, we also consider a more general case where $J(x)$ is non-convex.

Relevant prior works are by Le et al. [31], Brune et al. [6, 7, 8], and Sidky et al. [44]. Additionally, we refer to Jia et al. [23], Jung et al. [27], Setzer et al. [39], Jafarpour et al. [22], Harmany et al. [19], Willet et al. [48], among other work. We also refer to the Compressive Sensing Resources [55]. The difference of this work from those will be discussed later.

The paper is organized as follows. In section 2, we will give a short introduction of expectation maximization (EM) iteration, or Richardson-Lucy algorithm, used in image reconstruction without background emission from the view of optimization. In section 3, we will propose general EM-Type algorithms for image reconstruction without background emission when the measured data is corrupted by Poisson noise. This is based on the maximum *a posteriori* likelihood estimation and an EM step. In this section, these EM-Type algorithms are shown to be equivalent to EM algorithms with *a priori* information. In addition, these EM-Type algorithms are also considered as alternating minimization methods for equivalent optimization problems, and the convergence results are obtained from both convex and non-convex $J(x)$. When the noise is weighted Gaussian noise, we also have the similar EM-Type algorithms. Simultaneous algebraic reconstruction technique is shown to be EM algorithm in section 4, and EM-Type algorithms for weighted Gaussian noise are introduced in section 5. In section 5, we also show the convergence analysis of EM-Type algorithms for weighted Gaussian noise via EM algorithms with *a priori* information and alternating minimization methods. Some numerical experiments on image reconstruction are given in section 6 to show the efficiency of the EM-Type algorithms. We will end this work by a short conclusion section.

2. Expectation Maximization (EM) Iteration. A maximum likelihood (ML) method for image reconstruction based on Poisson data was introduced by Shepp and Vardi [42] in 1982 for image reconstruction in emission tomography. In fact, this algorithm was originally proposed by Richardson [37] in 1972 and Lucy [33] in 1974 for image deblurring in astronomy. The ML method is a method for solving the special case of problem (7) without regularization term, i.e., $J(x)$ is a constant, which means we do not have any *a priori* information about the image. From equation (6), for given measured data y , we have a function of x , the likelihood of x , defined by $p_Y(y|x)$. Then a ML estimation of the unknown image is defined as any maximizer x^* of $p_Y(y|x)$.

By taking the negative log-likelihood, one obtains, up to an additive constant,

$$f_0(x) = \sum_i ((Ax)_i - y_i \log(Ax)_i),$$

and the problem is to minimize this function $f_0(x)$ on the nonnegative orthant, because we have the constraint that the image x is nonnegative. In fact, we have

$$f(x) = D_{KL}(y, Ax) := \sum_i (y_i \log \frac{y_i}{(Ax)_i} + (Ax)_i - y_i) = f_0(x) + C,$$

where $D_{KL}(y, Ax)$ is the Kullback-Leibler (KL) divergence of Ax from y , and C is a constant independent of x . The KL divergence is considered as a data-fidelity function for Poisson data just like the standard least-square $\|Ax - y\|_2^2$ is the data-fidelity function for additive Gaussian noise. $D_{KL}(y, \cdot)$ is convex, nonnegative and coercive on the nonnegative orthant, so the minimizers exist and are global.

In order to find a minimizer of $f(x)$ with the constraint $x_j \geq 0$ for all j , we can solve the Karush-Kuhn-Tucker (KKT) conditions [30, 29],

$$\begin{aligned} \sum_i \left(A_{i,j} \left(1 - \frac{y_i}{(Ax)_i} \right) \right) - s_j &= 0, & j = 1, \dots, N, \\ s_j \geq 0, \quad x_j \geq 0, & & j = 1, \dots, N, \\ s^T x &= 0, \end{aligned}$$

where s_j is the Lagrange multiplier corresponding to the constraint $x_j \geq 0$. By the positivity of $\{x_j\}$, $\{s_j\}$ and the complementary slackness condition $s^T x = 0$, we have $s_j x_j = 0$ for every $j \in \{1, \dots, N\}$. Multiplying by x_j gives us

$$\sum_i \left(A_{i,j} \left(1 - \frac{y_i}{(Ax)_i} \right) \right) x_j = 0, \quad j = 1, \dots, N.$$

Therefore, we have the following iteration scheme

$$(8) \quad x_j^{k+1} = \frac{\sum_i \left(A_{i,j} \left(\frac{y_i}{(Ax^k)_i} \right) \right)}{\sum_i A_{i,j}} x_j^k.$$

This is the well-known EM iteration or Richardson-Lucy algorithm in image reconstruction, and an important property of it is that it preserves positivity. If x^k is positive, then x^{k+1} is also positive if A preserves positivity. It is also shown that for each iteration, $\sum_i (Ax)_i$ is fixed and equals $\sum_i y_i$. Since $\sum_i (Ax)_i = \sum_i \left(\sum_j A_{i,j} \right) x_j$, the minimizer has a weighted l_1 constraint.

Shepp and Vardi showed in [42] that this is equivalent to the EM algorithm proposed by Dempster, Laird and Rubin [13]. To make it clear, EM iteration

means the special EM method used in image reconstruction, while EM algorithm means the general EM algorithm for solving missing data problems.

3. EM-Type Algorithms for Poisson data. The method shown in the last section is also called maximum-likelihood expectation maximization (ML-EM) reconstruction, because it is a maximum likelihood approach without any Bayesian assumption on the images. If additional *a priori* information about the image is given, we have maximum a posteriori probability (MAP) approach [21, 32], which is the case with regularization term $J(x)$. Again we assume here that the detected data is corrupted by Poisson noise, and the regularization problem is

$$(9) \quad \begin{aligned} & \underset{x}{\text{minimize}} \quad E^P(x) := \beta J(x) + \sum_i ((Ax)_i - y_i \log(Ax)_i), \\ & \text{subject to } x_j \geq 0, \quad j = 1, \dots, N. \end{aligned}$$

This is still a convex constraint optimization problem if J is convex and we can find the optimal solution by solving the KKT conditions:

$$\begin{aligned} \beta \partial J(x)_j + \sum_i \left(A_{i,j} \left(1 - \frac{y_i}{(Ax)_i} \right) \right) - s_j &= 0, & j = 1, \dots, N, \\ s_j \geq 0, \quad x_j \geq 0, & & j = 1, \dots, N, \\ s^T x &= 0. \end{aligned}$$

Here s_j is the Lagrangian multiplier corresponding to the constraint $x_j \geq 0$. By the positivity of $\{x_j\}$, $\{s_j\}$ and the complementary slackness condition $s^T x = 0$, we have $s_j x_j = 0$ for every $j \in \{1, \dots, N\}$. Thus we obtain

$$\beta x_j \partial J(x)_j + \sum_i \left(A_{i,j} \left(1 - \frac{y_i}{(Ax)_i} \right) \right) x_j = 0, \quad j = 1, \dots, N,$$

or equivalently

$$\beta \frac{x_j}{\sum_i A_{i,j}} \partial J(x)_j + x_j - \frac{\sum_i \left(A_{i,j} \left(\frac{y_i}{(Ax)_i} \right) \right)}{\sum_i A_{i,j}} x_j = 0, \quad j = 1, \dots, N.$$

Notice that the last term on the left hand side is an EM step (8). After plugging the EM step into the equation, we obtain

$$(10) \quad \beta \frac{x_j}{\sum_i A_{i,j}} \partial J(x)_j + x_j - x_j^{EM} = 0, \quad j = 1, \dots, N,$$

which is the optimality condition for the following optimization problem

$$(11) \quad \underset{x}{\text{minimize}} \quad E_1^P(x, x^{EM}) := \beta J(x) + \sum_j \left(\sum_i A_{i,j} \right) (x_j - x_j^{EM} \log x_j).$$

Therefore we propose the general EM-Type algorithms in Algorithm 1. The initial guess x^0 can be any positive initial image, and ϵ , chosen for the stopping criteria, is a small constant. *Num_Iter* is the maximum number of iterations. If $J(x)$ is constant, the second step is just $x^k = x^{k-\frac{1}{2}}$ and this is exactly the ML-EM from the previous section. When $J(x)$ is not constant, we have to solve an optimization problem at each iteration. In general, the problem can not be solved analytically, and we have to use iterative methods to solve it. However, in practice, we do not have to solve it exactly to the steady state, just a few iterations are

sufficient. We will show that the algorithm will also converge without solving it exactly.

Algorithm 1 Proposed EM-Type algorithms.

Input: x^0, ϵ
for $k = 0 : Num_Iter$ **do**
 $x^{k+\frac{1}{2}} = EM(x^k)$ using (8),
 $x^{k+1} = \operatorname{argmin}_x E_1^P(x, x^{k+\frac{1}{2}})$ by solving (11),
if $\|x^{k+1} - x^k\| / \|x^k\| < \epsilon$ **then**
 Break,
end if
end for

There are several other methods for solving this problem using EM iteration, we mention two of them and show the differences with our method. One of the methods is a modified EM iteration [17, 35] which is

$$x_j^{k+1} = \frac{\sum_i \left(A_{i,j} \left(\frac{y_i}{(Ax^k)_i} \right) \right)}{\sum_i A_{i,j} + \beta \partial J(x^k)_j} x_j^k.$$

This can be considered as solving a modified version of (10):

$$\beta \frac{x_j}{\sum_i A_{i,j}} \partial J(x^k)_j + x_j - x_j^{EM} = 0, \quad j = 1, \dots, N.$$

The convergence is given only when β is small [17]. It is easy to notice that when β is large, x_j will be negative for some j 's and the projection onto the non-negative cone is needed. In this case, there is no convergence result.

Another algorithm is also a two stage method, and the difference is in the second stage [6, 8, 7]. Instead of solving (10) directly, which is a weighted Poisson denoising problem, a semi-implicit scheme is developed and it becomes a problem of weighted Gaussian denoising. The semi-implicit scheme is

$$\beta \frac{x_j^k}{\sum_i A_{i,j}} \partial J(x)_j + x_j - x_j^{EM} = 0, \quad j = 1, \dots, N.$$

However, there is no convergence for this algorithm and the convergence provided in [6] is for the following damped version:

$$(12) \quad \beta \frac{x_j^k}{\sum_i A_{i,j}} \partial J(x)_j + x_j - (wx_j^{EM} + (1-w)x_j^k) = 0, \quad j = 1, \dots, N.$$

There is a bound on w for each iteration to show the convergence of damped algorithm [6], and the bound is difficult to find. In the following, we will show the convergence of our algorithm without any assumptions on β and additional parameters. It will converge to a global minimum for convex $J(x)$ and a subsequence will converge to a local minimum for non-convex $J(x)$.

3.1. Equivalence to EM Algorithms with *a priori* Information. In this subsection, the EM-Type algorithms are shown to be equivalent to EM algorithms with *a priori* information. The EM algorithm is a general approach for maximizing a *posterior* distribution when some of the data is missing [13]. It is an iterative method which alternates between expectation (E) steps and maximization (M) steps. For image reconstruction, we assume that the missing data is the latent variables $\{z_{ij}\}$, describing the intensity of pixel (or voxel) j observed by detector i . It is introduced as a realization of a multi-valued random variable Z , and for each (i, j) pair, z_{ij} follows a Poisson distribution with expected value $A_{i,j}x_j$. Then the observed data is $y_i = \sum_j z_{ij}$, because the summation of two Poisson distributed random variables also follows a Poisson distribution, whose expected value is the summation of the two expected values.

The original E-step is to find the expectation of the log-likelihood given the present variables x^k :

$$Q(x|x^k) = E_{z|x^k, y} \log p(x, z|y).$$

Then, the M-step is to choose x^{k+1} to maximize the expected log-likelihood $Q(x|x^k)$ found in the E-step:

$$\begin{aligned} x^{k+1} &= \operatorname{argmax}_x E_{z|x^k, y} \log p(x, z|y) = \operatorname{argmax}_x E_{z|x^k, y} \log(p(y, z|x)p(x)) \\ (13) \quad &= \operatorname{argmax}_x E_{z|x^k, y} \sum_{ij} (z_{ij} \log(A_{i,j}x_j) - A_{i,j}x_j) - \beta J(x) \\ &= \operatorname{argmin}_x \sum_{ij} (A_{i,j}x_j - E_{z|x^k, y} z_{ij} \log(A_{i,j}x_j)) + \beta J(x). \end{aligned}$$

From (13), what we need before solving it is just $\{E_{z|x^k, y} z_{ij}\}$. Therefore we can compute the expectation of missing data $\{z_{ij}\}$ given present x^k and the condition $y_i = \sum_j z_{ij}$, denoting this as an E-step. Because for fixed i , $\{z_{ij}\}$ are Poisson variables with mean $\{A_{i,j}x_j^k\}$ and $\sum_j z_{ij} = y_i$, the conditional distribution of z_{ij} is the binomial distribution $\left(y_i, \frac{A_{i,j}x_j^k}{(Ax^k)_i}\right)$. Thus we can find the expectation of z_{ij} with all these conditions by the following E-step

$$(14) \quad z_{ij}^{k+1} = E_{z|x^k, y} z_{ij} = \frac{A_{i,j}x_j^k y_i}{(Ax^k)_i}.$$

After obtaining the expectation for all z_{ij} , we can solve the M-step (13).

We will show that EM-Type algorithms are exactly the described EM algorithms with *a priori* information. Recalling the definition of x^{EM} , we have

$$x_j^{EM} = \frac{\sum_i z_{ij}^{k+1}}{\sum_i A_{i,j}}.$$

Therefore, the M-step is equivalent to

$$\begin{aligned} x^{k+1} &= \operatorname{argmin}_x \sum_{ij} (A_{i,j}x_j - z_{ij}^{k+1} \log(A_{i,j}x_j)) + \beta J(x) \\ &= \operatorname{argmin}_x \sum_j \left(\sum_i A_{i,j} \right) (x_j - x_j^{EM} \log(x_j)) + \beta J(x). \end{aligned}$$

We have shown that EM-Type algorithms are EM algorithms with *a priori* information. The convergence of EM-Type algorithms is shown in the next subsection from the equivalence of EM-Type algorithms with alternating minimization methods for equivalent problems.

3.2. EM-Type Algorithms are Alternating Minimization Methods. In this section, we will show that these algorithms can also be derived from alternating minimization methods of other problems with variables x and z . The new optimization problems are

$$(15) \quad \begin{aligned} & \underset{x, z}{\text{minimize}} \quad E^P(x, z) := \sum_{ij} \left(z_{ij} \log \frac{z_{ij}}{A_{i,j}x_j} + A_{i,j}x_j - z_{ij} \right) + \beta J(x), \\ & \text{subject to} \quad \sum_j z_{ij} = y_i, \quad i = 1, \dots, M, \\ & \quad \quad \quad z_{ij} \geq 0, \quad i = 1, \dots, M, j = 1, \dots, N. \end{aligned}$$

Here E^P is used again to define the new function. $E^P(\cdot)$ means the negative log-likelihood function of x , while $E^P(\cdot, \cdot)$ means the new function of x and z defined in new optimization problems. $x \geq 0$ is implicitly included in the formula.

Having initial guess x^0, z^0 of x and z , the iteration for $k = 0, 1, \dots$ is as follows:

$$\begin{aligned} z^{k+1} &= \underset{z}{\operatorname{argmin}} \quad E^P(x^k, z), \quad \text{subject to} \quad \sum_j z_{ij} = y_i, \\ x^{k+1} &= \underset{x}{\operatorname{argmin}} \quad E^P(x, z^{k+1}). \end{aligned}$$

Firstly, in order to obtain z^{k+1} , we fix $x = x^k$ and easily derive

$$(16) \quad z_{ij}^{k+1} = \frac{A_{i,j}x_j^k y_i}{(Ax^k)_i}.$$

After finding z^{k+1} , let $z = z^{k+1}$ fixed and update x , then we have

$$\begin{aligned} x^{k+1} &= \underset{x}{\operatorname{argmin}} \quad \sum_{ij} \left(A_{i,j}x_j + z_{ij}^{k+1} \log \frac{z_{ij}^{k+1}}{A_{i,j}x_j} \right) + \beta J(x) \\ &= \underset{x}{\operatorname{argmin}} \quad \sum_{ij} (A_{i,j}x_j - z_{ij}^{k+1} \log(A_{i,j}x_j)) + \beta J(x), \end{aligned}$$

which is the M-Step (13) in section 3.1. Thus EM-Type algorithms for (9) are alternating minimization methods for problem (15). The equivalence of problems (9) and (15) is provided in the following theorem.

Theorem 3.1. *If (x^*, z^*) is a local minimum of problem (15), then x^* is also a local minimum of (9). If x^* is a local minimum of (9), then we can find z^* from (16) and (x^*, z^*) is a local minimum of problem (15).*

Proof. If (x^*, z^*) is a local minimum of problem (15), we can find $\delta > 0$ so that for all (x, z) such that $\|(x - x^*, z - z^*)\| < \delta$ and $\sum_j z_{ij} = y_i, z_{ij} \geq 0$, the expression $E^P(x, z) \geq E^P(x^*, z^*)$ holds. Let us assume that x^* is not a local minimum of problem (9), then there exist a sequence x^k such that $\lim_{k \rightarrow \infty} x^k = x^*$ and $E^P(x^k) < E^P(x^*)$. Let z^{k+1} be chosen from (16) for corresponding x^k and we have $E^P(x^k, z^{k+1}) = E^P(x^k) < E^P(x^*) = E^P(x^*, z^*)$, and $\lim_{k \rightarrow \infty} z^k = z^*$. This is a contradiction with that (x^*, z^*) is a local minimum of (15).

In another way, if x^* is a local minimum of (9), there exist a $\delta > 0$ such that for all x such that $\|x - x^*\| \leq \delta$ and $x \geq 0$, we have $E^P(x) \geq E^P(x^*)$. If we choose (x, z) such that $\|(x - x^*, z - z^*)\| < \delta$ and $\sum_j z_{ij} = y_i$, then we have $E^P(x, z) \geq E^P(x) \geq E^P(x^*) = E^P(x^*, z^*)$. Thus (x^*, z^*) is a local minimum of problem (15). \square

From the equivalence of EM-Type algorithm with alternating minimization methods, we know that the function value is decreasing. Next, we will show that if $J(x)$ is convex, (x^k, z^k) will converge to a global minimum of problem (15), and if $J(x)$ is non-convex, a subsequence of $\{(x^k, z^k)\}$ will converge to a local minimum of problem (15). Before that, we show two lemmas which are used in the proof of the convergence results.

For $s > 0, t \geq 0$, let us define $d^P(s, t) = t \log \frac{t}{s} + s - t$, assuming that $0 \log 0 = 0$ and $\log 0 = -\infty$. In addition \mathcal{Z} is defined as $\mathcal{Z} := \{z : z_{ij} \geq 0, \sum_j z_{ij} = y_i\}$. We will have the following lemmas:

Lemma 3.2. For all $p > 0, s > 0, t \geq 0$, we have

$$(17) \quad d^P(s, t) - d^P(p, t) + d^P(p, s) = \log \frac{s}{p}(s - t).$$

Proof.

$$\begin{aligned} & d^P(s, t) - d^P(p, t) + d^P(p, s) \\ &= t \log \frac{t}{s} + s - t - \left(t \log \frac{t}{p} + p - t \right) + \left(s \log \frac{s}{p} + p - s \right) \\ &= \log \frac{s}{p}(s - t). \end{aligned}$$

\square

Lemma 3.3. For all $p > 0, q \geq 0, s > 0, t \geq 0$, we have

$$(18) \quad d^P(s, t) - d^P(p, t) + d^P(q, t) - \frac{p - q}{p}(s - p) \geq 0.$$

Proof. Since $d^P(s, t)$ is convex for $\{(s, t) : s > 0, t \geq 0\}$, we have

$$\begin{aligned} & d_{d^P}((s, t), (p, q)) \\ &:= d^P(s, t) - d^P(p, q) - \nabla_p d^P(p, q)(s - p) - \nabla_q d^P(p, q)(t - q) \\ (19) \quad &= d^P(s, t) - d^P(p, q) - \left(-\frac{q}{p} + 1 \right)(s - p) - \log \frac{q}{p}(t - q) \\ &= d^P(s, t) - d^P(p, q) - \frac{p - q}{p}(s - p) - d^P(p, t) + d^P(q, t) + d^P(p, q). \end{aligned}$$

The last equality comes from Lemma 3.2. Since $d^P(s, t)$ is convex, the result follows because $d_{d^P}((s, t), (p, q)) \geq 0$. \square

In order to show the convergence for general $J(x)$, we will need the following assumption on $J(x)$.

Assumption 1. $J(x)$ is lower semicontinuity, bounded below, and there exists $\delta > 0$ for any given $\bar{z} \in \mathcal{Z}$ and any local minimum point \bar{x} of $E^P(x, \bar{z})$, such that for all x satisfying $\|x - \bar{x}\| \leq \delta$, the following inequality holds:

$$\beta J(x) \geq \beta J(\bar{x}) - \sum_{ij} \frac{A_{i,j} \bar{x}_j - \bar{z}_{ij}}{A_{i,j} \bar{x}_j} A_{i,j} (x_j - \bar{x}_j) - \sum_{ij} d_{d^P}((A_{i,j} x_j, z_{ij}), (A_{i,j} \bar{x}_j, \bar{z}_{ij})),$$

where $z \in \mathcal{Z}$.

Remark 1. For convex $J(x)$, we have a stronger statement

$$(20) \quad \beta J(x) \geq \beta J(\bar{x}) - \langle \partial_x J(\bar{x}), x - \bar{x} \rangle = \beta J(\bar{x}) - \sum_{ij} \frac{A_{i,j} \bar{x}_j - \bar{z}_{ij}}{A_{i,j} \bar{x}_j} A_{i,j} (x_j - \bar{x}_j).$$

The equality comes from the optimality condition for \bar{x} .

With Assumption 1, we have the following theorem which will provide an alternative way to determine the local optimality.

Theorem 3.4. *If there exists (\bar{x}, \bar{z}) such that $\bar{z} = \operatorname{argmin}_z E^P(\bar{x}, z)$ and \bar{x} is a local minimum point of $E^P(x, \bar{z})$. Then (\bar{x}, \bar{z}) is a local minimum of problem (15).*

Proof. Since \bar{x} is a local minimum point of $E^P(x, \bar{z})$, there exists $\delta > 0$ such that for all x satisfying $\|x - \bar{x}\| \leq \delta$, we have, from Assumption 1, that

$$\begin{aligned} & \beta J(x) \\ & \geq \beta J(\bar{x}) - \sum_{ij} \frac{A_{i,j} \bar{x}_j - \bar{z}_{ij}}{A_{i,j} \bar{x}_j} A_{i,j} (x_j - \bar{x}_j) - \sum_{ij} d_{d^P}((A_{i,j} x_j, z_{ij}), (A_{i,j} \bar{x}_j, \bar{z}_{ij})) \\ & = \beta J(\bar{x}) + \sum_{ij} d^P(A_{i,j} \bar{x}_j, \bar{z}_{ij}) - \sum_{ij} d^P(A_{i,j} x_j, z_{ij}) + \sum_{ij} \log \frac{\bar{z}_{ij}}{A_{i,j} \bar{x}_j} (z_{ij} - \bar{z}_{ij}) \\ & = \beta J(\bar{x}) + \sum_{ij} d^P(A_{i,j} \bar{x}_j, \bar{z}_{ij}) - \sum_{ij} d^P(A_{i,j} x_j, z_{ij}), \end{aligned}$$

for all $z \in \mathcal{Z}$. The first equality comes from (19), and the second equality holds because for any $z \in \mathcal{Z}$, we have

$$\sum_{ij} \log \frac{\bar{z}_{ij}}{A_{i,j} \bar{x}_j} (z_{ij} - \bar{z}_{ij}) = \sum_i \log \frac{y_i}{(A\bar{x})_i} \sum_j (z_{ij} - \bar{z}_{ij}) = 0.$$

Therefore, we have

$$E^P(x, z) = \beta J(x) + \sum_{ij} d^P(A_{i,j} x_j, z_{ij}) \geq \beta J(\bar{x}) + \sum_{ij} d^P(A_{i,j} \bar{x}_j, \bar{z}_{ij}) = E^P(\bar{x}, \bar{z})$$

for all x satisfying $\|x - \bar{x}\| \leq \delta$ and $z \in \mathcal{Z}$, which means that (\bar{x}, \bar{z}) is a local minimum of problem (15). \square

Remark 2. In fact, (\bar{x}, \bar{z}) being a local minimum of problem (15) requires that there exist a constant $\delta > 0$ such that for all x satisfying $\|x - \bar{x}\| \leq \delta$, the following inequality holds:

$$\beta J(x) \geq \beta J(\bar{x}) - \sum_{ij} \frac{A_{i,j} \bar{x}_j - \bar{z}_{ij}}{A_{i,j} \bar{x}_j} A_{i,j} (x_j - \bar{x}_j) - \sum_{ij} d_{d^P}((A_{i,j} x_j, z_{ij}), (A_{i,j} \bar{x}_j, \bar{z}_{ij})),$$

where $z \in \mathcal{Z}$. Therefore, it is reasonable to make the assumptions on $J(x)$.

Theorem 3.5. *The algorithm will converge to a global minimum of problem (15) for convex $J(x)$, and there exists a subsequence which converges to a local minimum of problem (15) for non-convex $J(x)$.*

Proof. From previous theorem, we have to show that it converges to (x^*, z^*) with $z^* = \operatorname{argmin}_z E^P(x^*, z)$ and x^* being a local minimum point of $E^P(x, z^*)$.

$$\begin{aligned}
& \sum_{ij} d^P(A_{i,j}x_j, z_{ij}) + \sum_{ij} d^P(z_{ij}^k, z_{ij}) \\
\stackrel{(18)}{=} & \sum_{ij} d^P(A_{i,j}x_j^k, z_{ij}) + \sum_{ij} \frac{z_{ij}^k - A_{i,j}x_j^k}{A_{i,j}x_j^k} (A_{i,j}x_j^k - A_{i,j}x_j) \\
& + d_{d^P}((A_{i,j}x_j, z_{ij}), (A_{i,j}x_j^k, z_{ij}^k)) \\
\stackrel{(17)}{=} & \sum_{ij} d^P(z_{ij}^{k+1}, z_{ij}) + \sum_{ij} d^P(A_{i,j}x_j^k, z_{ij}^{k+1}) + \sum_{ij} \log \frac{A_{i,j}x_j^k}{z_{ij}^{k+1}} (z_{ij}^{k+1} - z_{ij}) \\
(21) \quad & + \sum_{ij} \frac{z_{ij}^k - A_{i,j}x_j^k}{A_{i,j}x_j^k} (A_{i,j}x_j^k - A_{i,j}x_j) + d_{d^P}((A_{i,j}x_j, z_{ij}), (A_{i,j}x_j^k, z_{ij}^k)) \\
= & \sum_{ij} d^P(z_{ij}^{k+1}, z_{ij}) + \sum_{ij} d^P(A_{i,j}x_j^k, z_{ij}^{k+1}) + \beta J(x^k) - \beta J(x) \\
& + \sum_{ij} \frac{z_{ij}^k - A_{i,j}x_j^k}{A_{i,j}x_j^k} (A_{i,j}x_j^k - A_{i,j}x_j) + \beta J(x) - \beta J(x^k) \\
& + d_{d^P}((A_{i,j}x_j, z_{ij}), (A_{i,j}x_j^k, z_{ij}^k)) \\
\geq & \sum_{ij} d^P(z_{ij}^{k+1}, z_{ij}) + \sum_{ij} d^P(A_{i,j}x_j^k, z_{ij}^{k+1}) + \beta J(x^k) - \beta J(x).
\end{aligned}$$

The last inequality comes from Assumption 1, and it holds only for x close to x^k . Therefore, we have $\sum_{ij} d^P(z_{ij}^k, z_{ij}) - \sum_{ij} d^P(z_{ij}^{k+1}, z_{ij}) \geq E^P(x^k, z^{k+1}) - E^P(x, z)$. Since $J(x)$ is bounded below and $d^P(+\infty, z) = +\infty$, x^k is bounded and there exists a subsequence denoted by $\{x^{o_k}\}$ converging to x^* . Let $z^* = \operatorname{argmin}_z E^P(x^*, z)$. Then we have $\lim_{k \rightarrow \infty} z^{o_k+1} = z^*$ and $E^P(x^*, z^*) \leq \lim_{k \rightarrow \infty} E^P(x^{o_k}, z^{o_k+1}) \leq E^P(x^{o_k}, z^{o_k+1})$. Since $E^P(x^k, z^{k+1})$ is decreasing, we have $E^P(x^*, z^*) < E^P(x^k, z^k)$ and $\lim_{k \rightarrow \infty} E^P(x^k, z^k) = E^P(x^*, z^*)$. From (21), $d^P(z_{ij}^k, z_{ij}^*)$ is monotone decreasing and $\{z^{o_k}\}$ is bounded, there exists a subsequence converging to \bar{z} , and we still denote it by $\{z^{o_k}\}$ for simplicity. Since $E^P(x^*, \bar{z}) \leq \lim_{k \rightarrow \infty} E^P(x^{o_k}, z^{o_k}) = E^P(x^*, z^*)$, we have $\bar{z} = z^*$. Next, we will show that x^* is a local minimum point of $E^P(x, z^*)$. From Assumption 1 we have

$$\begin{aligned}
\beta J(x) & \geq \beta J(x^{o_k}) - \sum_{ij} \frac{A_{i,j}x_j^{o_k} - z_{ij}^{o_k}}{A_{i,j}x_j^{o_k}} A_{i,j}(x_j - x_j^{o_k}) \\
& - \sum_{ij} d_{d^P}((A_{i,j}x_j, z_{ij}^{o_k}), (A_{i,j}x_j^{o_k}, z_{ij}^{o_k})) \\
& = \beta J(x^{o_k}) - \sum_{ij} d^P(A_{i,j}x_j, z_{ij}^{o_k}) + \sum_{ij} d^P(A_{i,j}x_j^{o_k}, z_{ij}^{o_k}).
\end{aligned}$$

Let $k \rightarrow \infty$ we have $\beta J(x) + \sum_{ij} d^P(A_{i,j}x_j, z_{ij}^*) \geq \beta J(x^*) + \sum_{ij} d^P(A_{i,j}x_j^*, z_{ij}^*)$, which means that x^* is a local minimum of $E^P(x, z^*)$ and (x^*, z^*) is a local minimum.

If $J(x)$ is convex, we can choose (x, z) in (21) to be a global minimum (\bar{x}, \bar{z}) , and we have that $d^P(z_{ij}^k, \bar{z}_{ij})$ is monotone decreasing. Thus z is bounded and for each convergent subsequence of z^{o_k} with $\lim_{k \rightarrow \infty} z^{o_k} = \bar{z}$, we can find x^{o_k} and

$\lim_{k \rightarrow \infty} x^{o_k} = \tilde{x}$, with \tilde{x} being a minimum point of $E^P(x, \tilde{z})$. We have $E^P(\tilde{x}, \tilde{z}) \leq \lim_{k \rightarrow \infty} E^P(x^{o_k}, z^{o_k}) = E^P(\tilde{x}, \tilde{z})$. Thus (\tilde{x}, \tilde{z}) is also a global minimum. Therefore, $\lim_{k \rightarrow \infty} z^k = z^*$, and let $x^* = \operatorname{argmin}_x E^P(x, z^*)$ satisfying $\lim_{k \rightarrow \infty} x^k = x^*$. The resulting (x^*, z^*) is a global minimum of problem (15). \square

Remark 3. Even if the second step is not solved exactly, we have $E^P(x^{k+1}) < E^P(x^k)$ if $E^P(x^{k+1}, z^{k+1}) < E^P(x^k, z^{k+1})$. In fact it is impossible to solve the second step exactly in many cases, and we have to approximately solve it using iterative methods.

Remark 4. The relations between these algorithms are shown in Figure 1. EM-Type algorithm is a special EM-algorithm with *a priori* information, and EM iteration is a special case of EM-Type algorithm without $J(x)$.

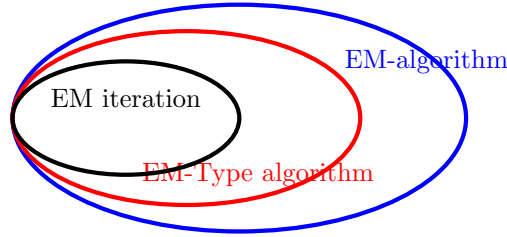


FIGURE 1. Relations between the algorithms.

4. Simultaneous Algebraic Reconstruction Technique (SART) is EM. Among all the iterative reconstruction algorithms, there are two important classes. One is EM from statistical assumptions mentioned above, and the other is algebraic reconstruction technique (ART)-Type algorithms [16, 20]. Simultaneous algebraic reconstruction technique (SART) [3, 4], as a refinement of ART, is used widely [5, 36, 52] and the convergence analysis of SART is well studied by Jiang and Wang [25, 24], Wang and Zheng [47], Censor and Elfving [9], and Yan [50]. In this section, we will show that SART is also an EM algorithm, building the connection between these two classes.

From the convergence analysis of SART in [50], SART is also an algorithm for solving a maximum likelihood problem

$$(22) \quad p_Y(y|x) = \prod_i \frac{1}{\sqrt{2\pi w_i}} e^{-\frac{(y_i - (Ax)_i)^2}{2w_i}},$$

where $w_i = \sum_j A_{i,j}$. Similarly, we assume that the missing data $\{z_{ij}\}$ follow normal distributions with expected values $\{A_{i,j}x_j\}$ and variances $\{A_{i,j}\}$ respectively. The original E-step is to find the expectation of the log-likelihood given the present variables x^k and the constraints $y_i = \sum_j z_{ij}$. It is easy to derive that under the constraints, $\{z_{ij}\}$ are still realizations of normally distributed random variables, but with different expected values $\{A_{i,j}x_j + \frac{A_{i,j}(y_i - (Ax)_i)}{w_i}\}$ and variances $\{\frac{A_{i,j}(w_i - A_{i,j})}{w_i}\}$ respectively.

In this section, we consider the special case without regularization function, i.e., there is no *a priori* information about the image to be reconstructed. The M-step

is to maximize the expected value of the log-likelihood function,

$$\begin{aligned} E_{z|x^k, y} \log p(y, z|x) &= -E_{z|x^k, y} \sum_{ij} \frac{(z_{ij} - A_{i,j}x_j)^2}{2A_{i,j}} + C \\ &= -\sum_{ij} \frac{(E_{z|x^k, y} z_{ij} - A_{i,j}x_j)^2}{2A_{i,j}} + C, \end{aligned}$$

where C is a constant independent of x and z . Therefore, for the E-step we have to just find the expected value of z_{ij} given x^k and the constraints, which is

$$z_{ij}^{k+1} = A_{i,j}x_j^k + \frac{A_{i,j}(y_i - (Ax^k)_i)}{w_i}.$$

For the M-step, we find x^{k+1} by maximizing $p(y, z^{k+1}|x)$ with respect to x , which has an analytical solution

$$(23) \quad x_j^{k+1} = \frac{\sum_i z_{ij}^{k+1}}{\sum_i A_{i,j}} = x_j^k + \frac{1}{\sum_i A_{i,j}} \sum_i \frac{A_{i,j}(y_i - (Ax^k)_i)}{w_i}.$$

This is the original SART algorithm proposed by Andersen [3].

From the convergence analysis of SART in [50], the result of SART depends on the initialization x^0 for both noiseless and noisy cases when A is underdetermined.

Remark 5. SART is just one example of Landweber-like schemes for solving systems of linear equations. By changing the variance of y_i and z_{ij} , different schemes can be proposed. For other Landweber-like schemes such as component averaging in [9, 10], they can also be derived from the EM algorithm similarly by choosing different variances. Furthermore, new schemes can be derived by choosing different variances.

5. EM-Type Algorithms for Gaussian Noise. It is shown in the last section that SART is an EM algorithm based on weighted Gaussian assumption for the problem without regularization. Without regularization, the original problem is ill-posed, and the result will depend on the initialization x^0 . In this section, we will consider the regularized problem

$$(24) \quad \underset{x}{\text{minimize}} \quad E^G(x) := \beta J(x) + \sum_i \frac{((Ax)_i - y_i)^2}{2w_i},$$

and derive EM-Type algorithms with Gaussian noise assumption for solving it. The E-step is the same as in the case without regularization,

$$(25) \quad z_{ij}^{k+1} = A_{i,j}x_j^k + \frac{A_{i,j}(y_i - (Ax^k)_i)}{w_i}.$$

However, the M-step is different because we have *a priori* information on the image x to be reconstructed. The new M-step is to solve the following optimization problems

$$(26) \quad \underset{x}{\text{minimize}} \quad \sum_{ij} \frac{(z_{ij}^{k+1} - A_{i,j}x_j)^2}{2A_{i,j}} + \beta J(x),$$

which is equivalent to

$$\underset{x}{\text{minimize}} \quad \frac{1}{2} \sum_j \left(\sum_i A_{i,j} \right) \left(x_j - \frac{\sum_i z_{ij}^{k+1}}{\sum_i A_{i,j}} \right)^2 + \beta J(x).$$

From the SART iteration (23) in the last section, we can define

$$(27) \quad x^{SART} = x_j^k + \frac{1}{\sum_i A_{i,j}} \sum_i \frac{A_{i,j} (y_i - (Ax^k)_i)}{w_i}.$$

and have

$$(28) \quad x^{k+1} = \underset{x}{\text{argmin}} \quad E_1^G(x, x^{SART}) := \frac{1}{2} \sum_j \left(\sum_i A_{i,j} \right) (x_j - x_j^{SART})^2 + \beta J(x).$$

Therefore, the proposed EM-Type algorithms for image reconstruction with Gaussian noise are as follows.

Algorithm 2 Proposed EM-Type algorithms for Gaussian noise.

Input: $x^0, \epsilon,$
for $k = 0 : Num_Iter$ **do**
 $x^{k+\frac{1}{2}} = SART(x^k)$ using (27)
 $x^{k+1} = \underset{x}{\text{argmin}} \quad E_1^G(x, x^{k+\frac{1}{2}})$ by solving (28)
if $\|x^{k+1} - x^k\| / \|x^k\| < \epsilon$ **then**
 Break,
end if
end for

The initial guess x^0 can be any initial image and ϵ , chosen for the stopping criteria, is very small. Num_Iter is the maximum number of iterations. When $J(x)$ is not constant, we have to solve an optimization problem for each iteration. The convergence analysis of these algorithms can be shown similarly as for the case with Poisson noise, which is described in the following subsection.

5.1. EM-Type Algorithms are Alternating Minimization Methods. Same as the algorithms for Poisson data, the algorithms can also be derived from an alternating minimization method of other problems with variables x and z . The new problems are

$$(29) \quad \begin{aligned} &\underset{x,z}{\text{minimize}} \quad E^G(x, z) := \sum_{ij} \frac{(z_{ij} - A_{i,j}x_j)^2}{2A_{i,j}} + \beta J(x), \\ &\text{subject to} \quad \sum_j z_{ij} = y_i, \quad i = 1, \dots, M. \end{aligned}$$

Here E^G is used again to define the new function. $E^G(\cdot)$ means the negative log-likelihood function of x , while $E^G(\cdot, \cdot)$ means the new function of x and z defined in new optimization problems. The iteration is as follows:

$$\begin{aligned} z^{k+1} &= \underset{z}{\text{argmin}} \quad E(x^k, z), \quad \text{subject to} \quad \sum_j z_{ij} = y_i. \\ x^{k+1} &= \underset{x}{\text{argmin}} \quad E(x, z^{k+1}). \end{aligned}$$

First, let us fix $x = x^k$ and update z . It is easy to derive

$$z_{ij}^{k+1} = A_{i,j}x_j^k + \frac{A_{i,j}}{w_i} (y_i - (Ax^k)_i).$$

Then, by fixing $z = z^{k+1}$ and updating x , we have

$$\begin{aligned} x^{k+1} &= \operatorname{argmin}_x \sum_{ij} \frac{(z_{ij} - A_{i,j}x_j)^2}{2A_{i,j}} + \beta J(x) \\ &= \operatorname{argmin}_x \frac{1}{2} \sum_j \left(\sum_i A_{i,j} \right) \left(x_j - \frac{\sum_i z_{ij}}{2 \sum_i A_{i,j}} \right)^2 + \beta J(x). \end{aligned}$$

If $J(x)$ is convex, problem (29) is convex, and we can find the minimizer with respect to z for fixed x first and obtain a function of x as follows,

$$\sum_i \frac{((Ax)_i - y_i)^2}{2w_i} + \beta J(x),$$

which is also convex and equals $E^G(x)$. Therefore EM-Type algorithms will converge to the solution of (24).

The convergence can be shown similarly with small changes in the definition of $d^P(s, t)$, for weighed Gaussian noise, we can define $d^G(s, t) = \frac{1}{2}(s - t)^2$. In addition \mathcal{Z} is defined as $\mathcal{Z} := \{z : z_{ij} = 0 \text{ if } A_{i,j} = 0, \sum_j z_{ij} = y_i\}$. The corresponding lemmas are provided without proof.

Lemma 5.1. *For all $p, s, t \in \mathbf{R}$, we have*

$$d^G(s, t) - d^G(p, t) + d^G(p, s) = (s - p)(s - t).$$

Lemma 5.2. *For all $p, q, s, t \in \mathbf{R}$, we have*

$$d_{d^G}((s, t), (p, q)) := d^G(s, t) - d^G(p, t) + d^G(q, t) - (p - q)(s - p) \geq 0.$$

As for the Poisson noise case, we have the following assumption on $J(x)$.

Assumption 2. $J(x)$ is lower semicontinuity, bounded below, and there exists $\delta > 0$ for any given $\bar{z} \in \mathcal{Z}$ and any local minimum point \bar{x} of $E^G(x, \bar{z})$, such that for all x satisfying $\|x - \bar{x}\| \leq \delta$, the following inequality holds:

$$\begin{aligned} \beta J(x) &\geq \beta J(\bar{x}) - \sum_{ij} \frac{1}{A_{i,j}} (A_{i,j}\bar{x}_j - \bar{z}_{ij}) A_{i,j} (x_j - \bar{x}_j) \\ &\quad - \sum_{ij} \frac{1}{A_{i,j}} d_{d^G}((A_{i,j}x_j, z_{ij}), (A_{i,j}\bar{x}_j, \bar{z}_{ij})), \end{aligned}$$

where $z \in \mathcal{Z}$.

Remark 6. For convex $J(x)$, we have a stronger statement

$$\beta J(x) \geq \beta J(\bar{x}) - \langle \partial_x J(\bar{x}), x - \bar{x} \rangle = \beta J(\bar{x}) - \sum_{ij} \frac{1}{A_{i,j}} (A_{i,j}\bar{x}_j - \bar{z}_{ij}) A_{i,j} (x_j - \bar{x}_j).$$

The equality comes from the optimality condition for \bar{x} .

With this assumption, we have the following theorem which will provide an alternative way to determine the local optimality. The proof is similar to the case of Poisson noise and we omit it here.

Theorem 5.3. *If there exists (\bar{x}, \bar{z}) such that $\bar{z} = \operatorname{argmin}_z E^G(\bar{x}, z)$ and \bar{x} is a local minimum point of $E^G(x, \bar{z})$. Then (\bar{x}, \bar{z}) is a local minimum of problem (29).*

The convergence result is similar and we put it here without proof.

Theorem 5.4. *The algorithm will converge to a global minimum of problem (29) for convex $J(x)$, and there exists a subsequence which converges to a local minimum of problem (29) for non-convex $J(x)$.*

5.2. Relaxation. In practice, some authors use a relaxation of SART reconstruction, which is

$$x_j^{k+1} = x_j^k + \frac{w}{\sum_i A_{i,j}} \sum_i \frac{A_{i,j}(y_i - (Ax^k)_i)}{w_i},$$

with a relaxant coefficient w . The convergence of this relaxation is shown in [25, 24, 50] for any $w \in (0, 2)$. Inspired by this strategy, we have a relaxation of the EM-Type algorithms for image reconstruction with Gaussian noise. The EM-step is the relaxed SART with relaxant coefficient w :

$$x_j^{k+\frac{1}{2}} = x_j^k + \frac{w}{\sum_i A_{i,j}} \sum_i \frac{A_{i,j}(y_i - (Ax^k)_i)}{w_i}.$$

The corresponding regularization step is

$$x^{k+1} = \operatorname{argmin}_x \frac{1}{2} \sum_j \left(\sum_i A_{i,j} \right) (x_j - x_j^{k+\frac{1}{2}})^2 + w\beta J(x).$$

When $w = 1$, we have already discussed the convergence in the previous subsections by EM algorithms with *a priori* information and alternating minimization methods. For $w \neq 1$, we will show the convergence of the relaxed EM-Type algorithms for $w \in (0, 1)$ by alternating minimization methods.

We will show that the relaxed EM-Type algorithms are equivalent to solving the unconstrained problems

$$(30) \quad \operatorname{minimize}_{x,z} E_R^G(x, z) := \sum_{ij} \frac{(z_{ij} - A_{i,j}x_j)^2}{2A_{i,j}} + \gamma \sum_i \frac{(\sum_j z_{ij} - y_i)^2}{2w_i} + w\beta J(x),$$

where $\gamma = \frac{w}{1-w}$, by alternating minimization between x and z . First, fix $x = x^k$, we can solve the problem of z only, and the analytical solution is

$$(31) \quad z_{ij}^{k+1} = A_{i,j}x_j^k + \frac{\gamma}{1+\gamma} \frac{A_{i,j}}{w_i} (y_i - (Ax^k)_i) = A_{i,j}x_j^k + w \frac{A_{i,j}}{w_i} (y_i - (Ax^k)_i).$$

Then let $z = z^{k+1}$ fixed, and we can find x^{k+1} by solving

$$\begin{aligned} \operatorname{minimize}_x \quad & \sum_{ij} \frac{(z_{ij} - A_{i,j}x_j)^2}{2A_{i,j}} + w\beta J(x) \\ & = \frac{1}{2} \sum_j \left(\sum_i A_{i,j} \right) \left(x_j - \frac{\sum_i z_{ij}}{\sum_i A_{i,j}} \right)^2 + w\beta J(x) + C, \end{aligned}$$

where C is a constant independent of x . Having z^{k+1} from (31), we can calculate

$$\frac{\sum_i z_{ij}^{k+1}}{\sum_i A_{i,j}} = x_j^k + \frac{w}{\sum_i A_{i,j}} \sum_i \frac{A_{i,j}(y_i - (Ax^k)_i)}{w_i} = x_j^{k+\frac{1}{2}}.$$

Therefore this relaxed EM-Type algorithm is an alternating minimization method. We will show next that the result of this relaxed EM-Type algorithm is the solution to (24).

Because the objective function $E_R^G(x, z)$ in (30) is convex if $J(x)$ is convex, we can first minimize the function with respect to z with x fixed. Then the problem becomes

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \frac{\gamma}{1 + \gamma} \sum_i \frac{((Ax)_i - y_i)^2}{2w_i} + w\beta J(x) \\ & = w \sum_i \frac{((Ax)_i - y_i)^2}{2w_i} + w\beta J(x). \end{aligned}$$

We have shown in this subsection that the relaxed EM-Type algorithm will also converge to the solution of the original problem (24) when $\alpha \in (0, 1]$.

6. Numerical Experiments. As mentioned in the introduction, the main focus of this paper is the convergence result of EM-Type algorithms, and some of the numerical experiments in this section have been included in [49, 51]. The newly added experiment is using a non-convex $J(x)$. In addition, this algorithm has been implemented in different hardwares and the speedups can be found in [11].

In this section, several numerical experiments are provided to show the efficiency of EM-Type algorithms. Though these EM-Type algorithms can be used in many applications, we choose Computed Tomography (CT) image reconstruction as our application in this work. CT is a medical imaging method which utilizes X-ray equipment to produce a two dimensional (or three dimensional) image of the inside of an object from a large series of one dimensional (or two dimensional) X-ray images taken along a single axis of rotation [20]. In CT reconstruction, the operator A is the discrete Radon transform, and the discrete version of A is constructed by Siddon's algorithm [43, 53]. The problem is to reconstruct the image from the measurements, which is equivalent to solve $Ax = b$. Poisson noise is assumed [45] and two regularizations are used: the total variation (TV), and an approximation to a non-convex Mumford-Shah TV-like version.

6.1. CT Reconstruction using EM+TV (2D). At first, we illustrate one method (EM+TV) on a simple synthetic object (two dimensional Shepp-Logan phantom), see Figure 2.

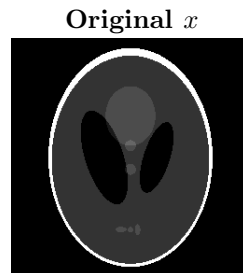


FIGURE 2. 2D Shepp-Logan phantom

The most common method used in commercial CT (computerized tomography) is the filtered back projection (FBP), which can be implemented in a straight forward

way and it can be computed rapidly. However, FBP has limitations due to the presence of streak artifacts and the noise enhancement, which is inherent in the reconstruction. Furthermore, in order to obtain an accurate image, many views are taken. Algorithms that can perform accurate image reconstruction from fewer views are very important in reducing patient radiation dose and speeding up scans. Optimization based methods, including EM+TV, can reconstruct an image from fewer views, but require more computing time. However, with the development of graphics processing units (GPUs), the computing time has reduced greatly and this kind of techniques becomes more and more important.

In the following experiment, we will compare the reconstruction results obtained by EM+TV with those obtained by filtered back projection. Firstly we obtain the measurements using Siddon's algorithm. We consider both the noise-free and noisy cases. For the FBP method, we present results using 36 views (every 10 degrees; for each view there are 301 measurements), 180 views, and 360 views. In order to show that we can reduce the number of views by using EM+TV, we only use 36 views for the proposed method. The results are shown in Figure 3. We notice the much improved results obtained with EM+TV using only 36 views, by comparison with FBP using 36, 180 or even 360 views. To solve the regularization step, we use semi-implicit scheme for some iterations [49].

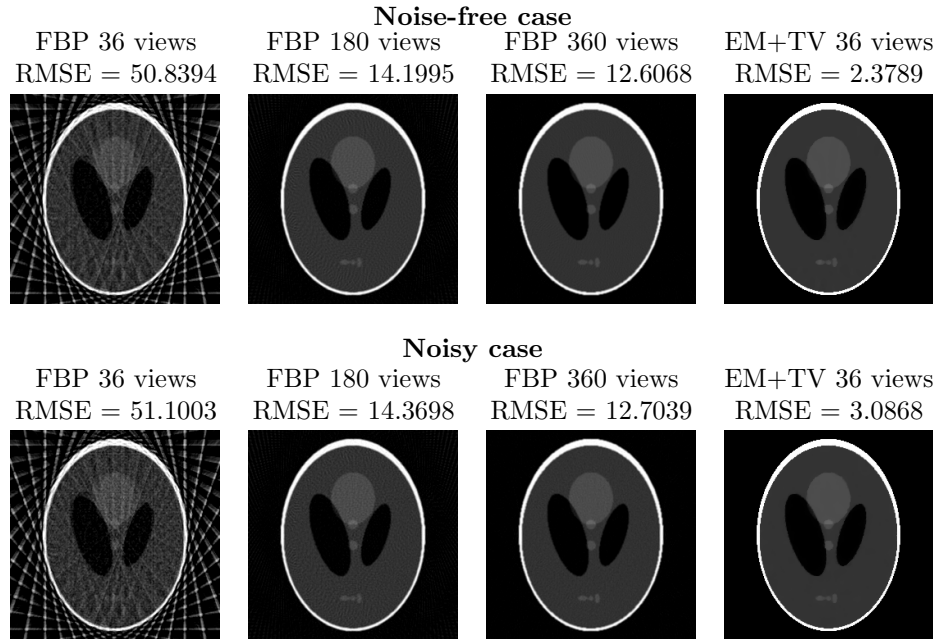


FIGURE 3. Top from left to right: reconstruction result in the noise-free case using FBP with 36, 180 and 360 views, and result using EM+TV with 36 views. Bottom from left to right: reconstruction result in the noisy case using FBP with 36, 180 and 360 views, and result using EM+TV with 36 views. The root mean square errors are also given.

6.2. CT Reconstruction using EM+MSTV (2D). Instead of TV regularization, we will show the results using a modified TV-like regularization, which is called Mumford-Shah TV (MSTV) [40]. This regularization is

$$J(x, v) = \int_{\Omega} v^2 |\nabla x| + \alpha \int_{\Omega} \left(\epsilon |\nabla v|^2 + \frac{(v-1)^2}{4\epsilon} \right),$$

which has two variables x and v , and Ω is the image domain. In the continuous case, it is shown by Alicandro et al. [2] that $J(x, v)$ will Γ -converge to

$$\int_{\Omega \setminus K} |\nabla x| + \alpha \int_K \frac{|x^+ - x^-|}{1 + |x^+ - x^-|} d\mathcal{H}^1 + |D^c x|(\Omega),$$

where x^+ and x^- denote the image values on two sides of the edge set K , \mathcal{H}^1 is the one-dimensional Hausdorff measure and $D^c x$ is the Cantor part of the measure-valued derivative Dx .

The comparisons between EM+TV and EM+MSTV in both noise-free and noisy cases are shown in Figure 4. From the results, we can see that with MSTV, the reconstructed images will be better than with TV only, visually and according to the root-mean-squared-error (RMSE).

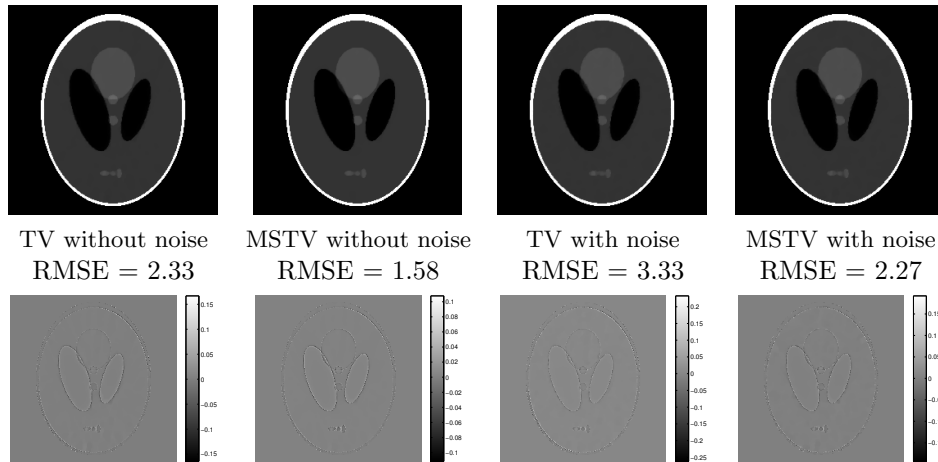


FIGURE 4. Comparisons of TV regularization and MSTV regularization for both without and with noise cases. Top row: reconstructed images by these two methods in both cases. Bottom row: differences (errors) between the reconstructed images and original phantom image. The RMSEs and differences show that MSTV can provide better results than TV only.

6.3. CT Reconstruction using EM+TV (3D). In this experiment, we will show the reconstruction results using EM+TV for a three dimensional phantom image. The image chosen is the $128 \times 128 \times 128$ Shepp-Logan phantom, and the sinogram data is obtained from 36 views. The result is compared with the one obtained using the EM step only in Figure 5.

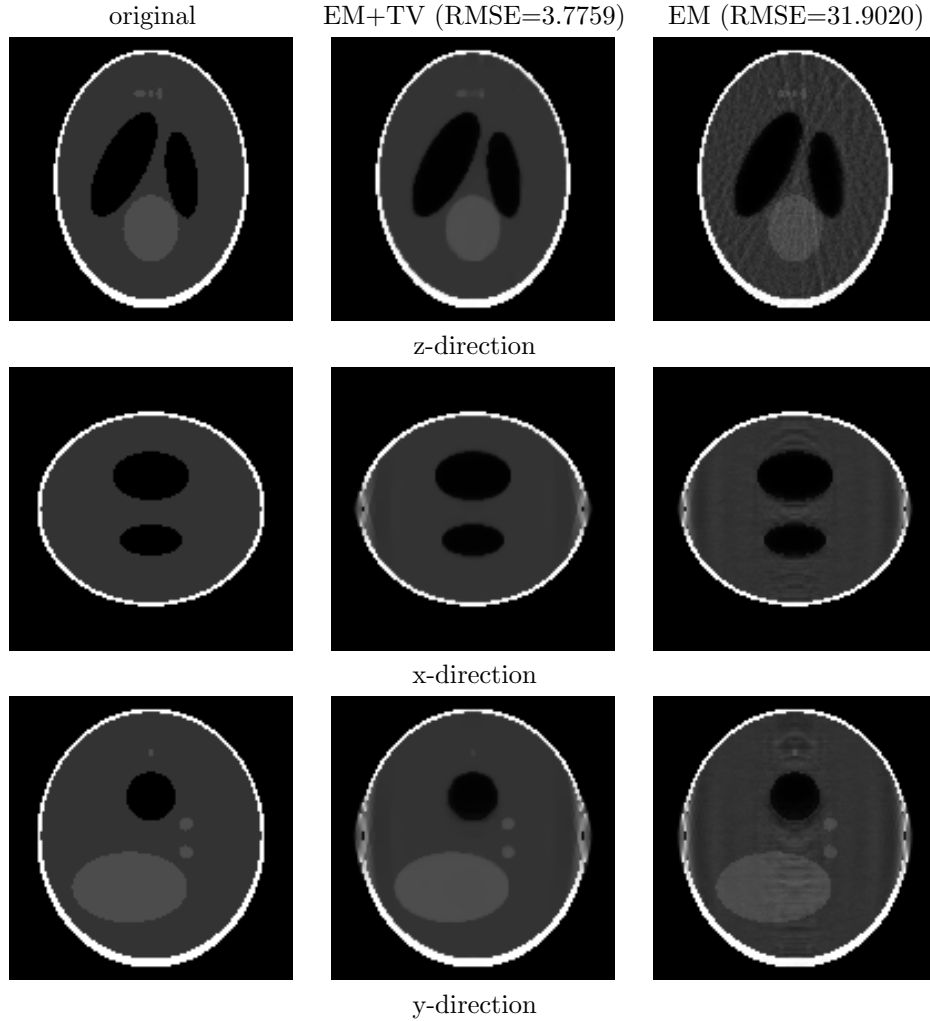


FIGURE 5. Reconstruction results in three dimensions in the noise-free case. First column: two-dimensional views of the original three-dimensional Shepp-Logan phantom. Middle column: two-dimensional views of reconstruction results obtained using EM+TV algorithm. Last column: two-dimensional views of reconstruction results obtained using EM iteration. The root mean square errors are also given.

7. Conclusion. In this paper, we proposed general robust EM-Type algorithms for image reconstruction without background emission. Both Poisson noise and Gaussian noise are considered. The EM-Type algorithms are performed using iteratively EM (or SART for weighted Gaussian noise) and regularization in the image domain. The convergence of these algorithms is proved in several ways: EM with *a priori* information and alternating minimization methods. To show the efficiency of EM-Type algorithms, the application in CT computerized tomography reconstruction is chosen. We compared EM+TV and EM+MSTV for 2D CT reconstruction. Both

methods can give us good results by using undersampled data comparing to the filtered back projection. Results from EM+MSTV have sharper edges than those from EM+TV. Also EM+TV is used for 3D CT reconstruction and the performance is better than using EM only (without regularization term) for undersampled data.

REFERENCES

- [1] R. Acar and C. R. Vogel, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, **10** (1994), 1217–1229.
- [2] R. Alicandro, A. Braides and J. Shah, *Free-discontinuity problems via functionals involving the L^1 -norm of the gradient and their approximation*, Interfaces and Free Boundaries, **1** (1999), 17–37.
- [3] A. Andersen, *Algebraic reconstruction in CT from limited views*, IEEE Transactions on Medical Imaging, **8** (1989), 50–55.
- [4] A. Andersen and A. Kak, *Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm*, Ultrasonic Imaging, **6** (1984), 81–94.
- [5] C. Atkinson and J. Soria, *An efficient simultaneous reconstruction technique for tomographic particle image velocimetry*, Experiments in Fluids, **47** (2009), 553–568.
- [6] C. Brune, M. Burger, A. Sawatzky, T. Kusters and F. Wubbeling, *Forward-Backward EM-TV methods for inverse problems with Poisson noise*, Preprint, august 2009.
- [7] C. Brune, A. Sawatzky and M. Burger, *Bregman-EM-TV methods with application to optical nanoscopy*, Lecture Notes in Computer Science, **5567** (2009), 235–246.
- [8] C. Brune, A. Sawatzky and M. Burger, *Primal and dual Bregman methods with application to optical nanoscopy*, International Journal of Computer Vision, **92** (2011), 211–229.
- [9] Y. Censor and T. Elfving, *Block-iterative algorithms with diagonally scaled oblique projections for the linear feasibility problem*, SIAM Journal on Matrix Analysis and Applications, **24** (2002), 40–58.
- [10] Y. Censor, D. Gordon and R. Gordon, *Component averaging: An efficient iterative parallel algorithm for large and sparse unstructured problems*, Parallel Computing, **27** (2001), 777–808.
- [11] J. Chen, J. Cong, L. A. Vese, J. D. Villasenor, M. Yan and Y. Zou, *A hybrid architecture for compressive sensing 3-D CT reconstruction*, IEEE Journal on Emerging and Selected Topics in Circuits and Systems, **2** (2012), 616–625.
- [12] J. A. Conchello and J. G. McNally, *Fast regularization technique for expectation maximization algorithm for optical sectioning microscopy*, in “Proceeding of SPIE Symposium on Electronic Imaging Science and Technology”, **2655** (1996), 199–208.
- [13] A. Dempster, N. Laird and D. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society Series B, **39** (1977), 1–38.
- [14] N. Dey, L. Blanc-Feraud, C. Zimmer, P. Roux, Z. Kam, J. C. Olivo-Marin and J. Zerubia, *Richardson-Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution*, Microscopy Research and Technique, **69** (2006), 260–266.
- [15] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **6** (1984), 721–741.
- [16] R. Gordon, R. Bender and G. Herman, *Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography*, Journal of Theoretical Biology, **29** (1970), 471–481.
- [17] P. J. Green, *On use of the EM algorithm for penalized likelihood estimation*, Journal of the Royal Statistical Society Series B, **52** (1990), 443–452.
- [18] U. Grenander, *Tutorial in pattern theory*, Lecture Notes Volume, Division of Applied Mathematics, Brown University, 1984.
- [19] Z. T. Harmany, R. F. Marcia and R. M. Willett, *Sparse Poisson intensity reconstruction algorithms*, in “Proceedings of IEEE/SP 15th Workshop on Statistical Signal Processing”, (2009), 634–637.
- [20] G. Herman, “Fundamentals of Computerized Tomography: Image Reconstruction From Projection”, Springer, 2009.
- [21] H. Hurwitz, *Entropy reduction in Bayesian analysis of measurements*, Physics Review A, **12** (1975), 698–706.

- [22] S. Jafarpour, R. Willett, M. Raginsky and R. Calderbank, *Performance bounds for expander-based compressed sensing in the presence of Poisson noise*, in “Proceedings of the IEEE Forty-Third Asilomar Conference on Signals, Systems and Computers”, (2009), 513–517.
- [23] X. Jia, Y. Lou, R. Li, W. Y. Song and S. B. Jiang, *GPU-based fast cone beam CT reconstruction from undersampled and noisy projection data via total variation*, *Medical Physics*, **37** (2010), 1757–1760.
- [24] M. Jiang and G. Wang, *Convergence of the simultaneous algebraic reconstruction technique (SART)*, *IEEE Transaction on Image Processing*, **12** (2003), 957–961.
- [25] M. Jiang and G. Wang, *Convergence studies on iterative algorithms for image reconstruction*, *IEEE Transactions on Medical Imaging*, **22** (2003), 569–579.
- [26] S. Joshi and M. I. Miller, *Maximum a posteriori estimation with Good’s roughness for three-dimensional optical sectioning microscopy*, *Journal of the Optical Society of America A*, **10** (1993), 1078–1085.
- [27] M. Jung, E. Resmerita and L. A. Vese, *Dual norm based iterative methods for image restoration*, *Journal of Mathematical Imaging and Vision*, **44** (2012), 128–149.
- [28] A. Kak and M. Slaney, “*Principles of Computerized Tomographic Imaging*”, Society of Industrial and Applied Mathematics, 2001.
- [29] W. Karush, “*Minima of Functions of Several Variables With Inequalities as Side Constraints*”, Master’s thesis, Department of Mathematics, University of Chicago, Chicago, Illinois, 1939.
- [30] H. Kuhn and A. Tucker, *Nonlinear programming*, in “Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability”, (1951), 481–492.
- [31] T. Le, R. Chartrand and T. J. Asaki, *A variational approach to reconstructing images corrupted by Poisson noise*, *Journal of Mathematical Imaging and Vision*, **27** (2007), 257–263.
- [32] E. Levitan and G. T. Herman, *A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography*, *IEEE Transactions on Medical Imaging*, **6** (1987), 185–192.
- [33] L. B. Lucy, *An iterative technique for the rectification of observed distributions*, *Astronomical Journal*, **79** (1974), 745–754.
- [34] J. Markham and J. A. Conchello, *Fast maximum-likelihood image-restoration algorithms for three-dimensional fluorescence microscopy*, *Journal of the Optical Society America A*, **18** (2001), 1052–1071.
- [35] F. Natterer and F. Wubbeling, “*Mathematical Methods in Image Reconstruction*”, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2001.
- [36] Y. Pan, R. Whitaker, A. Cheryauka and D. Ferguson, *Feasibility of GPU-assisted iterative image reconstruction for mobile C-arm CT*, in “Proceedings of International Society for Photonics and Optics” *SPIE* **7258** (2009), 72585J.
- [37] W. H. Richardson, *Bayesian-based iterative method of image restoration*, *Journal of the Optical Society America*, **62** (1972), 55–59.
- [38] L. Rudin, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, *Physica D*, **60** (1992), 259–268.
- [39] S. Setzer, G. Steidl and T. Teuber, *Deblurring poissonian images by split Bregman techniques*, *Journal of Visual Communication and Image Representation*, **21** (2010), 193–199.
- [40] H. Shah, *A common framework for curve evolution, segmentation and anisotropic diffusion*, in “Proceeding of IEEE Conference on Computer Vision and Pattern Recognition”, (1996), 136–142.
- [41] L. Shepp and B. Logan, *The Fourier reconstruction of a head section*, *IEEE Transaction on Nuclear Science*, **21** (1974), 21–34.
- [42] L. Shepp and Y. Vardi, *Maximum likelihood reconstruction for emission tomography*, *IEEE Transaction on Medical Imaging*, **1** (1982), 113–122.
- [43] R. Siddon, *Fast calculation of the exact radiological path for a three-dimensional CT array*, *Medical Physics*, **12** (1986), 252–255.
- [44] E. Y. Sidky, R. Chartrand and X. Pan, *Image reconstruction from few views by non-convex optimization*, in “IEEE Nuclear Science Symposium Conference Record”, **5** (2007).
- [45] E. Y. Sidky, J. H. Jorgensen and X. Pan, *Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm*, *Physics in Medicine and Biology*, **57** (2012), 3065.
- [46] A. N. Tychonoff and V. Y. Arsenin, “*Solution of Ill-posed Problems*”, Winston & Sons, Washington, 1977.

- [47] J. Wang and Y. Zheng, *On the convergence of generalized simultaneous iterative reconstruction algorithms*, IEEE Transaction on Image Processing, **16** (2007), 1–6.
- [48] R. M. Willett, Z. T. Harmany and R. F. Marcia, *Poisson image reconstruction with total variation regularization*, *Proceedings of 17th IEEE International Conference on Image Processing*, (2010), 4177–4180.
- [49] M. Yan and L. A. Vese, *Expectation maximization and total variation based model for computed tomography reconstruction from undersampled data*, in *Proceeding of SPIE Medical Imaging: Physics of Medical Imaging*, **7961** (2011), 79612X.
- [50] M. Yan, *Convergence analysis of SART: optimization and statistics*, International Journal of Computer Mathematics, **90** (2013), 30–47.
- [51] M. Yan, J. Chen, L. A. Vese, J. D. Villasenor, A. A. T. Bui and J. Cong, *EM+TV based reconstruction for cone-beam CT with reduced radiation*, in “Lecture Notes in Computer Science” **6938** (2011), 1–10.
- [52] H. Yu and G. Wang, *SART-type image reconstruction from a limited number of projections with the sparsity constraint*, Journal of Biomedical Imaging, **2010** (2010), 1–9.
- [53] H. Zhao and A. J. Reader, *Fast ray-tracing technique to calculate line integral paths in voxel arrays*, in IEEE Nuclear Science Symposium Conference Record, **4** (2003), M11–197.
- [54] D. Zhu, M. Razaz and R. Lee, *Adaptive penalty likelihood for reconstruction of multi-dimensional confocal microscopy images*, Computerized Medical Imaging and Graphics, **29** (2005), 319–331.
- [55] Compressive Sensing Resources, <http://dsp.rice.edu/cs>,

Received xxxx 20xx; revised xxxx 20xx.