General Convergent Expectation Maximization (EM)-Type Algorithms for Image Reconstruction with Background Emission and Poisson Noise

Ming Yan*

September 2, 2011

Abstract

Obtaining high quality images is very important in many areas of applied sciences. In this paper, we proposed general convergent and robust expectation maximization (EM)-Type algorithms for image reconstruction with background emission when the measured data is corrupted by Poisson noise. These methods are separated into two steps: EM step and regularization step, and these algorithms are shown to be equivalent to EM algorithms with *a priori* information and alternating minimization methods. The convergence of the algorithms is shown in several different ways. To overcome contrast reduction introduced by regularizations, EM-Type algorithms with Bregman iterations are introduced. The numerical experiments using different regularizations show the performance of these methods.

Keywords: Expectation Maximization, Image Reconstruction, Background Emission, Alternating Minimization Method, Poisson Noise.

1 Introduction

Obtaining high quality images is very important in many areas of applied sciences, 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 equipment [1]. However, it is sensitive to noise and suffers from streak artifacts (star artifacts). An alternative to analytical reconstruction is the use of iterative reconstruction technique, which is quite different from FBP. The main advantages of iterative reconstruction technique over FBP are insensitivity to noise and flexibility [2]. 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. We will focus on the iterative reconstruction technique.

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

$$y = Ax + b + n. \tag{1}$$

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), which is different for different applications, x is the desired exact image (vector in $\mathbf{R}^N$ in the discrete case), b is the background and n is the noise. We will consider only the case with background emission ($b \neq 0$) in this paper. In astronomy, this is due to sky emission [3, 4]. In fluorescence microscopy, it is due to auto-fluorescence and reflections of the excitation light. The computation of x directly by finding the inverse of A is not reasonable because (1) is ill-posed and n is unknown. 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 [5], and Grenander [6]. The idea is to use *a priori*

*Department of Mathematics, University of California, Los Angeles, CA 90095. Email: yanm@math.ucla.edu

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 the Bayesian formula gives us

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

This is a conditional probability of having $X = x$ given that y is the measured data. After inserting the detected value of 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 as a Gibbs random field, which is a random variable with the following probability distribution

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

where $J(x)$ is a given convex energy functional, 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 $J(x) = \|x\|_2^2/2$ [7, 8], quadratic gradient $J(x) = \|\nabla x\|_2^2/2$ [9], total variation $J(x) = \||\nabla x|\|_1$ [10, 11, 12, 13, 14], and Good's roughness penalization $J(x) = \||\nabla x|^2/x\|_1$ [15].

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

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

in the case of additive Gaussian noise, and the minimization of the negative log-likelihood function gives us the famous Tikhonov regularization method [16]

$$\underset{x\geq 0}{\text{minimize}} \ \frac{1}{2}\|Ax+b-y\|_2^2 + \beta J(x). \tag{5}$$

If the random variable Y of the detected values y follows a Poisson distribution [17, 18, 19] with an expectation value provided by $Ax + b$ instead of Gaussian distribution, we have

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

By minimizing the negative log-likelihood function $-\log p_Y(y|x)$, we obtain the following optimization problem

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

In this paper, we will focus on solving (7). It is easy to see that the objective function in (7) is convex. Additionally, with suitably chosen regularization $J(x)$, the objective function is strictly convex, and the solution to this problem is unique.

The work is organized as follows. The well-posedness of the problems (5) and (7) are shown in section 2. In section 3, we will give a short introduction of expectation maximization (EM) iteration, or Richardson-Lucy algorithm, used in image reconstruction with background emission from the view of optimization. In section 4, we will propose general EM-Type algorithms for image reconstruction with background emission when the measured data is corrupted by Poisson noise. This is based on the maximum a posteriori likelihood estimation and EM step. In this section, these EM-Type algorithms are shown to be equivalent to EM algorithms with *a priori* information, and the convergence of them is shown in a different way. In addition, these EM-Type algorithms are also considered as alternating minimization methods. For the case without regularization, more analysis on the convergence (the distance to the solution is decreasing) is provided. However, for some regularizations, the reconstructed images will lose contrast. To overcome this problem, EM-Type algorithms with Bregman iterations are introduced in section 5. Some numerical experiments are given in section 6 to show the efficiency of the EM-Type algorithms with different regularizations. We will end this work by a short conclusion section.

2 Well-Posedness of the Problems

As mentioned in the introduction, the original problem without regularization is ill-posed. Therefore at least one of these three properties: (i) a solution of the problem exists, (ii) the solution is unique, and (iii) the solution depends continuously on the data, are not fulfilled. For the well-posedness of the continuous modeling, the analysis will be different depending on different regularizations. If $J(x) = \int |\nabla x|$, i.e. the regularization is total variation, the well-posedness of regularization is shown in [20] and [13] for Gaussian and Poisson noise respectively. However, for the discrete modeling, the well-posedness of the problem is easy to prove, because problems (5) and (7) are convex. We have to just show that the solution is unique.

In the discrete modeling, the operator A is a matrix and x is a vector. After imposing some reasonable assumptions on $J(x)$ and A, the objective function is strictly convex, therefore the solution is unique. The strict convexity means that given two different vectors x^1 and x^2, then for any $w \in (0,1)$, the new vector $x_w = wx^1 + (1-w)x^2$ satisfies

$$\frac{1}{2}\|Ax_w - y\|_2^2 + \beta J(x_w) < w\frac{1}{2}\|Ax^1 - y\|_2^2 + w\beta J(x^1) + (1-w)\frac{1}{2}\|Ax^2 - y\|_2^2 + (1-w)\beta J(x^2). \tag{8}$$

If the objective function is not strictly convex, then we can find two different vectors x^1 and x^2 and $w \in (0,1)$ such that

$$\frac{1}{2}\|Ax_w - y\|_2^2 + \beta J(x_w) \geq w\frac{1}{2}\|Ax^1 - y\|_2^2 + w\beta J(x^1) + (1-w)\frac{1}{2}\|Ax^2 - y\|_2^2 + (1-w)\beta J(x^2). \tag{9}$$

From the convexity of the objective function, we have

$$\frac{1}{2}\|Ax_w - y\|_2^2 + \beta J(x_w) = w\frac{1}{2}\|Ax^1 - y\|_2^2 + w\beta J(x^1) + (1-w)\frac{1}{2}\|Ax^2 - y\|_2^2 + (1-w)\beta J(x^2), \tag{10}$$

for all $w \in (0,1)$. Since $\frac{1}{2}\|Ax - y\|_2^2$ and $J(x)$ are convex, we have

$$\frac{1}{2}\|Ax_w - y\|_2^2 = w\frac{1}{2}\|Ax^1 - y\|_2^2 + (1-w)\frac{1}{2}\|Ax^2 - y\|_2^2, \tag{11}$$

$$J(x_w) = wJ(x^1) + (1-w)J(x^2), \tag{12}$$

for all $w \in (0,1)$. From equation (11), we have $Ax^1 = Ax^2$. If A is injective, i.e. the null space of A is trivial, x^1 and x^2 have to be equal, then the objective function is strictly convex. If A is not injective (for instance, in reconstructions from positron emission tomography (PET) or computed tomography (CT)), we have to also consider equation (12). The equality in (12) depends on the regularization $J(x)$. For quadratic penalization, $J(x)$ is strictly convex, which implies $x^1 = x^2$, while for quadratic gradient, equation (12) gives us $\nabla x^1 = \nabla x^2$. If $J(x)$ is total variation, we obtain, from the equality, that $\nabla x^1 = k\nabla x^2$ with $k \geq 0$ and depends on the pixel (or voxel). When Good's roughness penalization is used, we have $\frac{\nabla x^1}{x^1} = \frac{\nabla x^2}{x^2}$ from the equality. Thus, if the matrix A is chosen such that we can not find two different vectors (images) satisfying $Ax^1 = Ax^2$ and $\nabla x^1 = k\nabla x^2$, the objective function is strictly convex. Actually, this assumption is reasonable and in the applications mentioned above, it is satisfied. Therefore, for the discrete modeling the optimization problem has a unique solution. If Poisson noise, instead of Gaussian noise, is assumed, the objective function is still strictly convex, and the problem has a unique solution.

3 Expectation Maximization (EM) Iteration

A maximum likelihood (ML) method for image reconstruction based on Poisson data was introduced by Shepp and Vardi [18] in 1982 for applications in emission tomography. In fact, this algorithm was originally proposed by Richardson [21] in 1972 and Lucy [22] in 1974 for astronomy. In this section, we consider the special case without regularization term, i.e. $J(x)$ is a constant, 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 estimate 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

$$f_0(x) = \sum_i \big((Ax+b)_i - y_i \log(Ax+b)_i\big), \tag{13}$$

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

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

where $D_{KL}(Ax+b, y)$ is the Kullback-Leibler (KL) divergence of $Ax+b$ 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 $\|Ax+b-y\|_2^2$ is the data-fidelity function for additive Gaussian noise. It is convex, nonnegative and coercive on the nonnegative orthant, so the minimizers exist and are global.

In order to find the minimizer, we can solve the Karush-Kuhn-Tucker (KKT) conditions [23, 24],

$$\begin{aligned} \sum_i \left(a_{ij}\left(1 - \frac{y_i}{(Ax+b)_i}\right)\right) - s_j &= 0, \qquad j = 1, \cdots, N, \\ s_j \geq 0, \quad x_j &\geq 0, \qquad j = 1, \cdots, N, \\ s^T x &= 0. \end{aligned}$$

Here s_j is the Lagrange multiplier corresponding to the constraint $x_j > 0$. 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 $j = 1, \cdots, N$. Multiplying by x_j gives us

$$\sum_i \left(a_{ij}\left(1 - \frac{y_i}{(Ax+b)_i}\right)\right) x_j = 0, \qquad j = 1, \cdots, N. \tag{14}$$

Therefore, we have the following iterative scheme

$$x_j^{k+1} = \frac{\sum_i \left(a_{ij}\left(\frac{y_i}{(Ax^k+b)_i}\right)\right)}{\sum_i a_{ij}} x_j^k. \tag{15}$$

This is the well-known EM iteration or Richardson-Lucy algorithm in image reconstruction, and an important property of it is that it will preserve positivity. If x^k is positive, then x^{k+1} is also positive if A preserves positivity.

Shepp and Vardi showed in [18] that when $b = 0$, this is equivalent to the EM algorithm proposed by Dempster, Laird and Rubin [25]. Actually, when $b \neq 0$, this is also equivalent to the EM algorithm and this will be shown in the next section. 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.

4 EM-Type Algorithms for Image Reconstruction

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 [26, 27], 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

$$\begin{cases} \underset{x}{\text{minimize}} & E^p(x) \equiv \beta J(x) + \sum_i \left((Ax+b)_i - y_i \log(Ax+b)_i\right), \\ \text{subject to} & x_j \geq 0, \quad j = 1, \cdots, N. \end{cases} \tag{16}$$

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

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

Here s_j is the Lagrange multiplier corresponding to the constraint $x_j > 0$. 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 $j = 1, \cdots, N$. Thus we obtain

$$\beta x_j \partial J(x)_j + \sum_i \left(a_{ij}(1 - \frac{y_i}{(Ax+b)_i}) \right) x_j = 0, \qquad j = 1, \cdots, N,$$

or equivalently

$$\beta \frac{x_j}{\sum\limits_i a_{ij}} \partial J(x)_j + x_j - \frac{\sum\limits_i \left(a_{ij}(\frac{y_i}{(Ax+b)_i}) \right)}{\sum\limits_i a_{ij}} x_j = 0, \qquad j = 1, \cdots, N.$$

Notice that the last term on the left hand side is an EM step (15). After plugging the EM step into the KKT condition, we obtain

$$\beta \frac{x_j}{\sum\limits_i a_{ij}} \partial J(x)_j + x_j - x_j^{EM} = 0, \qquad j = 1, \cdots, N,$$

which is the optimality for the following optimization problem

$$\underset{x}{\text{minimize}} \ \ E_1^p(x, x^{EM}) \equiv J(x) + \sum_j (\sum_i a_{ij}) \left(x_j - x_j^{EM} \log x_j \right). \tag{17}$$

Therefore we propose the general EM-Type algorithm in Algorithm **1**. The initial guess x^0 can be any positive intial image, and ϵ, chosen for the stopping criteria, is very small. 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 last section. When $J(x)$ is not constant, we have to solve an optimization problem for each iteration. In general, the problem can not be solved analytically, and we have to use iterative methods to solve it. In practice, we do not have to solve it exactly by stopping it after a few iterations. We will show that the algorithm will also converge without solving it exactly.

```
Input: Given x^0, ε, k = 0;
while k < Num_Iter & ||x^k − x^{k−1}|| < ε do
    k = k + 1;
    x^{k−1/2} = EM(x^{k−1}) using (15) ;
    x^k = argmin E_1^p(x, x^{k−1/2}) by solving (17);
end
```

Algorithm 1: Proposed EM-Type algorithm.

4.1 Equivalence to EM Algorithm 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 [25]. 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 $\{z_{ij}\}$, describing the intensity of pixel (or voxel) j observed by detector i and $\{\bar{b}_i\}$, the intensity of background observed by detector i. Therefore the observed data are $y_i = \sum\limits_j z_{ij} + \bar{b}_i$. We can have the assumption that z is a realization of multi-valued random variable Z, and for each (i, j) pair, z_{ij} follows a Poisson distribution with expected value $a_{ij} x_j$, and $\bar{b}_i$ follows a

Poisson distribution with expected value b_i, because the summation of two Poisson distributed random variables also follows a Poisson distribution, whose expected value is 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:

$$\begin{aligned} x^{k+1} &= \underset{x}{\operatorname{argmax}}\ E_{z|x^k,y} \log p(x, z|y) = \underset{x}{\operatorname{argmax}}\ E_{z|x^k,y} \log(p(y, z|x)p(x)) \\ &= \underset{x}{\operatorname{argmax}}\ E_{z|x^k,y} \sum_{ij} (z_{ij} \log(a_{ij}x_j) - a_{ij}x_j) - \beta J(x) \\ &= \underset{x}{\operatorname{argmin}}\ \sum_{ij} (a_{ij}x_j - E_{z|x^k,y} z_{ij} \log(a_{ij}x_j)) + \beta J(x). \end{aligned} \tag{18}$$

From (18), what we need before solving it is just $\{E_{z|x^k,y}z_{ij}\}$. Therefore we compute the expectation of missing data $\{z_{ij}\}$ given present x^k, denoted this as an E-step. Because for fixed i, $\{z_{ij}\}$ are Poisson variables with mean $\{a_{ij}x_j^k\}$ and $\bar{b}_i$ is Poisson variable with mean b_i, then the distribution of z_{ij}, is binomial distribution $\left(y_i, \frac{a_{ij}x_j^k}{(Ax^k+b)_i}\right)$, thus we can find the expectation of z_{ij} with all these conditions by the following E-step

$$z_{ij}^{k+1} \equiv E_{z|x^k,y} z_{ij} = \frac{a_{ij}x_j^k y_i}{(Ax^k + b)_i}. \tag{19}$$

After obtaining the expectation for all z_{ij}, then we can solve the M-step (18).

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

$$x_j^{EM} = \frac{\sum_i z_{ij}^{k+1}}{\sum_i a_{ij}}. \tag{20}$$

Therefore, the M-step is equivalent to

$$\begin{aligned} x^{k+1} &= \underset{x}{\operatorname{argmin}}\ \sum_{ij} (a_{ij}x_j - z_{ij}^{k+1} \log(a_{ij}x_j)) + \beta J(x) \\ &= \underset{x}{\operatorname{argmin}}\ \sum_{j} (\sum_i a_{ij})(x_j - x_j^{EM} \log(x_j)) + \beta J(x). \end{aligned}$$

4.2 Convergence of EM-Type Algorithms

In this subsection, we will show that the negative log-likelihood is decreasing in the following theorem.

Theorem 4.1 *The objective functional (negative log-likelihood) $E^p(x^k)$ in (16) with x^k given by Algorithm* **1** *will decrease until it attaints a minimum.*

Proof: For all k and i, we always have the constraint satisfied

$$\sum_j z_{ij}^k + \bar{b}_i = y_i.$$

Therefore, we have the following inequality

$$\begin{aligned}
& y_i \log\big((Ax^{k+1}+b)_i\big) - y_i \log\big((Ax^k+b)_i\big) \\
= & y_i \log\left(\frac{(Ax^{k+1}+b)_i}{(Ax^k+b)_i}\right) = y_i \log\left(\frac{\sum\limits_j a_{ij}x_j^{k+1}+b_i}{(Ax^k+b)_i}\right) \\
= & y_i \log\left(\sum_j \frac{a_{ij}x_j^k x_j^{k+1}}{(Ax^k+b)_i x_j^k} + \frac{b_i}{(Ax^k+b)_i}\right) \\
= & y_i \log\left(\sum_j \frac{z_{ij}^{k+1} a_{ij} x_j^{k+1}}{y_i a_{ij} x_j^k} + \frac{\bar{b}_i}{y_i}\right) \\
\geq & y_i \sum_j \frac{z_{ij}^{k+1}}{y_i} \log\left(\frac{a_{ij}x_j^{k+1}}{a_{ij}x_j^k}\right) \qquad \text{(Jensen's inequality)} \\
= & \sum_j z_{ij}^{k+1} \log(a_{ij}x_j^{k+1}) - \sum_j z_{ij}^{k+1} \log(a_{ij}x_j^k).
\end{aligned} \tag{21}$$

This inequality gives us

$$\begin{aligned}
& E^p(x^{k+1}) - E^p(x^k) \\
= & \sum_i ((Ax^{k+1}+b)_i - y_i \log(Ax^{k+1}+b)_i) + \beta J(x^{k+1}) \\
& - \sum_i \big((Ax^k+b)_i - y_i \log(Ax^k+b)_i\big) - \beta J(x^k) \\
\leq & \sum_{ij} (a_{ij}x_j^{k+1} - z_{ij}^{k+1} \log(a_{ij}x_j^{k+1})) + \beta J(x^{k+1}) \\
& - \sum_{ij} (a_{ij}x_j^k - z_{ij}^{k+1} \log(a_{ij}x_j^k)) - \beta J(x^k) \\
\leq & \, 0.
\end{aligned}$$

The first inequality comes from (21) and the second inequality comes from the M-step (18). When $E(x^{k+1}) = E(x^k)$, these two equalities have to be satisfied. The first equality is satisfied if and only if $x_j^{k+1} = x_j^k$ for all j, while the second one is satisfied if and only if x^k and x^{k+1} are minimizers of the M-step (18). The functional to be minimized in M-step (18) is strictly convex, which means that we have

$$\beta x_j^k \partial J(x^k)_j + \sum_i a_{ij}x_j^k - \sum_i z_{ij}^{k+1} = 0, \qquad j = 1, \cdots, N.$$

After plugging the E-step (19) into these equations, we have

$$\beta x_j^k \partial J(x^k)_j + \sum_i a_{ij}x_j^k - \sum_i \frac{a_{ij}x_j^k y_i}{(Ax^k+b)_i} = 0, \qquad j = 1, \cdots, N.$$

Therefore, x^k is one minimizer of the original problem. ∎

The log-likelihood function will increase for each iteration until the solution is found, and from the proof, we do not fully use the M-step. Even if the M-step is not solved exactly, it will still increase as long as $Q(x^{k+1}|x^k) > Q(x^k|x^k)$ is satisfied.

The increasing of log-likelihood function can be proved in another way by using the M-step. From $x^{k+1} = \underset{x}{\operatorname{argmax}}\ Q(x|x^k)$, we have

$$\beta x_j^{k+1} \partial J(x^{k+1})_j + \sum_i a_{ij} x_j^{k+1} - \sum_i z_{ij}^{k+1} = 0, \qquad j = 1, \cdots, N.$$

Multiplying by $(x_j^{k+1} - x_j^k)/x_j^{k+1}$ and taking summation over j give us

$$\beta \sum_j (x_j^{k+1} - x_j^k) \partial J(x^{k+1})_j + \sum_{ij} a_{ij}(x_j^{k+1} - x_j^k) - \sum_{ij} z_{ij}^{k+1} \frac{x_j^{k+1} - x_j^k}{x_j^{k+1}} = 0.$$

From the convexity of $J(x)$, we have

$$J(x^k) \geq J(x^{k+1}) + (x^k - x^{k+1}) \partial J(x^{k+1}).$$

Therefore we have

$$\begin{aligned}
0 &\geq \beta J(x^{k+1}) - \beta J(x^k) + \sum_{ij} a_{ij}(x_j^{k+1} - x_j^k) - \sum_{ij} z_{ij}^{k+1} \frac{x_j^{k+1} - x_j^k}{x_j^{k+1}} \\
&= E^p(x^{k+1}) - E^p(x^k) + \sum_i y_i \log\left(\frac{(Ax^{k+1} + b)_i}{(Ax^k + b)_i}\right) - \sum_{ij} z_{ij}^{k+1} \frac{x_j^{k+1} - x_j^k}{x_j^{k+1}} \\
&\geq E^p(x^{k+1}) - E^p(x^k) + \sum_i y_i \left(1 - \frac{(Ax^k + b)_i}{(Ax^{k+1} + b)_i}\right) - \sum_{ij} z_{ij}^{k+1} \frac{x_j^{k+1} - x_j^k}{x_j^{k+1}} \\
&= E^p(x^{k+1}) - E^p(x^k) - \sum_i y_i \frac{\sum_j a_{ij} x_j^k + b_i}{\sum_j a_{ij} x_j^{k+1} + b_i} + \sum_{ij} z_{ij}^{k+1} \frac{x_j^k}{x_j^{k+1}} + \sum_i (y_i - \sum_j z_{ij}^{k+1}) \\
&\geq E^p(x^{k+1}) - E^p(x^k).
\end{aligned}$$

The second inequality comes from $\log(x) \geq 1 - 1/x$ for $x > 0$, and the last inequality comes from Cauchy-Schwarz inequality. If $E^P(x^{k+1}) = E^P(x^k)$, from the last inequality, we have $x_j^{k+1} = x_j^k$ for all j. Therefore, the log-likelihood function will increase untill the solution is found.

4.3 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 problem is

$$\begin{aligned}
\underset{x,z}{\operatorname{minimize}}\ E^p(x,z) \quad := \quad & \sum_{ij} \left(z_{ij} \log \frac{z_{ij}}{a_{ij} x_j} + a_{ij} x_j - z_{ij} \right) \\
& + \sum_i \left(\bar{b}_i \log \frac{\bar{b}_i}{b_i} + b_i - \bar{b}_i \right) + \beta J(x),
\end{aligned} \tag{22}$$

where $\bar{b}_i = y_i - \sum\limits_j z_{ij},$ for all $i = 1, \cdots, M$. Having initial guess $x^0,\ z^0$ of x and z, the iteration for $k = 0, 1, \cdots$ is as follows:

$$\begin{aligned}
z^{k+1} &= \underset{z}{\operatorname{argmin}}\ E^p(x^k, z), \\
x^{k+1} &= \underset{x}{\operatorname{argmin}}\ E^p(x, z^{k+1}).
\end{aligned}$$

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

$$z_{ij}^{k+1} = \frac{a_{ij}x_j^k y_i}{(Ax^k+b)_i}. \tag{23}$$

After we find z^{k+1}, fix $z = z^{k+1}$ and update x, then we have

$$\begin{aligned} x^{k+1} &= \operatorname*{argmin}_x \sum_{ij}\left(a_{ij}x_j + z_{ij}^{k+1}\log\frac{z_{ij}^{k+1}}{a_{ij}x_j}\right) + \beta J(x) \\ &= \operatorname*{argmin}_x \sum_{ij}\left(a_{ij}x_j - z_{ij}^{k+1}\log(a_{ij}x_j)\right) + \beta J(x), \end{aligned}$$

which is the M-Step in section **4.1**. The equivalence of problems (16) and (22) is provided in the following theorem.

Theorem 4.2 *If (x^*, z^*) is a solution of problem (22), then x^* is also a solution of (16), i.e. $x^* = \operatorname*{argmin}_x \; E^p(x)$. If x^* is a solution of (16), then we can find z^* from (23) and (x^*, z^*) is a solution of problem (22).*

Proof: The equivalence can be proved in two steps. Firstly, we will show that $E^P(x,z) \geq E^P(x) + c$ for all z, here c is constant dependent on y only.

$$\begin{aligned} E^p(x,z) &= \sum_{ij}\left(z_{ij}\log\frac{z_{ij}}{a_{ij}x_j} + a_{ij}x_j - z_{ij}\right) + \sum_i\left(\bar{b}_i\log\frac{\bar{b}_i}{b_i} + b_i - \bar{b}_i\right) + \beta J(x) \\ &= \sum_{ij}\left(\frac{z_{ij}}{y_i}\log\frac{z_{ij}}{a_{ij}x_j}\right)y_i + \sum_i \frac{\bar{b}_i}{y_i}\log\frac{\bar{b}_i}{b_i}y_i + \sum_i((Ax+b)_i - y_i) + \beta J(x) \\ &\geq \sum_i y_i \log\left(\frac{y_i}{(Ax+b)_i}\right) + \sum_i\left((Ax+b)_i - y_i\right) + \beta J(x) \\ &= E^p(x) + \sum_i (y_i\log y_i - y_i). \end{aligned}$$

The inequality comes from Jensen's inequality, and the equality is satisfied if and only if

$$\frac{z_{ij}}{a_{ij}x_j} = \frac{\bar{b}_i}{b_i} = c_i, \qquad \forall j = 1, \cdots, N, \tag{24}$$

where c_i are constant. Therefore $\min_z E^P(x,z) = E^P(x) + c$, which means that problems (22) and (16) are equivalent. ∎

From these two convergence analyses, if the second part of the EM-Type algorithm can not be solved exactly, we can choose the initial guess to be the result from previous iteration, then use any method for solving convex optimization problem to obtain a better result.

4.4 More Analysis for the Case Without Regularization

For the case without regularization, we will show that for each limit point $\tilde{x}$ of the sequence $\{x^k\}$, we have $D_{KL}(\tilde{x}, x^{k+1}) \leq D_{KL}(\tilde{x}, x^k)$ if $\sum_i a_{ij} = 1$ for all j. If this condition is not fulfilled, similarly, we can show that $D_{KL}(\tilde{x}', {x^{k+1}}') \leq D_{KL}(\tilde{x}', {x^k}')$, where $\tilde{x}'_j = \sum_i a_{ij}\tilde{x}_j$ and ${x^k}'_j = \sum_i a_{ij}x_j^k$ for all j.

Theorem 4.3 *If $\sum_i a_{ij} = 1$ for all j, $D_{KL}(\tilde{x}, x^k)$ is decreasing for the case without regularization.*

Proof: Define vectors f^j, g^j such that their components are

$$f_i^j = \frac{a_{ij}y_i/(A\tilde{x}+b)_i}{(A^T(y/(A\tilde{x}+b)))_j}, \qquad g_i^j = \frac{a_{ij}y_i/(Ax^k+b)_i}{(A^T(y/(Ax^k+b)))_j}, \qquad i = 1, \cdots n, \tag{25}$$

then we have $\sum_i f_i^j = \sum_i g_i^j = 1$ and

$$\begin{aligned}
0 &\le \sum_j \tilde{x}_j D_{KL}(f^j, g^j) \\
&= \sum_j \tilde{x}_j \sum_i f_i^j \log \frac{f_i^j}{g_i^j} \\
&= \sum_j \tilde{x}_j \sum_i \frac{a_{ij} y_i/(A\tilde{x}+b)_i}{(A^T(y/(A\tilde{x}+b)))_j} \log \frac{(Ax^k+b)_i (A^T(y/(Ax^k+b)))_j}{(A\tilde{x}+b)_i (A^T(y/(A\tilde{x}+b)))_j} \\
&= \sum_j \tilde{x}_j \sum_i \frac{a_{ij} y_i/(A\tilde{x}+b)_i}{(A^T(y/(A\tilde{x}+b)))_j} \log \frac{(Ax^k+b)_i x_j^{k+1} \tilde{x}_j}{(A\tilde{x}+b)_i \tilde{x}_j x_j^k}.
\end{aligned}$$

Since

$$\tilde{x}_j = \frac{(A^T(y/(A\tilde{x}+b))_j)}{(A^T 1)_j} \tilde{x}_j,$$

we have

$$\frac{(A^T(y/(A\tilde{x}+b)))_j}{(A^T 1)_j} = 1,$$

It follows that

$$\begin{aligned}
0 &\le \sum_j \tilde{x}_j \sum_i \frac{a_{ij} y_i}{(A\tilde{x}+b)_i} \log \frac{(Ax^k+b)_i x_j^{k+1}}{(A\tilde{x}+b)_i x_j^k} \\
&= \sum_j \tilde{x}_j \sum_i \frac{a_{ij} y_i}{(A\tilde{x}+b)_i} \left(\log \frac{(Ax^k+b)_i}{(A\tilde{x}+b)_i} + \log \frac{x_j^{k+1}}{x_j^k} \right) \\
&= \sum_j \tilde{x}_j \sum_i \frac{a_{ij} y_i}{(A\tilde{x}+b)_i} \log \frac{(Ax^k+b)_i}{(A\tilde{x}+b)_i} + \sum_j \tilde{x}_j \log \frac{x_j^{k+1}}{x_j^k} \\
&= \sum_i \frac{(A\tilde{x})_i y_i}{(A\tilde{x}+b)_i} \log \frac{(Ax^k+b)_i}{(A\tilde{x}+b)_i} + \sum_j \tilde{x}_j \log \frac{x_j^{k+1}}{x_j^k} \\
&= D_{KL}(y, A\tilde{x}+b) - D_{KL}(y, Ax^k+b) + D_{KL}(\tilde{x}, x^k) - D_{KL}(\tilde{x}, x^{k+1}) \\
&\quad - \sum_i \frac{b_i y_i}{(A\tilde{x}+b)_i} \log \frac{(Ax^k+b)_i}{(A\tilde{x}+b)_i} - \sum_j \tilde{x}_j + \sum_j x_j^{k+1}
\end{aligned}$$

Since $\sum_i y_i - \sum_j x_j^{k+1} = \sum_i (y_i - \sum_j z_{ij}^{k+1}) = \sum_i \frac{b_i y_i}{(Ax^k+b)_i}$, we have

$$\begin{aligned}
&- \sum_i \frac{b_i y_i}{(A\tilde{x}+b)_i} \log \frac{(Ax^k+b)_i}{(A\tilde{x}+b)_i} - \sum_j \tilde{x}_j + \sum_j x_j^{k+1} \\
&= - \sum_i \frac{b_i y_i}{(A\tilde{x}+b)_i} \log \frac{(Ax^k+b)_i}{(A\tilde{x}+b)_i} + \sum_i \frac{b_i y_i}{(A\tilde{x}+b)_i} - \sum_i \frac{b_i y_i}{(Ax^k+b)_i} \\
&\le - D_{KL}\left(\frac{b_i y_i}{(A\tilde{x}+b)_i}, \frac{b_i y_i}{(Ax^k+b)_i}\right) \le 0.
\end{aligned}$$

The decreasing of the objective function $D_{KL}(y, Ax^k+b)$ gives us $D_{KL}(y, A\tilde{x}+b) \le D_{KL}(y, Ax^k+b)$ and it follows that

$$0 \leq D_{KL}(\tilde{x}, x^k) - D_{KL}(\tilde{x}, x^{k+1})$$

which is $D_{KL}(\tilde{x}, x^{k+1}) \leq D_{KL}(\tilde{x}, x^k)$. ∎

If $\sum_i a_{ij} = 1$ is not satisfied, we have the same property for $\tilde{x}'$ and $x^{k'}$, which are just weighted vectors with the j^{th} weight being $\sum_i a_{ij}$, from the same proof.

5 EM-Type Algorithms with Bregman Iteration

In the previous section, the EM-Type algorithms are presented to solve problem (16). However, the regularization may lead to reconstructed images suffering from contrast reduction [28]. Therefore, we suggest a contrast improvement in EM-Type algorithms by Bregman iterations, which is introduced in [29, 30, 31]. An iterative refinement is obtained from a sequence of modified EM-Type algorithms.

For the problem with Poisson noise, we start with the basic EM-Type algorithm, i.e. finding the minimum x^1 of (16). After that, variational problems with a modified regularization term

$$x^{k+1} = \underset{x}{\operatorname{argmin}} \ \beta(J(x) - \langle p^k, x\rangle) + \sum_i ((Ax+b)_i - y_i \log(Ax+b)_i) \tag{26}$$

where $p^k \in \partial J(x^k)$, are solved sequentially. From the optimality of (26), we have the following formula for updating p^{k+1} from p^k and x^{k+1}:

$$p^{k+1} = p^k - \frac{1}{\beta} A^T \left(1 - \frac{y}{Ax^{k+1}+b}\right). \tag{27}$$

Therefore the EM-Type algorithms with Bregman iterations are as follows:

Input: Given x^0, δ, ϵ, $k=1$ and $p^0 = 0$;
while $k \leq Num\_outer$ & $D_{KL}(Ax^{k-1}+b, y) < \delta$ **do**
  $x^{temp,0} = x^{k-1}$;
  $l = 0$;
  while $l \leq Num\_inner$ & $\|x^{temp,l} - x^{temp,l-1}\| \leq \epsilon$ **do**
    $l = l+1$;
    $x^{temp,l-\frac{1}{2}} = EM(x^{temp,l-1})$ using (15) ;
    $x^{temp,l} = \underset{x}{\operatorname{argmin}} \ E_1^p(x, x^{temp,l-\frac{1}{2}})$ with $J(x) - \langle p^{k-1}, x\rangle$;
  end
  $x^k = x^{temp,l}$;
  $p^k = p^{k-1} - \frac{1}{\beta} A^T \left(1 - \frac{y}{Ax^k+b}\right)$;
  $k = k+1$;
end

Algorithm 2: Proposed EM-Type algorithm with Bregman iterations.

The initial guess x^0 can be any positive image, and $\delta = D_{KL}(Ax^*+b, y)$, where x^* is the ground truth, is assumed to be known. Here, $\epsilon > 0$ is a small parameter used in the stopping criterion.

6 Numerical Experiments

In this section, we will illustrate the proposed EM-Type algorithms for image reconstruction (more specifically, image deblurring). At the beginning, we present some deblurring results on a phantom with the proposed EM-TV algorithm, one example of EM-Type algorithms with total variation (TV) regularization, and the Bregman version of it. The phantom used in this section is a synthetic 200×200 phantom. It consists of circles with intensities 65, 110 and 170, enclosed by a square frame of intensity 10. For the experiment, we choose the backgroud $b = 20$. Firstly, we consider the case without noise. The blurred image is obtained from original image using a Gaussian blur kernel K with standard deviation $\sigma = 100$. The result is shown in Figure 1. The root mean square error (RMSE) is 2.5629 and the KL distance is 0.0080.

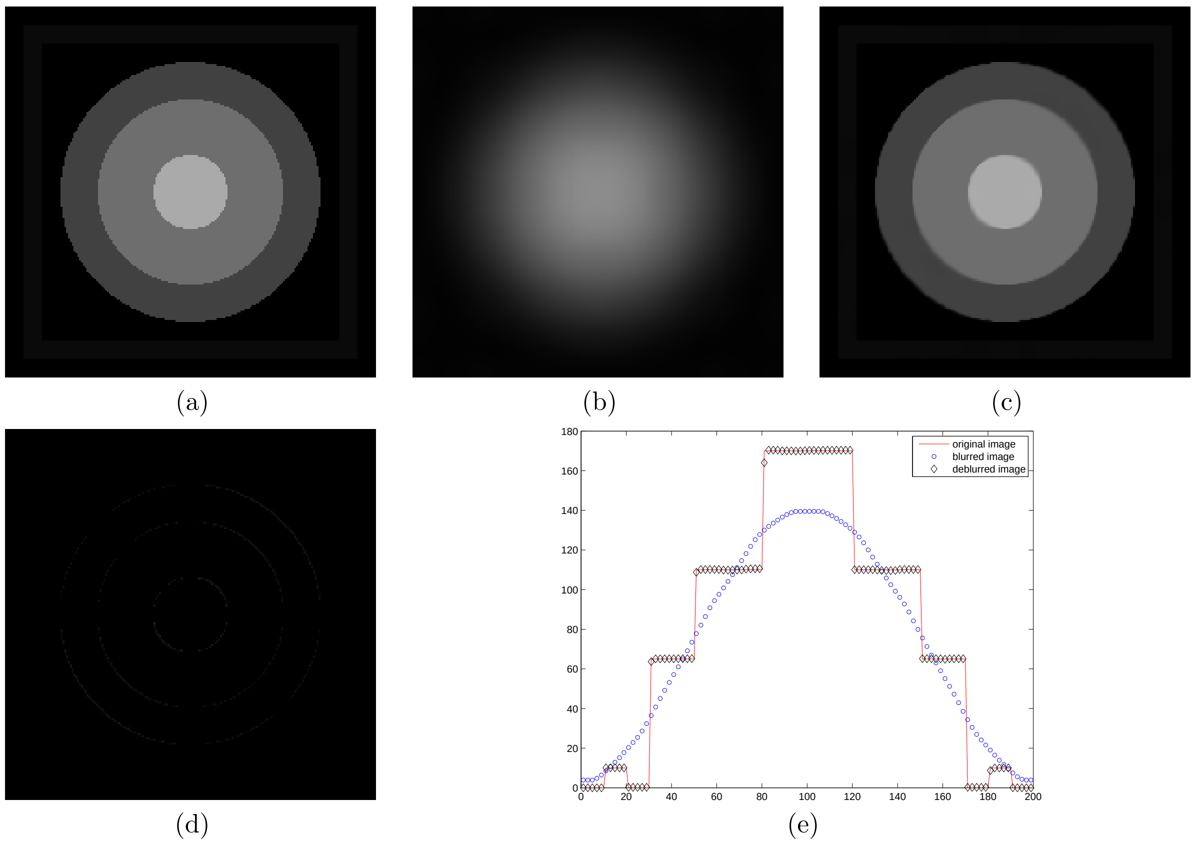

Figure 1: (a) The orginal image u_*. (b) Blurred image $K * u_*$ using a Gaussian blur kernel K. (c) The deblurred image using the proposed EM-TV with Bregman iterations. (d) The difference between the deblurred image and the original image. (e) The lineouts of original image, blurred image and deblurred image. Some parameters chosen are $\beta = 5$, Num_inner = 1 and Num_outer = 10000.

To illustrate the advantage of Bregman iterations, we show the comparison results in Figure 2. The RMSE for 2(a), 2(b) and 2(c) are 11.9039, 5.0944 and 2.5339, respectively. The corresponding KL distances are 93.0227, 0.8607 and 0.0172, respectively.

EM-TV with Bregman iterations can provide us with very good result if there is no noise in the blurred images. However, the noise is unavoidable in applications. The next experiment is to illustrate the EM-TV algorithm on noisy case. The RMSE for 3(b) and 3(c) are 12.9551 and 4.1176, respectively.

Next, the EM-TV algorithm is used to perform deconvolution on an image of a satellite (4(a)), and the point spread function (PSF) is shown in Figure 4(b). In order to make the algorithm fast, we choose the initial guess x^0 to be the result from solving $Ax = y - b$ using conjugate gradient (CG). The negative values are changed into zero before applying EM-TV algorithm. The corresponding RMSE for x^0 and the result are 13.6379 and 11.8127, respectively. By using the EM-TV with Bregman iterations, we get a better image with sharp edges and artifacts are removed.

The same EM-TV algorithm is also tested on an image of text (5(a)) and the point spread function (PSF) is shown in Figure 5(b). In order to make the algorithm fast, we choose the initial guess x^0 to be the result from solving $Ax = y - b$ using Hybrid Bidiagonalization Regularization (HyBR) [32]. The negative values are changed into zero before applying EM-TV algorithm. The corresponding RMSE for x^0 and the result are 45.8918 and 37.8574, respectively. By using the EM-TV with Bregman iteration, we get a better image with sharp edges and artifacts are removed.

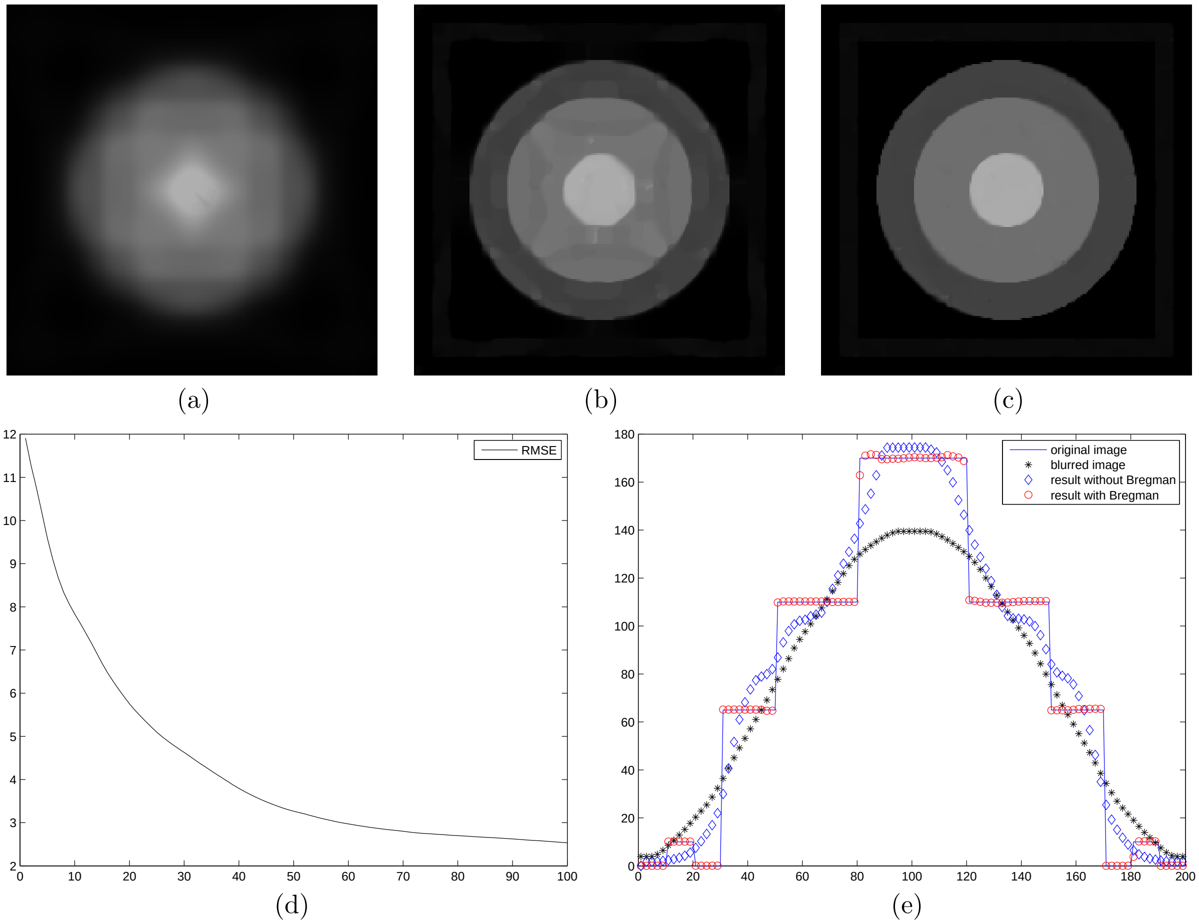

Figure 2: (a) The result without Bregman iteration. (b) The result with 25 Bregman iterations. (c) The result with 100 Bregman iterations. (d) The plot of RMSE versus Bregman iterations. (e) The lineouts of original image, blurred image, the results with and without Bregman iterations. Some parameters chosen are $\beta = 0.001$, Num_inner $= 100$ and Num_outer $= 100$.

The convergence analysis of EM-Type algorithms is for the case when $J(x)$ is convex. When $J(x)$ is not convex, we still have the same algorithms, and from the equivalence with alternating minimization method, the algorithm will converge to a local minimum of the functional. For the last experiment (Figure 6), we will try to seperate the spare objects in lensfree fluorescent imaging [33] using EM-Type algorithm with a non-convex $J(x)$. The result of EM (or Richardson-Lucy) method will tend to be sparse, because the l_1 norm is almost fixed for all the iterations (when $b = 0$, the l_1 norm is fixed), but EM method can not seperate the particles when they are close to each other ($13\mu m$ and $10\mu m$ in this experiment). Therefore, we can choose $J(x) = \sum_j |x_j|^p$ for $p \in (0, 1)$, and these two particles can be separated even when the distance is very small). For the numerical experiment, top row shows the lensfree raw images. As the distance between particles become smaller, their signatures become indistinguishable to the bare eye. The PSF is measured using small diameter fluorescent particles that are imaged at a low concentration. we choose the same number of iterations for EM-l_p and EM method, and the results show that with $p < 1$, we can obtain better results.

7 Conclusion

In this paper, we proposed general robust EM-Type algorithms for image reconstruction when the measured data is corrupted by Poisson noise: iteratively performing EM and regularization in the image domain. The convergence of these algorithms is proved in several ways. For the case without regularization, the KL distance to the limit of the sequence of iterations is decreasing. The problem with regularization will lead to contrast

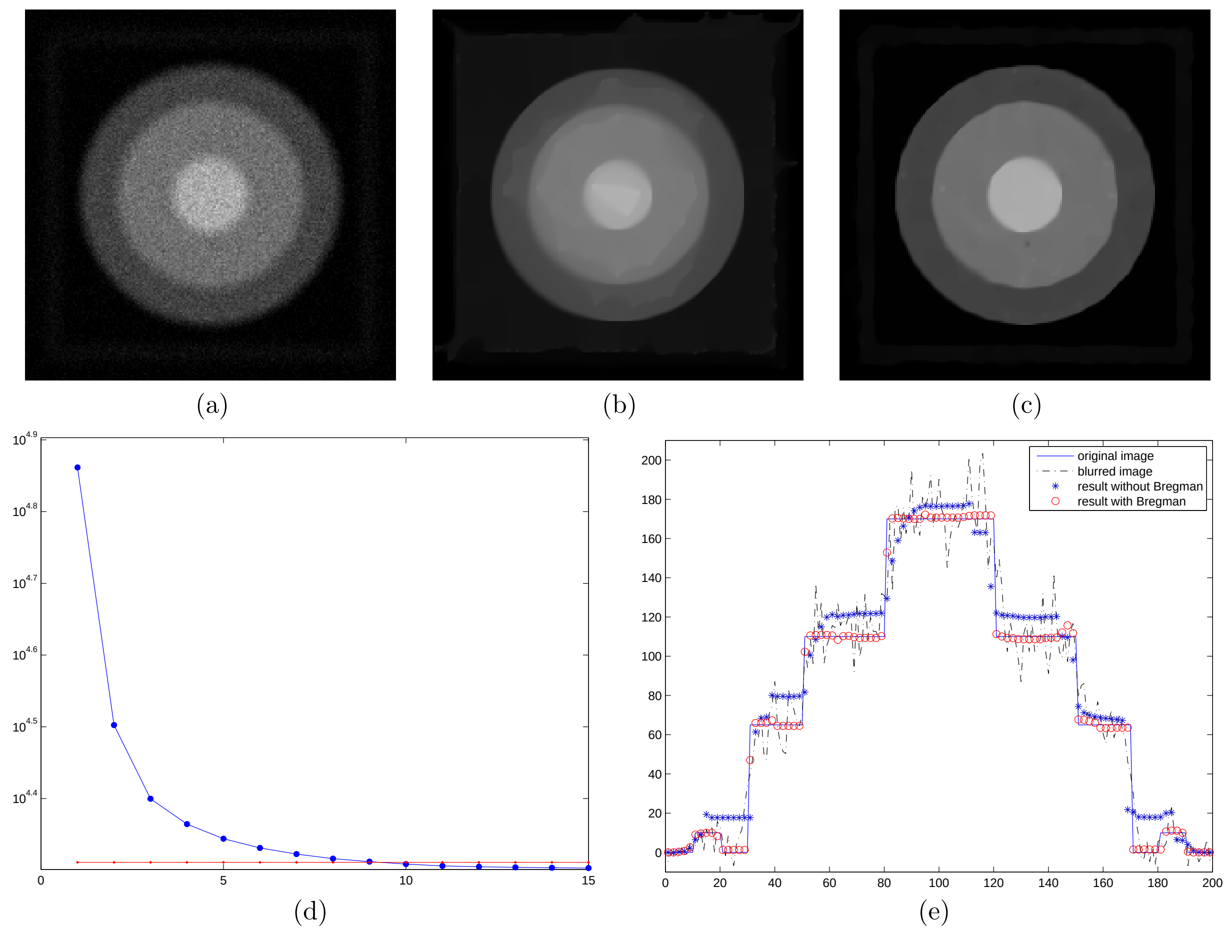

Figure 3: (a) The noisy blurred image. (b) The result without Bregman iteration. (c) The result with 9 Bregman iterations. (d) The plot of KL distances versus Bregman iterations. (e) The lineouts of original image, blurred image, the results with and without Bregman iterations. Some parameters chosen are $\beta = 1$, Num_inner $= 200$ and Num_outer $= 15$.

reduction in the reconstructed images. Therefore, in order to improve the contrast, we suggested EM-Type algorithms with Bregman iterations by applying a sequence of modified EM-Type algorithms. We tried EM-Type algorithms with different $J(x)$. With TV regularization, this EM-TV algorithm can provied images with preserved edges and artifacts are removed. With l_p regularization, this EM-l_p algorithm can be used to separate sparse particles even when the distance is small (better than in the case of EM only method).

Acknowledgement

The author would like to thank Luminita A. Vese for valuable discussions and suggestions. This work is supported by the Center for Domain-Specifi Computing (CDSC) funded by the NSF Expedition in Computing Award CCF-0926127.

References

[1] L. Shepp and B. Logan, “The Fourier reconstruction of a head section”, *IEEE Transaction on Nuclear Science*, **21** (1974), pp. 21–34.

[2] A. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, Society of Industrial and Applied Mathematics, (2001).

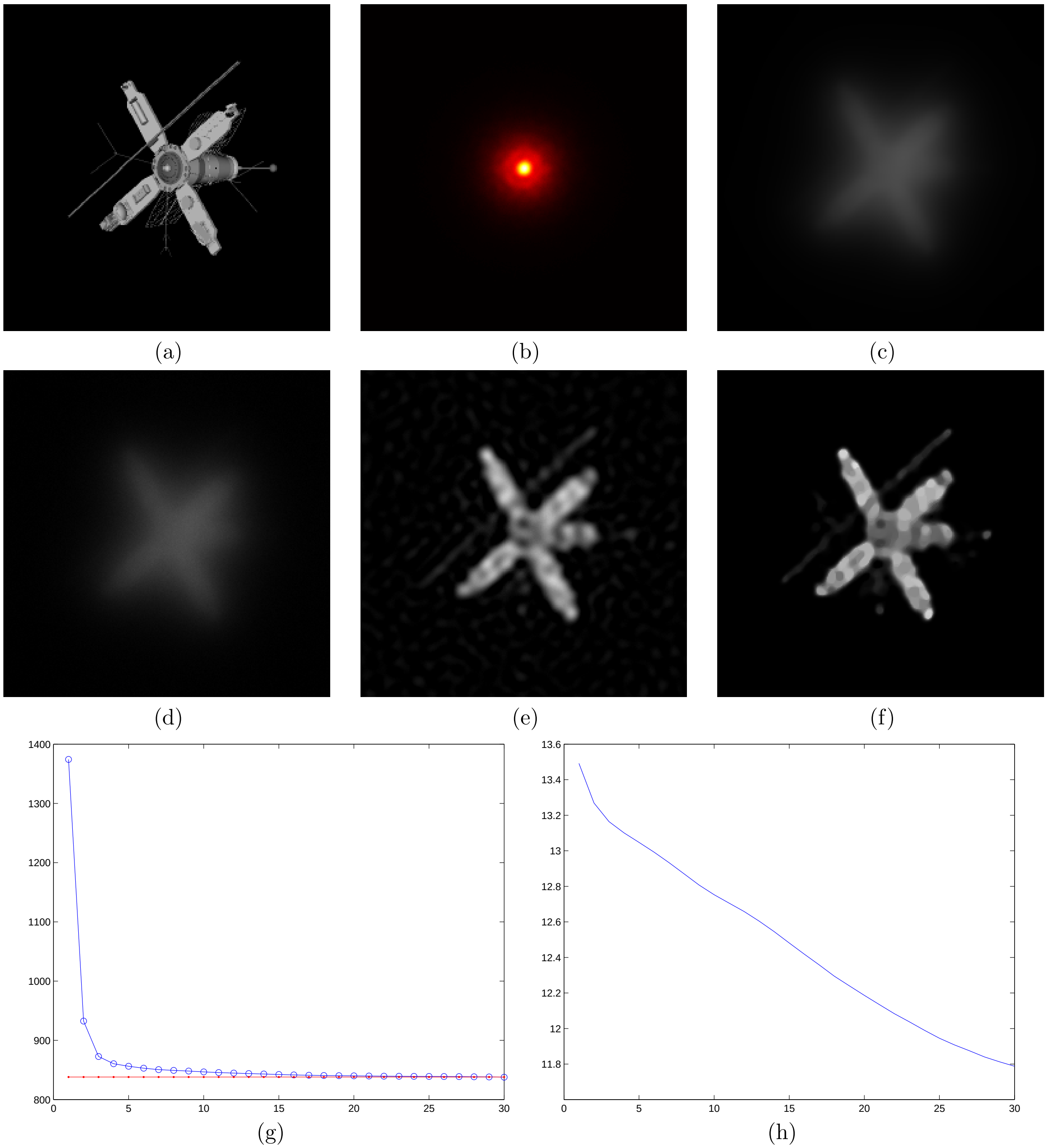

Figure 4: (a) The original image. (b) The PSF image. (c) The blurred image. (d) The noisy blurred image. (e) Initial guess from CG. (f) The result of EM-Type algorithm with Bregman iterations. (g) The plot of KL divergence versus Bregman iterations. (h) The RMSE versus Bregman iterations. Some parameters chosen are $\beta = 1$, Num_inner = 200 and Num_outer = 30.

[3] M. Bertero, P. Boccacci, G. Desidera and G. Vicidomini, "Image deblurring with Poisson data: from cells to galaxies", *Inverse Problem*, **25** (2009), 123006.

[4] D. G. Politte and D. L. Snyder, "Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron-emission tomography", *IEEE Transaction on Medical Imaging*, **10** (1991), pp. 82–89.

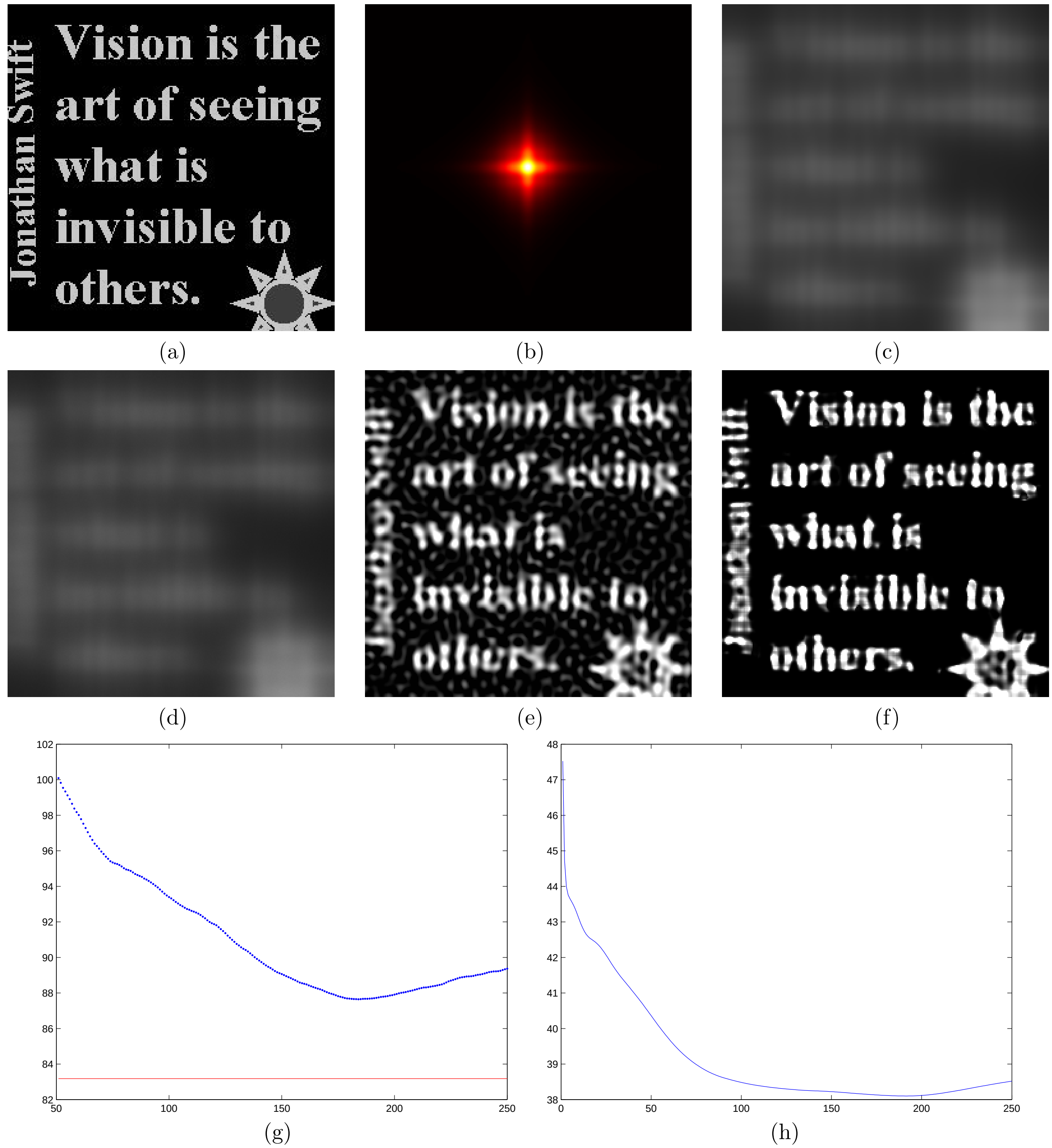

Figure 5: (a) The original image. (b) The PSF image. (c) The blurred image. (d) The noisy blurred image. (e) Initial guess from HyBR. (f) The result of EM-Type algorithm with Bregman iterations. (g) The plot of KL divergence versus Bregman iterations. (h) The RMSE versus Bregman iterations. Some parameters chosen are $\beta = 10^{-5}$, Num_inner = 10 and Num_outer = 250.

[5] 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), pp. 721–741.

[6] U. Grenander, "Tutorial in pattern theory", Lecture Notes Volume, Division of Applied Mathematics, Brown University, (1984).

[7] J. A. Conchello and J. G. McNally, "Fast regularization technique for expectation maximization algorithm

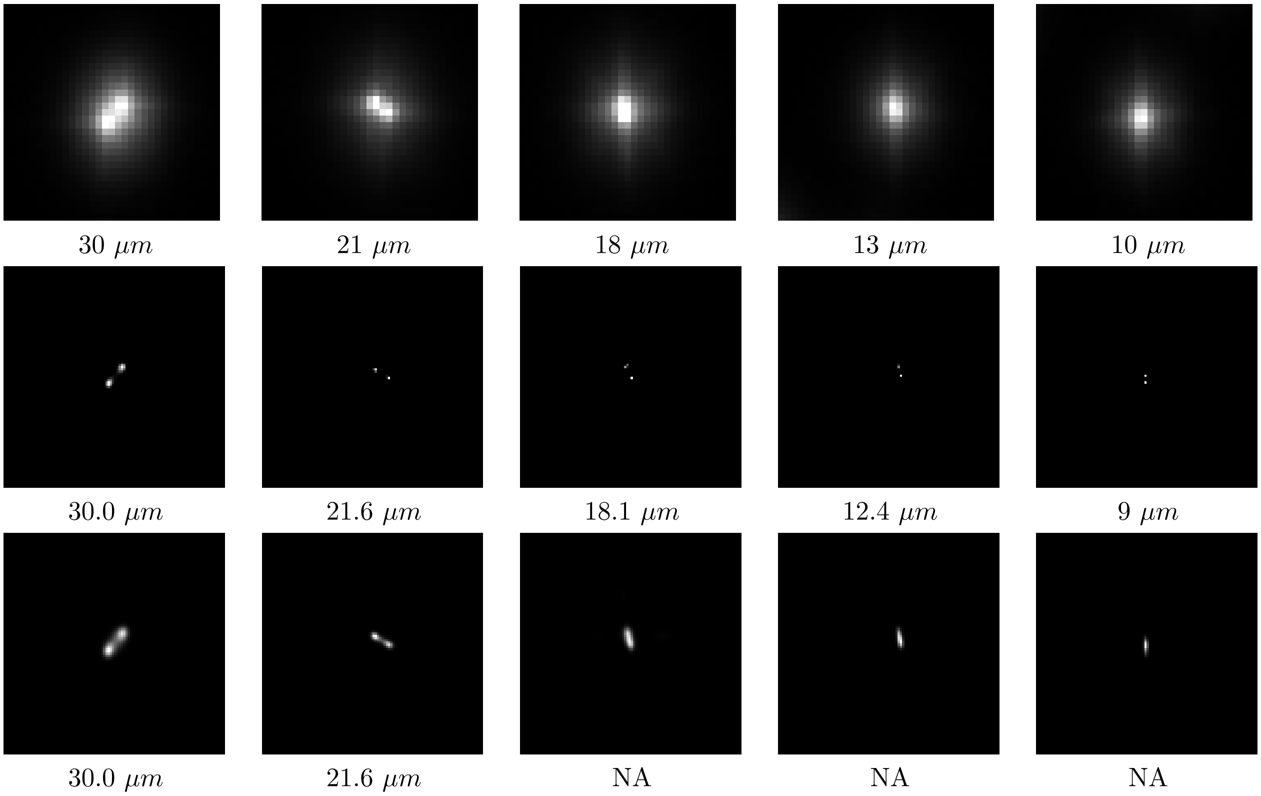

Figure 6: Top row shows raw lensfree fluorescent images of different pairs of particles. The distances betweens thes two particles are $30\mu m$, $21\mu m$, $18\mu m$, $13\mu m$ and $10\mu m$, from left to right. Middle row shows the results of EM-Type algorithm with $p = 0.5$, Bottom row shows the results for EM (or Richardson-Lucy) method only.

for optical sectioning microscopy", In: *Proceeding of SPIE Symposium on Electronic Imaging Science and Technology*, Volume 2655, (1996), pp. 199–208.

[8] 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), pp. 1052–1071.

[9] 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), pp. 319–331.

[10] L. Rudin, S. Osher and E. Fatemi, "Nonlinear total variation based noise removal algorithms", *Phys D,*, **60** (1992), pp. 259–268.

[11] L.I. Rudin and S. Osher, "Total variation based image restoration with free local constraints", In: *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, Volume 1, (nov 1994), pp. 31 –35 vol.1.

[12] 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), pp. 260–266.

[13] C. Brune, A. Sawatzky, F. Wubbeling, T. Kosters and M. Burger, "An analytical view on EM-TV based methods for inverse problems with Poisson noise", Preprint, University of Münster, (August 2009).

[14] 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*, Volume 7961, (2011), 79612X.

[15] S. Joshi and M. I. Miller, "Maximum a posteriori estimation with Good's roughness for optical sectioning microscopy", *Journal of the Optical Society of America A*, **10** (1993), pp. 1078–1085.

[16] A. N. Tychonoff and V. Y. Arsenin, *Solution of Ill-posed Problems*, Washington, Winston, (1977).

[17] 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), pp. 257–263.

[18] L. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography", *IEEE Transaction on Medical Imaging*, **1** (1982), pp. 113–122.

[19] M. Jung, E. Resmerita and L. A. Vese, "Dual norm based iterative methods for image restoration", *Journal of Mathematical Imaging and Vision*, to appear.

[20] R. Acar and C. R. Vogel, "Analysis of bounded variation penalty methods for ill-posed problems", *Inverse Problem*, **10** (1994), pp. 1217–1229.

[21] W. H. Richardson, "Bayesian-based iterative method of image restoration", *Journal of the Optical Society America*, **62** (1972), pp. 55–59.

[22] L. B. Lucy, "An iterative technique for the rectification of observed distributions", *Astronomical Journal*, **79** (1974), pp. 745–754.

[23] H. Kuhn and A. Tucker, "Nonlinear programming", In: *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, (1951), pp. 481–492.

[24] 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).

[25] 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), pp. 1–38.

[26] H. Hurwitz, "Entropy reduction in Bayesian analysis of measurements", *Physics Review A*, **12** (1975), pp. 698–706.

[27] E. Levitan and G. T. Herman, "A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography", *IEEE Transactions on Medial Imaging*, **6** (1987), pp. 185–192.

[28] Y. Meyer, *Oscillating Patterns in Image Processing and in some Nonlinear Evolution Equations*, American Mathematical Society, (2001).

[29] L. Bregman, "The relaxation method for finding common points of convex sets and its application to the solution of problems in convex programming", *USSR Computational Mathematics and Mathematical Physics*, **7** (1967), pp. 200–217.

[30] S. Osher, Y. Mao, B. Dong and W. Yin, "Fast linearized Bregman iteration for compressed sensing and sparse denoising", *Communications in Mathematical Sciences*, **8** (2010), pp. 93–111.

[31] W. Yin, S. Osher, D. Goldfarb and J. Darbon, "Bregman iterative algorithms for l1- minimization with applications to compressed sensing", *Journal on Imaging Sciences*, **1** (2008), pp. 143–168.

[32] J. Chung, J. Nagy and D. O'Leary, "A weighted GCV method for Lanczos hybrid regularization", *Electronic Transactions on Numerical Analysis*, **28** (2008), pp. 149–167.

[33] A. F. Coskun, I. Sencan, T. W. Su and A. Ozcan, "Lensless wide field fluorescent imaging on a chip using compressive decoding of sparse objects", *Optics Express*, **18** (2010), pp. 10510–10523.