SUPERIORIZATION OF EM ALGORITHM AND ITS APPLICATION IN SINGLE-PHOTON EMISSION COMPUTED TOMOGRAPHY (SPECT)

SHOUSHENG LUO AND TIE ZHOU

LMAM, School of Mathematical Sciences, Peking University
No.5 Yiheyuan Road Haidian District, Beijing, 100871, China

(Communicated by Hao-Min Zhou)

ABSTRACT. In this paper, we presented an efficient algorithm to implement the regularization reconstruction of SPECT. Image reconstruction with priori assumptions is usually modeled as a constrained optimization problem. However, there is no efficient algorithm to solve it due to the large scale of the problem. In this paper, we used the superiorization of the expectation maximization (EM) iteration to implement the regularization reconstruction of SPECT. We first investigated the convergent conditions of the EM iteration in the presence of perturbations. Secondly, we designed the superiorized EM algorithm based on the convergent conditions, and then proposed a modified version of it. Furthermore, we gave two methods to generate desired perturbations for two special objective functions. Numerical experiments for SPECT image reconstruction were conducted to validate the performance of the proposed algorithms. The experiments show that the superiorized EM algorithms are more stable and robust for noised projection data and initial image than the classic EM algorithm, and outperform the classic EM algorithm in terms of mean square error and visual quality of the reconstructed images.

1. Introduction. Single-photon emission computed tomography (SPECT), which can visualize the physiological information of various organs with the help of radiopharmaceuticals [1, 2, 3], a biochemical molecule labeled with radioactivity. The gamma-rays emitted by the injected radioactive material are recorded by a gamma camera rotated around the patient. The goal of SPECT is to reconstruct the radionuclide distribution from the measurements numerically. The variety of existing SPECT image reconstruction algorithms can be split into a family of analytical methods [4, 5, 6, 7, 8] and a wide class of iterative techniques [9, 10, 11].

From the analytic point of view, the SPECT image reconstruction problem [3, 4, 7] is to invert the attenuated Radon transform (aRt) of $f$ (distribution of radiopharmaceutical)

$$R_a f(s, \varphi) = \int_R f(s\theta + t\theta^\perp)e^{-\int_0^\infty \mu(s\theta + r\theta^\perp)dr}dt,$$  

(1)

where $\mu$ is a known function, referred to as the attenuation map of gamma-rays, $\theta = (\cos \varphi, \sin \varphi)$ and $\theta^\perp = (-\sin \varphi, \cos \varphi)$. In practice, $f$ and $\mu$ are two functions with compact support $\Omega$. Therefore, the integrand in (1) is zero outside a bounded
interval, and this integral is written over \((-\infty, +\infty)\) for convenience.

For the iterative methods, the SPECT image reconstruction problem is to solve the following linear system \([9]\),

\[
Ax = b,
\]

where the elements of the observed data \(b = (b_1, b_2, \ldots, b_M)^t \in \mathbb{R}^M\), the unknown image \(x = (x_1, x_2, \ldots, x_N)^t \in \mathbb{R}^N\) and the system matrix \(A = (a_{ij}) \in \mathbb{R}^{M \times N}\) are all nonnegative. Here and in the following, the superscript \(t\) denotes the transpose of vectors. The aim is to reconstruct the unknown \(x\) as an image from the projection data \(b\) via stable algorithms. A solution is not feasible with conventional methods directly because of the noisy projection data \(b\), the ill-posedness and large scale of the problem. For the SPECT image reconstruction, the system matrix \(A\) not only can model the attenuation of the gamma-rays, but also can fuse some realistic factors, such as photon scattering and camera blurring.

The expectation maximization (EM) algorithm \([9, 12]\) and the algebraic reconstruction technique (ART)\([13, 14]\) are two widely used technologies in imaging sciences, due to their simplicity, efficiency and performance. In practice, the EM algorithm is more appropriate for emission tomography including SPECT. Firstly, the EM algorithm maintains the nonnegative constraint in the iteration procedure. Secondly, the EM algorithm is relatively robust against data inconsistencies introduced by Poisson noise, because it seeks to minimize the Kullback-Liebler (K-L) distance between the measured data \(b\) and the projection of the estimated image \(Ax\), which is equivalent to maximizing the likelihood of Poisson distribution.

For SPECT image reconstruction, it is one of the main issues to estimate the radionuclide distribution from low-counts projection data. This issue occurs quite frequently because of practical constraints, such as imaging hardware and scanning geometry. Furthermore, in order to reduce acquisition time, radiation dose and imaging cost efficiently, we should decrease the counts of the projection data. However, this would cause the strong deterioration of the observed data and the under-determinacy \((m \ll n)\) of the linear system (2). In these situations, the reconstructed images by the EM algorithm are usually dominated by various distortions, because the EM algorithm accepts any solution which minimizes the K-L distance and depends on the initial point.

The qualities of the reconstructed images can be improved by regularization methods. Regularization methods are often used techniques to improve the quality of reconstructed images. There are two regularization reconstruction models: unconstraint optimization\([15, 16, 17, 18, 19, 20]\)

\[
\min_x \left\{ \phi(x) + \frac{1}{\lambda} D(b, Ax) \right\}
\]

and constraint optimization \([21, 22, 23]\)

\[
\min \phi(x) \quad \text{s.t.} \quad x \in \{x | D(b, Ax) < \epsilon\}
\]

where \(\phi\) is a convex function, representing the prior knowledge. \(D(\cdot, \cdot)\) denotes a distance function, such as \(l^2\) and K-L distances. The parameters \(\lambda, \epsilon\) are nonnegative and related to the noise level. The lower the noise level is, the smaller the parameters are.

For the first optimization problem (3), there are different algorithms \([17, 18, 20, 24]\). In this paper, we focus on the second optimization problem (4). To our knowledge, there is no optimal algorithm to solve the constraint optimization problem
Superiorization of EM and Its Application in SPECT

(4) efficiently due to the large scale. In this paper, we resort to an emerging approach called superiorization [25] to implement the regularization reconstruction of SPECT.

The superiorization of iterative methods, which was first proposed by the authors of [21], is a relaxation technology for the constrained optimization problem. The superiorized algorithm lies between the feasibility-seeking algorithms, which seek a feasible point in the constrained set, and the optimal algorithms, which seek the minimum point of objective function in the constrained set. The aim of superiorization algorithm is to look for a superior instead of the optimal point of the objective function or just a feasible point in the constrained set. The basic idea of superiorization is to do the feasibility-seeking algorithms with perturbations about the objective function.

The superiorization of ART algorithms has been studied and applied to the regularization reconstruction of computed tomography (CT) [21, 22, 25, 26, 27]. The authors of [21, 26] first investigated the convergence of the two variants of ART under summable perturbations for consistent case. For inconsistent case, the authors of [27] proved the convergence of symmetric version of ART in the presence of summable perturbations. The superiorization of the EM algorithm was firstly proposed in [28], and applied to bioluminescence tomography. However, to our knowledge, it is still an open problem about the convergence of the superiorized EM algorithm.

In this paper, we first discussed the convergence of the EM algorithm in the presence of perturbation in section 2. A so-called bounded perturbation resilient (BPR) property of ART is vital in the proof of convergence for the perturbed version of ART. However, we cannot prove the BPR property of the EM iteration so far, because of the nonlinearity of the EM operator. Therefore, we investigated the convergence of the perturbed EM algorithm under the following assumptions. Firstly, the perturbations should maintain the positivity of iterations. Secondly, the perturbations should go to zero with the increase of iterates. Lastly, the perturbed EM iteration should gradually decrease the K-L distance between the observed data and the projection of estimation by each iteration.

Based on the convergent conditions, we presented the superiorized EM algorithm and its modified version in section 3. Furthermore, practicable techniques were given to produce ideal perturbations for two special objective functions, total variation (TV) [29] and $l^1$-norm [30], widely used in imaging sciences. While the proposed algorithms are applicable to diverse inverse problems, in this paper we restricted ourselves to demonstrate its usefulness to the SPECT image reconstruction. Numerical results for SPECT image reconstruction were given in section 4. As our expectation, the superiorization algorithms output superior image comparing with the classic EM algorithm in terms of MSE and visual quality. Some conclusions and discussions were given in section 5.

2. Perturbations resilience of EM iteration. For the sake of reference, we first introduce some notations and assumptions. In this paper, an image $x$ is described as a vector of length $N$ with individual elements $x_j$, $j = 1, 2, \cdots, N$. When it is necessary to refer to pixels in the context of a two-dimensional (2D) image we use the double subscript form $x_{p,q}$, where

\begin{equation}
    j = (q - 1)W + p, p = 1, 2, \cdots, H, q = 1, 2, \cdots, W,
\end{equation}

and integers $W$ and $H$ are, respectively, the width and height of the 2D image array, which has a total number of pixels $N = W \times H$. Denote by $\mathbb{R}^N_+$ the region
\(\{x \geq 0 | x \in \mathbb{R}^N\}\). For the system matrix \(A\), we set \(H_j = \sum_{i=1}^{M} a_{ij}\) and \(d_i(x) = \sum_{j=1}^{N} a_{ij} x_j\) for convenience. Furthermore, we introduce

\[
ge_{ij}(x) = \frac{a_{ij} b_i}{d_i(x)},
\]

and

\[
f_j(x) = \frac{1}{H_j} \sum_{i=1}^{M} \frac{a_{ij} b_i}{d_i(x)} = \frac{1}{H_j} \sum_{i=1}^{M} g_{ij}(x).
\]

Let \(P\) denote the EM operator, and then the EM iteration \(x^{k+1} = P(x^k) = (P_1(x^k), \ldots, P_N(x^k))^t\) for the problem (2) is defined as

\[
x_{j}^{k+1} = P_j(x^k) = \frac{x_{j}^{k}}{H_j} \sum_{i=1}^{M} \frac{b_i a_{ij}}{\sum_{t=1}^{N} a_{it} x_{t}^{k}} = x_{j}^{k} \cdot f_j(x^k),
\]

with an initial image \(x^0 > 0\). Similarly, the perturbed version of the EM iteration is defined as

\[
x_{j}^{k+1} = P_j(x^k + \beta_k v^k) = \frac{y_{j}^{k}}{H_j} \sum_{i=1}^{M} \frac{b_i a_{ij}}{\sum_{t=1}^{N} a_{it} y_{t}^{k}} = y_{j}^{k} \cdot f_j(y^k),
\]

where \(y^k = x^k + \beta_k v^k\). Here, the vector \(v^k \in \mathbb{R}^N\) and number \(\beta_k \geq 0\) represent the direction and length of the perturbation of the \(k\)th iteration, respectively. Hereafter, we call the EM iteration without perturbation as the classic EM iteration to distinguish from each other.

It has been established that the sequence \(\{x^n\}\) generated by the classic EM iteration converges to a minimizer of the K-L distance \(I_B^k(x)\) between \(b\) and \(Ax\) on \(R^N_+\), where \(I_B^k(x)\) is defined as

\[
I_B^k(x) = I(b, Ax) = \sum_{i=1}^{M} b_i \ln \frac{b_i}{d_i(x)} - \sum_{i=1}^{M} (b_i - d_i(x)),
\]

where \(I(\cdot, \cdot)\) denote the K-L distance function of any two nonnegative vectors. An important inequality used in this paper is

\[
\ln t \geq 1 - \frac{1}{t},
\]

for \(t > 0\), and the inequality holds with equality if and only if \(t = 1\). By using the inequality (11), we have that \(I(x, y) \geq 0\) for all vectors \(x, y \geq 0\), and the inequality holds with equality if and only if \(x = y\).

Obviously, if \(\beta_k = 0\), the perturbed EM iteration is the same as the classic EM iteration, in which we are not interested. Therefore, in the following we assume \(\beta_k > 0\). A natural question is that under what assumptions the sequence \(\{x^n\}\) generated by the perturbed EM iteration (9) converges to a minimizer of (10) as well. Before discussing the convergent conditions of the perturbed EM iteration, we first summarize some propositions of the classic EM iteration [9, 12, 31]. Without loss of generality, we assume that all the elements \(b_i > 0\) for all \(i\), and \(H_j = \sum_{i=1}^{M} a_{ij} = 1\) for all \(j\) for simplicity.

**Theorem 2.1.** For any initial image \(x^0 > 0\), denote by \(x^k\) the estimate of the classic EM algorithm after \(k\) iterations. Then the following propositions hold:

1. \(x^k > 0\) and \(\sum_{j=1}^{N} x_{j}^{k} = \sum_{i=1}^{M} b_i\) for all \(k > 0\).
2. \( I_A^b(x^{k+1}) \leq I_A^b(x^k) \), and \( I(x^{k+1}, x^k) \leq I_A^b(x^k) - I_A^b(x^{k+1}) \).
3. \( \{x^k\} \) converges to a minimizer \( x^* \) of \( I_A^b(x) \) on \( \mathbb{R}_+^N \), and \( x^* \) is a fixed point of the EM operator.
4. \( I(x^*, x^{k+1}) \leq I(x^*, x^k) \).

The inequalities in second and fourth items above hold with equalities if and only if \( x^k \) is a fixed point of the EM operator.

**Remark 1.** Obviously, \( x \) is a fixed point of the EM operator if and only if \( f_j(x) = 1 \) for \( x_j \neq 0 \).

From the propositions above, we have that the sequence \( \{I_A^b(x^k)\} \) monotonically converges to the minimum of \( I_A^b(\cdot) \) on \( \mathbb{R}_+^N \), and \( \{x^k\} \) approximates to the minimizer \( x^* \) gradually. Next, we investigate the convergence of the perturbed EM iteration, and prove the similar propositions as far as we can.

**Theorem 2.2.** Given any initial image \( x^0 \), denote by \( \{x^k\}_{k \in \mathbb{N}} \) the sequence generated by the perturbed EM iteration (9).

\[
\begin{align*}
\tag{12}
x_j^{k+1} &= (x_j^k + \beta_k v_j^k) \cdot f_j(x^k + \beta_k v^k) = y_j^k \cdot f_j(y^k),
\end{align*}
\]

where \( y_j^k = x_j^k + \beta_k v_j^k \), and \( \{v^k\} \) is a bounded sequence. Suppose that

1. **Positivity:** \( y_j^k > 0 \).
2. **Vanishing:** \( \beta_k \rightarrow 0 \).
3. **Decreasing:**

\[
\tag{13}
\beta_k \max_{j \in S_k^-} \{ -\frac{v_j^k}{y_j^k} \} B_k^- - \beta_k \min_{j \in S_k^+} \{ -\frac{v_j^k}{y_j^k} \} B_k^+ + \beta_k \sum_{j=1}^N v_j^k \ < \ I_A^b(y^k) - I_A^b(x^{k+1}),
\]

where \( B_k^- = \max \{ \rho + \sum_{j \in S_k^-} \lambda_j x_j^{k+1}, \sum_{j \in S_k^-} \hat{x}_j \} \), \( B_k^+ = \min \{ \sum_{j \in S_k^+} \lambda_j x_j^{k+1}, \sum_{j \in S_k^+} \hat{x}_j \} \), and \( S_k^- = \{ j | v_j^k < 0 \}, \ S_k^+ = \{ j | v_j^k > 0 \} \). Here \( \rho \) is a sufficiently small positive number. Then, the following propositions hold:

1. \( I_A^b(x^{k+1}) \leq I_A^b(x^k) \).
2. \( \{x^k\} \) has a convergent subsequence \( \{x^{m_k}\} \), and the limit \( \hat{x} \) is a fixed point of the EM operator.
3. \( I(\hat{x}, x^{k+1}) \leq I(\hat{x}, x^k) \), and \( x^k \rightarrow \hat{x} \).
4. \( \hat{x} \) is a minimizer of \( I_A^b(\hat{x}) \).

Furthermore, the inequalities above hold with equalities iff \( x^k \) is a fixed point of the EM operator and \( \beta_k v^k = 0 \).

Before presenting the proof, we explain the necessities of the conditions of theorem 2.2. The positive condition of \( y_j^k \) is necessary for the nonnegative constraints. The second condition is required by the convergence of the sequence \( \{x^k\} \). Intuitively, the last condition is used to guarantee that \( I_A^b(x^{k+1}) \leq I_A^b(x^k) \), which implies the convergence of \( I_A^b(x^k) \).

An important concern is about the existence of the perturbations satisfying the conditions of theorem 2.2, especially the third condition. Because the sequence \( \{\hat{x}^k\} \) generated by the classic EM iteration (\( \beta_k = 0 \)) satisfies the conditions in theorem 2.2, the sequence \( \{x^k\} \) generated by the perturbed version also satisfies them when \( \beta_k \) is small enough for each iteration due to the continuities of the EM operator (9) and K-L distance (10). Therefore, there exist perturbations which satisfy the assumptions of theorem 2.2.
Proof. We prove this theorem step by step, which follows the proving procedure in [31].

- **Proof of proposition 1:** By the definition of \( P_h^b(\cdot) \) and \( d_i(\cdot) \), we have

\[
I_A^b(x^k) - I_A^b(x^{k+1})
\]

\[
= \sum_{i=1}^{M} b_i \ln \frac{b_i}{d_i(x^k)} - \sum_{i=1}^{M} b_i \ln \frac{b_i}{d_i(x^{k+1})}
\]

\[
= \sum_{i=1}^{M} b_i \ln \frac{d_i(x^{k+1})}{d_i(x^k) - \beta_k d_i(v^k)}
\]

\[
= \sum_{i=1}^{M} b_i \ln \frac{d_i(x^k)}{d_i(x^k)} + \sum_{i=1}^{M} b_i \ln \frac{1}{1 - \beta_k d_i(v^k)/d_i(x^k)}
\]

(14) \[
\geq I_A^b(y^k) - I_A^b(x^{k+1}) - \beta_k \sum_{j=1}^{N} v_j^k + \sum_{i=1}^{M} b_i \left[ 1 - \left( 1 - \frac{d_i(v^k)}{d_i(x^k)} \right) \right]
\]

\[
= I_A^b(y^k) - I_A^b(x^{k+1}) - \beta_k \sum_{j=1}^{N} v_j^k + \sum_{i=1}^{M} b_i \beta_k \frac{d_i(v^k)}{d_i(x^k)}
\]

\[
\geq \beta_k \max_{j \in S_k^+} \left\{ -\frac{v_j^k}{y_j^k} \right\} B_k^+ - \beta_k \min_{j \in S_k^-} \left\{ \frac{v_j^k}{y_j^k} \right\} B_k^- + \beta_k \sum_{j=1}^{N} \frac{v_j^k}{y_j^k} \sum_{i=1}^{M} a_{ij} b_i
\]

\[
= \beta_k \max_{j \in S_k^+} \left\{ -\frac{v_j^k}{y_j^k} \right\} B_k^+ - \beta_k \min_{j \in S_k^-} \left\{ \frac{v_j^k}{y_j^k} \right\} B_k^- + \beta_k \sum_{j=1}^{N} \frac{v_j^k}{y_j^k} x_j^{k+1}
\]

\[
\geq \beta_k \max_{j \in S_k^+} \left\{ -\frac{v_j^k}{y_j^k} \right\} (B_k^- - \sum_{j \in S_k^+} x_j^{k+1}) - \beta_k \min_{j \in S_k^-} \left\{ \frac{v_j^k}{y_j^k} \right\} (B_k^+ - \sum_{j \in S_k^-} x_j^{k+1})
\]

(15) \[
\geq \beta_k \max_{j \in S_k^+} \left\{ -\frac{v_j^k}{y_j^k} \right\} \rho - \beta_k \min_{j \in S_k^-} \left\{ \frac{v_j^k}{y_j^k} \right\} (B_k^- - \sum_{j \in S_k^+} x_j^{k+1}).
\]

The first and second inequalities hold because of the inequality (11) and the condition (13), respectively. Therefore, we have proved that \( I_A^b(x^{k+1}) \leq I_A^b(x^k) \) by (13) and (15).

Finally, we should prove that \( I_A^b(x^{k+1}) = I_A^b(x^k) \) iff \( \beta_k v^k = 0 \) and \( x^k \) is a fixed point of EM operator. The sufficiency is obvious by theorem 2.1. The necessity is also true by the following facts. If \( I_A^b(x^{k+1}) = I_A^b(x^k) \), we have that all the inequalities in the derivation above hold with equalities, which implies that \( \frac{1}{1 - \beta_k d_i(v^k)/d_i(x^k)} = 1 \) for all \( i \) and \( S_k^- = 0 \) by inequalities (14) and (15), respectively, i.e. \( \beta_k d_i(v^k) = 0 \) and \( v_j^k \geq 0 \). In fact, we have that \( v_j^k = 0 (\forall j) \), i.e. \( y_j^k = x_j^k \), and \( x^k \) is a fixed point by theorem 2.1. Otherwise, assume \( v_j^k > 0 \) for some \( j_0 \), there must exist \( i_0 \) such that \( a_{i_0 j_0} > 0 \) and \( d_{i_0}(v^k) = \sum_{j=1}^{N} a_{i_0 j} v_j^k \geq a_{i_0 j_0} v_{j_0} > 0 \) by the assumption \( \sum_{i=1}^{M} a_{ij} = 1 \neq 0 \), which is contradict to \( d_i(v^k) = 0 (\forall i) \).
• **Proof of proposition 2:** Because $0 < \sum_{j=1}^{N} x_j^k = \sum_{i=1}^{M} b_i + \text{(constant)}$ and $x^k > 0$, the sequence $\{x^k\}$ has a convergent subsequence $\{x^{m_k}\}$. Denoting by $\hat{x}$ the limit of $\{x^{m_k}\}$, we have $y^{m_k} \rightarrow \hat{x}$ because $\beta_k \rightarrow 0$ and $u^k$ is bounded. Next we prove $\hat{x}$ is a fixed point of the EM operator. To this end, we define a function for $x \geq 0$

$$D(x) = I(P(x), x).$$

By the second proposition of theorem 2.1, we have

$$D(y^{m_k}) = I(P(y^{m_k}), y^{m_k}) = I(x^{m_k+1}, y^{m_k}) \leq I_A^b(y^{m_k}) - I_A^b(x^{m_k+1}).$$

Due to the proposition 1 of theorem 2.2, $\{I_A^b(x^k)\}$ is a convergent sequence. Furthermore, $\{I_A^b(y^k)\}$ converges to the same limit as $\{I_A^b(x^k)\}$ because of the continuity of $I_A^b(\cdot)$ and $\beta_k \rightarrow 0$. Therefore, we have $I_A^b(y^{m_k}) - I_A^b(x^{m_k+1}) \rightarrow 0$, $m_k \rightarrow \infty$. Thus, we have that

$$D(\hat{x}) = \lim_{m_k \rightarrow \infty} D(y^{m_k}) = I(P(\hat{x}), \hat{x}) = 0,$$

i.e. $\hat{x}$ is a fixed point of the EM operator, since $I(x, y) = 0 \iff x = y$. Therefore, we have $f_j(\hat{x}) = 1$ if $\hat{x}_j \neq 0$ by the EM iteration formula.

• **Proof of proposition 3:** Before going further, we first prove a general inequality that $I(\hat{x}, x) - I(\hat{x}, P(x)) \geq I_A^b(\hat{x}) - I_A^b(\hat{x})$ for any point $x > 0$. Assuming $z = P(x)$, we have

$$I(\hat{x}, x) - I(\hat{x}, z) = \sum_{j=1}^{N} \hat{x}_j \ln \frac{\hat{x}_j}{x_j} - \sum_{j=1}^{N} (\hat{x}_j - x_j) + \sum_{j=1}^{N} \hat{x}_j \ln \frac{\hat{x}_j}{z_j} - \sum_{j=1}^{N} (\hat{x}_j - z_j) - \sum_{j=1}^{N} \hat{x}_j \ln \frac{z_j}{x_j} - \sum_{j=1}^{N} (\hat{x}_j - x_j) + \sum_{j=1}^{N} \hat{x}_j \ln \left(\frac{z_j}{x_j} g(x_j)\right) - \sum_{j=1}^{N} (\hat{x}_j - x_j)

\geq - \sum_{j=1}^{N} \hat{x}_j \ln \left(\frac{z_j}{x_j} g(x_j)\right) - \sum_{j=1}^{N} \hat{x}_j \ln \left(\frac{z_j}{x_j} g(x_j)\right) - \sum_{j=1}^{N} \hat{x}_j \ln \left(\frac{z_j}{x_j} g(x_j)\right) - \sum_{j=1}^{N} (\hat{x}_j - x_j) - \sum_{j=1}^{N} \hat{x}_j \ln \left(\frac{z_j}{x_j} g(x_j)\right) - \sum_{j=1}^{N} \hat{x}_j \ln \left(\frac{z_j}{x_j} g(x_j)\right) - \sum_{j=1}^{N} (\hat{x}_j - x_j)

= \sum_{i=1}^{M} b_i \ln \frac{d_i(\hat{x})}{d_i(x)} - \sum_{j=1}^{N} (\hat{x}_j - x_j) - \sum_{i=1}^{M} b_i \ln \frac{d_i(\hat{x})}{d_i(x)} - \sum_{j=1}^{N} (\hat{x}_j - x_j)

= \sum_{i=1}^{M} b_i \ln \frac{d_i(\hat{x})}{d_i(x)} - \sum_{j=1}^{N} (\hat{x}_j - x_j) - \sum_{i=1}^{M} b_i \ln \frac{d_i(\hat{x})}{d_i(x)} - \sum_{j=1}^{N} (\hat{x}_j - x_j)$$
\begin{equation}
\begin{aligned}
I(x) &= I(x) + \sum_{i=1}^M (d_i(x) - d_i(x)) - \sum_{j=1}^N (\hat{x}_j - x_j) \\
\Rightarrow \quad I(x) &= I(x) - I(\hat{x}) ,
\end{aligned}
\end{equation}

where we used the relationship \( \sum_{j=1}^N \hat{x}_j = \sum_{j=1}^N x_j = \sum_{i=1}^M b_i = \sum_{j=1}^N z_j \). The inequality and the second equality are implied by Jensen’s inequality and the fact \( f_j(\hat{x}) = \sum_{i=1}^M g_{ij}(\hat{x}) = 1 \) for \( \hat{x} \neq 0 \), respectively. The last equality follows the fact that

\begin{equation}
\sum_{i=1}^M d_i(x) = \sum_{i=1}^M \sum_{j=1}^N a_{ij} x_j = \sum_{j=1}^N x_j ,
\end{equation}

by the assumption that \( \sum_{i=1}^M a_{ij} = 1 \). Let \( x = y^k \), then \( z = P(y^k) = x^{k+1} \) and \( I(\hat{x}, y^k) - I(\hat{x}, x^{k+1}) \geq I_A(y^k) - I_A(\hat{x}) \) by (19). Furthermore, we have

\begin{equation}
\begin{aligned}
I(\hat{x}, y^k) - I(\hat{x}, x^{k+1}) &\geq I_A(y^k) - I_A(\hat{x}) \\
&= \left[ I_A(y^k) - I_A(x^{k+1}) \right] + I_A(x^{k+1}) - I_A(\hat{x}) \\
\geq & I_A(y^k) - I_A(x^{k+1}).
\end{aligned}
\end{equation}

The last inequality holds because \( I_A(x^{k+1}) \geq I_A(\hat{x}) \). Now, we prove \( I(\hat{x}, x^{k+1}) \leq I(\hat{x}, x^k) \). By the definitions of the K-L distance and the perturbed EM iteration, we have

\begin{equation}
\begin{aligned}
I(\hat{x}, x^k) &= I(\hat{x}, x^{k+1}) \\
&= \sum_{j=1}^N \hat{x}_j \ln \frac{(x_j^k + \beta_k v_j^k) f_j(y^k)}{x_j^k} \\
&= \sum_{j=1}^N \hat{x}_j f_j(y^k) + \sum_{j=1}^N \hat{x}_j \ln(1 + \beta_k x_j^k) \\
\geq & \sum_{j=1}^N \hat{x}_j f_j(y^k) y_j^k + \sum_{j=1}^N \hat{x}_j \left( 1 - \frac{1}{1 + \beta_k v_j^k/x_j^k} \right) \\
&= I(\hat{x}, y^k) - I(\hat{x}, x^{k+1}) + \sum_{j=1}^N \hat{x}_j \frac{\beta_k v_j^k}{x_j^k + \beta_k v_j^k} - \beta_k \sum_{j=1}^N v_j^k \\
\geq & I_A(y^k) - I_A(x^{k+1}) \\
&= -\beta_k \max_{j \in S_k^-} \left\{ -\frac{v_j^k}{y_j^k} \right\} \sum_{j \in S_k^-} \hat{x}_j + \beta_k \min_{j \in S_k^+} \left\{ \frac{v_j^k}{y_j^k} \right\} \sum_{j \in S_k^+} \hat{x}_j - \beta_k \sum_{j=1}^N v_j^k \\
&\geq \beta_k \max_{j \in S_k^-} \left\{ -\frac{v_j^k}{y_j^k} \right\} (B_k^- - \sum_{j \in S_k^-} \hat{x}_j) + \beta_k \min_{j \in S_k^+} \left\{ \frac{v_j^k}{y_j^k} \right\} \left( \sum_{j \in S_k^+} \hat{x}_j - B_k^+ \right)
\end{aligned}
\end{equation}
Proof of proposition 4:

•

Therefore, we have proved that \( I_{A} = I_{B} \) since \( \beta_{k} v_{j}^{k} = 0 \) and \( I_{A}^{b}(y^{k}) = I_{B}^{b}(x^{k+1}) \) by inequalities (22) and (23), respectively. Therefore, we have that \( y^{k} = x^{k} \) and \( x^{k} \) is a fixed point of EM operator by theorem 2.1.

Finally, we need to prove that \( x^{k} \rightarrow \hat{x} \). Firstly, we have \( I(\hat{x}, x^{k}) \rightarrow 0 \) since \( I(\hat{x}, x^{mk}) \rightarrow 0 \) and the sequence \( \{I(\hat{x}, x^{k})\} \) monotonously decreases. Therefore, we have proved \( x^{k} \rightarrow \hat{x} \) by \( I(x, y) = 0 \iff x = y \), and \( y^{k} \rightarrow \hat{x} \) as well by \( \beta_{k} v_{j}^{k} \rightarrow 0 \).

• Proof of proposition 4: In order to prove this proposition, we should prove that \( \hat{x} \) satisfies the Kuhn-Tucker (K-T) conditions. For the EM algorithm and the K-L function, the K-T conditions are equivalent to (1) \( f_{j} (\hat{x}) = 1 \) if \( \hat{x}_{j} \neq 0 \), and (2) \( 0 \leq f_{j} (\hat{x}) \leq 1 \) if \( \hat{x}_{j} = 0 \) [9, 31]. Obviously, the first condition holds because \( \hat{x} \) is a fixed point of the EM operator. For the second condition, the nonnegativity of \( f_{j} (\hat{x}) \) is satisfied since \( x_{j}^{k} \rightarrow \hat{x}_{j} \geq 0 \), \( b_{i} \geq 0 \) and \( a_{ij} \geq 0 \).

Next we need to prove that \( f_{j} (\hat{x}) \leq 1 \) for \( \hat{x}_{j} = 0 \). Suppose \( f_{j} (\hat{x}) \geq 1 + \epsilon_{1} > 1 \) and \( \hat{x}_{j} = 0 \) for some \( j \). There exists a sufficiently large integer \( L > 0 \) such that \( f_{j} (y^{k}) > 1 + \epsilon_{1} \) for all \( k \geq L \) since \( y^{k} \rightarrow \hat{x} \). If \( x_{j}^{k} \rightarrow 0 \), there must be infinite iterations such that \( x_{j}^{k+1} < x_{j}^{k} \). In the following, we assume that \( k > L \). For each iteration satisfying \( x_{j}^{k+1} = y_{j}^{k} \), \( f_{j} (y^{k}) < x_{j}^{k} \), we have that \( y_{j}^{k} < x_{j}^{k+1} < x_{j}^{k} \) since \( f_{j} (y^{k}) > 1 \), and

\[
1 > \frac{(x_{j}^{k} + \beta_{k} v_{j}^{k}) f_{j} (y^{k})}{x_{j}^{k}}
\]

(24)

\[
= (1 + \beta_{k} \frac{y_{j}^{k}}{x_{j}^{k}}) f_{j} (y^{k}).
\]

By the assumption that \( f_{j} (y^{k}) > 1 + \epsilon_{1} \), we have

\[
1 + \beta_{k} \frac{y_{j}^{k}}{x_{j}^{k}} < \frac{1}{1 + \epsilon_{1}}.
\]

(25)

Abstracting 1 on both sides of (25), we have \( \beta_{k} \frac{y_{j}^{k}}{x_{j}^{k}} < - \frac{1}{1 + \epsilon_{1}} \). By equation (15), we have that

\[
I_{A}^{b}(x^{k}) - I_{A}^{b}(x^{k+1}) \geq \beta_{k} \max_{j \in S_{k}} \{ - \frac{y_{j}^{k}}{y_{j}^{k}} (B_{j}^{-} - \sum_{j \in S_{k}^{+}} x_{j}^{k+1}) - \beta_{k} \min_{j \in S_{k}} \{ \frac{y_{j}^{k}}{y_{j}^{k}} (B_{j}^{+} - \sum_{j \in S_{k}} x_{j}^{k+1})
\}
\]

\[
\geq \frac{1 + \epsilon_{1}}{1 + \epsilon_{1}} (B_{k}^{-} - \sum_{j \in S_{k}} x_{j}^{k+1})
\]

\[
(26)
\]

\[
\geq \rho \frac{1 + \epsilon_{1}}{1 + \epsilon_{1}}.
\]
The second inequality holds since \( y_j^k < x_j^k \) and \( B_k^+ \leq \sum_{j \in S} x_j^{k+1} \). The last inequality follows from the definition of \( B_k^- \).

From equation (26), we can see that the decrease of \( I^b_A(x^k) \) is larger than a positive number for each iteration satisfying \( x_j^{k+1} < x_j^k \). Furthermore, because the number of iterations satisfying \( x_j^{k+1} < x_j^k \) is infinite and \( I^b_A(x^k) \) is finite, we have \( I^b_A(x^l) \to -\infty \), which is contradict to the fact that \( I^b_A(x) \geq 0 \) for all \( x \geq 0 \).

3. Superiorization of the EM algorithm. The SPECT regularization reconstruction can be modeled as a constrained optimization problem

\[
\min_{x \in E} \phi(x), \quad E = \{x^*| x^* = \arg \min_{x \geq 0} I^b_A(x)\},
\]

where \( \phi \) is a convex function, which assigns each image \( x \) a number indicating the "undesirability" of the image in some sense. The set \( E \) is called feasible set, and it is called feasible problem to look for a point in \( E \).

To our knowledge, there is no efficient algorithm to deal with the constrained optimization problem (27) because of the large scale of it for SPECT image reconstruction. On the other hand, although the feasible problem is also a constrained optimization, we can solve it by the classic EM algorithm [9] and its variants [10, 11] efficiently. Based on the facts above, we use the superiorization methodology to implement the SPECT regularization reconstruction.

For the objective function \( \phi \), the superiorized EM algorithm is illustrated as algorithm 1 based on the conditions of theorem 2.2. In order to emphasize the objective function \( \phi \) for which we are superiorizing, we refer to the superiorized algorithm as the \( \phi \)-superiorization of the EM iteration.

**Algorithm 1** Framework of \( \phi \)-superiorization algorithm

| Initialization: \( \beta_0 > 0, x^0 > 0, k = 0, \text{ and } 0 < \gamma < 1. \) |
| repeat: logic=true |
| while logic |
| find a decreasing direction \( v^k \) of \( \phi \) at \( x^k \), such that \( y_j^k = x_j^k + \beta_k v_j^k > 0 \). |
| If \( \phi(y^k) \leq \phi(x^k) \) and inequality (13) holds, \( (\ast) \) |
| logic=false, \( x^{k+1} = P(y^k), \beta_{k+1} = \beta_k, k = k + 1. \) |
| end(if) |
| \( \beta_k = \gamma \beta_k. \) |
| end(while) |

Because the perturbation direction \( v^k \) is selected as the decreasing direction of \( \phi \) at \( x^k \), the size of \( \beta_k \) represents the strength of regularization in some sense. Because the condition 3 of theorem 2.2 is very strict, the numerical experiments show that \( \{\beta_k\} \) goes to zero very fast, which results in the regularization is very weak. Therefore, we propose a modified version of algorithm 1, which is shown in algorithm 2. In algorithm 2, we only validate \( I^b_A(x^{k+1}) < I^b_A(x^k) \), rather than the inequality (13) of theorem 2.2. Since the inequality (13) implies \( I^b_A(x^{k+1}) < I^b_A(x^k) \) by theorem 2.2, algorithm 2 can be seen as a relaxation of algorithm 1. Furthermore, we introduce a relative decrease of \( I^b_A(\cdot) \) to avoid the situation that
In order to confirm the inequality (13) of theorem 2.2 in algorithm 1, we should compute $\sum_{j \in S_k^-} \tilde{x}_j$ and $\sum_{j \in S_k^+} \tilde{x}_j$ for each iteration. However, $\tilde{x}$ is unknown in the iterative procedure. In practice, we estimate the values $\sum_{j \in S_k^-} \hat{x}_j$ and $\sum_{j \in S_k^+} \hat{x}_j$ by $\sum_{j \in S_k^-} \hat{x}_j \approx \frac{|S_k^-|}{N} B$, and $\sum_{j \in S_k^+} \hat{x}_j \approx \frac{|S_k^+|}{N} B$, where $|S_k^-|$ and $|S_k^+|$ denote the cardinalities of the sets $S_k^-$ and $S_k^+$, respectively. Here, the capital letters $N$ and $B = \sum_{i=1}^{M} b_i$ denote the length of $x$ and the total counts. Lastly, the parameter $\rho$ is chosen as $10^{-6}$ in this paper.

In the following, we will discuss how to generate desirable perturbation $\beta_k v^k$ or $y^k$ for two concrete objective functions, TV and $l^1$-norm, such that the perturbations satisfy the conditions $(\ast)$ and $(\ast)$ of the two $\phi$-superiorization algorithms. Since the TV regularization allows the reconstructed image to have sharp edges, TV based models are widely used in imaging sciences [15, 16, 26, 29]. For an $H \times W$ image $x$ whose pixel values are denoted by $x_{i,j}$, the TV of $x$ is defined as

$$TV(x) = \sum_{i=1}^{H-1} \sum_{j=1}^{W-1} \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2},$$

where $H, W$ are the height and width of $x$. In order to reduce the value of TV at $x^k$, we choose $v^k$ as

$$v^k = s^k / |s^k|_\infty$$

where $s^k \in \partial TV(x^k)$ is the sub-gradient of TV at point $x^k$, and $|s^k|_\infty$ is the maximum absolute value of the components of $s^k$. Therefore, the sequence $\{v^k\}$ is bounded. In fact, $v^k$ is the normalization of $s^k$ in the $l^\infty$ space, rather than in the $l^1$ space used in [25, 27].

In addition to the TV based models, $l^1$-norm minimization method is another
widely used technique in image sciences [20, 23, 32, 33]. Here, the \( l^1 \)-norm is about the wavelet coefficients of \( x \), which is defined as

\[
||T_1(\psi_j)(x)||_1 = \sum_{j=1}^{N} |\alpha_j|,
\]

where \( \alpha_j \) are the coefficients of \( x \) under a given wavelet basis \( \{ \psi_j \} \), and the letter \( T \) denotes the wavelet decomposition operator. Although we can use the same method for TV function to reduce the wavelet \( l^1 \)-norm, we introduce two more effective methods, soft and hard thresholding schemes, to reduce the wavelet \( l^1 \)-norm of \( x^k \). Let

\[
\gamma^k_j = \text{Hard}(\alpha^k_j) = \begin{cases} 0 & |\alpha^k_j| \geq \beta_k \\ -\frac{\alpha^k_j}{|\alpha^k_j|} \text{sign}(\alpha^k_j) & |\alpha^k_j| < \beta_k \end{cases},
\]

or

\[
\gamma^k_j = \text{Soft}(\alpha^k_j) = \begin{cases} -\text{sign}(\alpha^k_j) & |\alpha^k_j| \geq \beta_k \\ -\frac{\alpha^k_j}{|\alpha^k_j|} \text{sign}(\alpha^k_j) & |\alpha^k_j| < \beta_k \end{cases},
\]

where \( \alpha^k_j \) are the wavelet coefficients of \( x^k \). The perturbation direction for each iteration is defined as \( \nu^k = T_{(\psi_j)}^{-1}(\gamma^k_j) \), and \( y^k = x^k + \beta_k \nu^k \). In fact, we need not to compute \( \nu^k \) explicitly. By the linearity of wavelet transform, we can obtain the wavelet coefficients of \( y^k \) directly by

\[
\alpha^k_j = \text{Hard}^C(\alpha^k_j) = \begin{cases} \alpha^k_j & |\alpha^k_j| \geq \beta_k \\ 0 & |\alpha^k_j| < \beta_k \end{cases},
\]

or

\[
\alpha^k_j = \text{Soft}^C(\alpha^k_j) = \begin{cases} \alpha^k_j - \text{sign}(\alpha^k_j) \cdot \beta_k & |\alpha^k_j| \geq \beta_k \\ 0 & |\alpha^k_j| < \beta_k \end{cases},
\]

and \( y^k = T_{(\psi_j)}^{-1}(\alpha^k_j) \). In the numerical experiments, we use Daubechies 6.8 biorthogonal wavelets with symmetric extensions at the boundaries. For referred convenience, we use hard-superiorization and soft-superiorization to distinguish them for \( l^1 \)-superiorization of EM algorithm in this paper. Furthermore, in order to avoid \( y^k_j \leq 0 \), \( y^k_j \) is set as \( \frac{1}{2} x^k_j \) if \( y^k_j \leq 0 \) in practice.

4. **Numerical results.** In this section we investigate the performance of the proposed algorithms by several numerical experiments of SPECT image reconstruction. To this end, and the projection data were generated based on the following model.

As shown in figure 1, the activity phantom consists of an ellipsoidal background (body region) with axes of length 22.5cm and 30cm, which contains two smaller ellipsoidal regions (lungs) with axes of length 10cm and 8.8cm, and a ring (myocardium) of inner and outer diameters 6cm and 8cm, respectively. The activities in myocardium, background, and lungs are specified to be in the ratio 3:2:1.

To simulate the attenuation coefficient in chest, we utilized the phantom used in [10], which imitates a section of human thorax. Besides the body background and lungs, the attenuation map consists of two circular regions (bones) of diameter 2.5cm (see figure 1). The attenuation coefficients were 0.03cm\(^{-1}\) within ‘lung’ regions, 0.17cm\(^{-1}\) within ‘bone’ regions, 0.15cm\(^{-1}\) elsewhere within the body ellipse, and 0.00cm\(^{-1}\) outside the body.

In the experiments, the activity and attenuation maps were evenly sampled in \([-15, 15] \times [-15, 15]\) on a grid of 128 \times 128. A perfect parallel hole collimator
was assumed, and noise-free projection data were created via attenuated Radon transform formula (1), which included tissue attenuation, but neither scattering nor blurring. In order to simulate the quantum noise in the simulated data, the following procedure was implemented [34].

- The projection data are scaled (multiplied by a constant factor) so that the number of counts is a predefined integer.
- Each value in the data set is then replaced by a random realization of a Poisson variant with a mean equal to that value.

Two data sets, sixty and thirty projections, were generated over $180^\circ$ evenly with view angles $\phi_l = \frac{l-1}{N_0} \pi (l = 1, \cdots, N_0)$, and $N_0 = 60$ or 30). The counts were recorded in 128 bins per projection, and the total counts were approximately 500K and 100K for two projection data sets, respectively.

We illustrated four numerical experiments for the proposed algorithms. The first and second experiments are about the projection data set 1 (60 projections, 500K) and projection data set 2 (30 projections, 100K) to validate the performance of the algorithms for different counts level. The third experiment uses the randomly initial image to test the robustness of the proposed algorithms. In the fourth experiment, we relax the condition of the TV-superiorized EM algorithms 1 and 2 further.

In order to evaluate the qualities of the reconstructed images, we computed the mean images $x^*$ from 100 noise trials as the standard images of the two data sets, respectively.

\[
x^* = \frac{1}{100} \sum_{m=1}^{100} \hat{x}^m,
\]

where $\hat{x}^m (m = 1, 2, \cdots, 100)$ is the reconstructed image of the $m$th trial by the classic EM algorithm, and the number of iterations for each trial is 30.

As shown in figure 2, the mean images for the two data sets are clear visually, while those reconstructed by the classic EM algorithm are dominated by noise, especially the image for data set 2 due to the low count level. Thus, we use the mean square error (MSE) between the estimation $x$ and the mean image $x^*$ from...
Figure 2. Images reconstructed from the simulated data sets 1 and 2. EM \( k (k = 1, 2) \) are the reconstructed images from data set \( k \ (k = 1, 2) \) by the classic EM algorithm, respectively. Avg \( k \ (k = 1, 2) \) are the average images of data sets \( k \ (k = 1, 2) \).

the same data set to measure the image qualities, where the MSE is computed by

\[
\text{MSE}(x) = \frac{1}{N} \sum_{j=1}^{N} (x_j - x_j^*)^2.
\]

(36)

The outputs of the different algorithms were taken as the best estimations in terms of MSE. In order to compare the qualities of images reconstructed from different data sets, we introduced the relative MSE (RMSE) defined as

\[
\text{RMSE}(x) = \sqrt{\frac{\sum_{j=1}^{N} (x_j - x_j^*)^2}{\sum_{j=1}^{N} (x_j^*)^2}}.
\]

(37)

In the numerical experiments, an uniformly initial image \( x^0 \) of value \( c = \frac{1}{N} \sum_{i=1}^{M} b_i \) was used for all experiments, unless there was a further explanation. The parameters \( \beta_0 \) were chosen as \( c/2 \) and \( c/10 \) for the TV- and \( l^1 \)-superiorized EM algorithms, respectively. The thresholding \( Q_1 \) and the parameter \( \gamma \) were chosen as 0.01 and 1/2, respectively. In the following, we use TV-, hard- and soft-alg \( n(n = 1, 2) \) as the abbreviations of the TV-, Hard- and Soft-superiorized EM algorithm \( n(n = 1, 2) \) for simplicity.

**Experiment 1: sixty projections and 500K counts**
As expected, the images by the superiorized EM algorithms are superior to the one by the classic EM algorithm in terms of TV value, $l^1$-norm and RMSE (see figure 3 and table 1), though the images by the superiorized algorithm 1 are visually indistinguishable from the one by the classic EM algorithm. Furthermore, the superiorized EM algorithm 2 are superior to the superiorized EM algorithm 1, because the regularization parameter $\beta_k$ goes to zero very fast for the superiorized EM algorithm 1.

As figure 4 shown, the evolutions of $I_A^h(x^k)$ and MSE($x^k$) of the superiorized EM
algorithms 1 validate the conclusions of theorem 2.2. Furthermore, the evolutions of $I^A_k(x^k)$ and $\text{MSE}(x^k)$ show the convergence of the superiorized EM algorithms 2, although we can not prove it theoretically. Because the inequality of condition 3 is very strict, the parameters $\beta_k$ went to zero very fast for the superiorized algorithm 1. This results in the reconstructed images and the evolutions of $I^A_k(x^k)$ and $\text{MSE}(x^k)$ of the superiorized EM algorithms 1 and the classic EM algorithm are indistinguishable from each other, i.e. the regularization of the superiorized EM algorithm 1 is not remarkable.

Although the number of iterations of the superiorized EM algorithm 2 is larger...
than the classic EM and the superiorized EM algorithm 1 (Table 1), the superiorized algorithm 2 is superior to the superiorized algorithm 1 in terms of RMSE: \( \text{RMSE}(x^{tv,13}) = 0.1668 \), \( \text{RMSE}(x^{h,13}) = 0.1741 \) and \( \text{RMSE}(x^{s,13}) = 0.1797 \), where \( x^{tv,13}, x^{h,13}, x^{s,13} \) denote the results of tv-, hard- and soft-superiorized algorithms 2 after 13 iterations, respectively.

**Experiment 2: thirty projections and 100K counts**

From the observations of Figures 5 and 6, and Table 2, we can draw the same

![Figure 5](image.png)

**Figure 5.** Reconstructed images of experiment 2 by different superiorized EM algorithms.

|          | EM 2 | TV-alg 1 | Hard-alg 1 | Soft-alg 1 |
|----------|------|----------|------------|------------|
| TV\((\times 10^3)\) | 13.739 | 13.105 | 13.396 | 13.061 |
| \(l^1(\times 10^3)\) | 6.729 | 6.452 | 6.451 | 6.277 |
| RMSE     | 0.2822 | 0.2778 | 0.2802 | 0.2795 |
| Iteration| 8    | 8        | 8         | 8          |

|          | Avg 2 | TV-alg 2 | Hard-alg 2 | Soft-alg 2 |
|----------|-------|----------|------------|------------|
| TV\((\times 10^3)\) | 7.167 | 9.769   | 5.733    | 5.576     |
| \(l^1(\times 10^3)\) | 4.037 | 5.069   | 1.434   | 1.283     |
| RMSE     | 0     | 0.2490  | 0.2091   | 0.2064    |
| Iteration| 9     | 12      | 10       |            |

conclusions as the experiment 1. Comparing Figures 3 and 5, we can see that the
Figure 6. Evolutions of experiment 2. The evolutions of $P(x^k)$ (top row), MSE($x^k$) (middle row) and ln($1+\beta_k$) (bottom row) of different algorithms.

visual quality of the images reconstructed from data set 2 are inferior to those reconstructed from data set 1 because of the low count level. And the performance of the superiorized EM algorithms 2 are much more remarkable than the superiorized EM algorithms 1 for data set 2.

Comparing the results of the two experiments above, we can observe that the TV-superiorized EM algorithm 2 is better in terms of RMSE, while the $l^1$-superiorized EM algorithms 2 in terms of TV and $l^1$-norm values. In addition, the thresholding operations cause the Gibbs oscillations in the reconstructed images by $l^1$-superiorized EM algorithms 2.
Because the detectors rotated from top to bottom through left side, the gamma rays emitted from the right part pixels are more likely to be absorbed. Therefore, the right part of the reconstructed images are blurred strongly (see figures 2, 3 and 5).

**Experiment 3: initial image $x^0$ with random values on interval $[1, 2]$ for data set 1**

In this experiment, an initial image $x^0$ with random values on interval $[1, 2]$ and the projection data set 1 are used. The reconstructed images by different algorithms are displayed in figure 7, and the corresponding TV values, $l^1$-norms and RMSEs are tabulated in table 3. Because the evolutions of $I^k_A(x^k)$ and $\text{MSE}(x^k)$ are very similar to these of experiment 1, we only plot the evolution of $\ln(1 + \beta_k)$ in figure 8.

This experiment shows that the superiorized EM algorithms 2 are stable and robust for initial image. By comparing figures 3 and 7, and tables 1 and 3, we have that the effect of initial image is very strong to the classic EM algorithm and the superiorized EM algorithm 1, but very weak to the superiorized EM algorithm 2.

A surprising observation is that the randomly initial image is superior to the uniformly initial image for the superiorized EM algorithms 2 in term of RMSE by comparing table 1 and table 3. This changes the long-standing opinion about the selection of the initial image for EM-like algorithm, and present a new method to improve the qualities of the reconstructed image.

|                | EM-r | TV-alg 1 | Hard-alg 1 | Soft-alg 1 |
|----------------|------|----------|------------|------------|
| TV ($\times 10^4$) | 40.591 | 29.804 | 40.364 | 34.214 |
| $l^1$ ($\times 10^3$) | 19.098 | 14.632 | 18.907 | 16.138 |
| RMSE            | 0.2538 | 0.2112 | 0.2529 | 0.2266 |
| iteration       | 13    | 13       | 13         | 13         |

|                | TV-alg 2 | Hard-alg 2 | Soft-alg 2 |
|----------------|----------|------------|------------|
| TV ($\times 10^4$) | 17.133  | 16.203     | 16.372   |
| $l^1$ ($\times 10^3$) | 9.403    | 3.417   | 3.207   |
| RMSE            | 0.1469 | 0.1683 | 0.1873 |
| iteration       | 19     | 30      | 23       |

Comparing the evolutions of $\ln(1 + \beta_k)$ in figures 4, 6 and 8, we can observe the following facts. Firstly, the parameter $\beta_k$ for the superiorized EM algorithms 1 go to zero fast because of the strict condition 3 of theorem 2.2. Secondly, the reconstructed images of uniformly initial image is more smooth than these of randomly initial image by the TV-superiorized algorithm at low iterations, which results in the parameter $\beta_k$ decrease at fast and low rates for experiments 1 and 3, respectively, because of the decreasing condition of the TV function in the conditions (\ast) and (\divide). Thirdly, the thresholding operations (hard and soft) always reduce the $l^1$-norm, which implies that the conditions (\ast) and (\divide) for the $l^1$-superiorized algorithm 1 and 2 become one condition about the decreasing of K-L distance.

The last observation enlightens us to modify the TV-superiorized EM algorithm further, which discards the deceasing condition of TV function in the conditions (\ast) and (\divide).
Figure 7. Reconstructed Images of experiment 3 by EM algorithm and different superiorized EM algorithms with randomly initial values.

Experiment 4: modification of the TV-superiorized algorithms 1 and 2

Figure 9 displays the reconstructed images by the modified versions of TV-superiorized EM algorithms 1 and 2 in absence of the decreasing condition of TV function. The TV values, $l^1$-norms, RMSEs and iterations are tabulated in table 4. And figure 10 plots the evolutions of $\ln(1 + \beta_k)$.

As our expectation, the evolution of parameter $\beta_k$ of the modified version of TV-superiorized algorithm 2 is similar to the $l^1$-superiorized EM algorithm 2, decreasing at much slower rate. However, this modification has very little effect on the TV-superiorized algorithm 1.

It is amazing that the modified algorithms also reduce the TV function (see table 3), even the reconstructed images are better than those reconstructed by the TV-superiorized EM algorithm 2, although we do not validate the decreasing condition of it. The reasons include two aspects. In superiorized algorithms 1 and 2,
there are two conditions to control the decreasing of $\beta_k$, which causes the size of $\beta_k$ is very small at large iterations. Therefore, the strength of regularization is very weak, and the reconstructed image is not good enough. For the modified superiorization algorithms, there is only one condition to control the decreasing of $\beta_k$, and $\beta_k$ decreases at a lower rate. Therefore, the modified superiorization algorithms can maintain stronger regularization at large iterations, and the reconstructed image is much better. The further study about this algorithm is future work.

**Table 4.** TV values, $l^1$-norm, RMSEs and iterations of the images of experiment 4 (images in figure 9).

|          | (a)   | (b)   | (c)   | (d)   |
|----------|-------|-------|-------|-------|
| TV($\times 10^3$) | 25.318 | 12.931 | 12.955 | 7.080 |
| $l^1$($\times 10^3$) | 12.404 | 6.408 | 7.013 | 3.399 |
| RMSE     | 0.1873 | 0.2772 | 0.1231 | 0.1999 |
| iteration| 13     | 8     | 29    | 18    |

**Remark 2.** The experiments show that the superiorized EM algorithm 2 is convergent, though we cannot prove it theoretically so far.

**Remark 3.** The parameter $\beta_k$ represents the strength of regularization in a sense, so we can obtain the regularization reconstruction by terminating the algorithms as long as $\beta_k$ is smaller than a predefined threshold. As explanation above, the size of $\beta_k$ represents the strength of regularization. Therefore, the threshold should be related to the noise level. Intuitively, in order to maintain the regularization strength, the higher the noise level is, the larger the threshold is. The further discussion about the selection of it will be studied in future work.

5. **Conclusions and discussions.** In this paper, the convergence of the EM algorithm in the presence of perturbations is discussed, and the superiorized EM algorithm based on the convergent conditions and its modified version are proposed. The numerical experiments validate the correction of theorem 2.2. The superiorized EM algorithms could efficiently reduce the corresponding objective functions which we are superiorizing. Furthermore, The proposed algorithms are more stable and
Figure 9. Images of experiment 4. The images in left and right columns are reconstructed from data sets 1 and 2, while the images in first and second rows are reconstructed by modified versions of TV-superiorized EM algorithms 1 and 2, respectively. From left to right and from top to bottom, the images are labeled as (a), (b), (c) and (d) for reference.

Figure 10. Evolutions of $\ln(1+\beta_k)$ of experiment 4. Left for data set 1 and right for data set 2, respectively.

robust than the classic EM algorithm for low counts projection data and randomly initial image.

Although the numerical experiments show the convergence of algorithm 2, we
cannot prove the convergence of it theoretically. A more challenging work is about
the amazing observation of the experiment 4, which enlightens us to modify the
superiorized EM algorithm further. In addition, we could not prove \( \phi(x^*) \leq \phi(\hat{x}) \)
theoretically, where \( \hat{x} \) and \( x^* \) are the solutions by the classic iteration algorithm
and the \( \phi \)-superiorization version [25].

**Acknowledgments.** The authors are grateful for the helpful discussions with Pro-

fessor Ming Jiang (School of Mathematical Sciences, Peking University).

**REFERENCES**

[1] M. N. Wernick and J. N. Aarsvold, *Emission Tomography: the Fundamentals of PET and
SPECT*, Elsevier Academic Press, California, 2004.

[2] G. T. Herman, *Image Reconstruction from Projection: the Fundamentals of Computerized
Tomography*, Academic Press, New York 1980.

[3] F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction*, Society for
Industrial and Applied Mathematics, Philadelphia, 2001.

[4] O. Tretiak and C. Metz, *The exponential Radon transform*, *SIAM Journal on Applied Math-
ematics*, 39(1980), 341–354.

[5] F. Natterer, *Inversion of the attenuated Radon transform*, *Inverse Problems*, 17 (2001), 113–
119.

[6] R. G. Novikov, *An inversion formula for the attenuated X-ray transformation*, *Arkiv för
Matematik*, 40 (2002), 145–167.

[7] F. Noo and J.-M. Wagner, *Image reconstruction in 2D SPECT with 180\(^\circ\) acquisition*, *Inverse
Problems*, 17 (2001), 1357–1371.

[8] F. Noo, M. Defrise, J. D. Pack and R. Clackdoyle, *Image reconstruction from truncated data
in single-photon emission computed tomography with uniform attenuation*, *Inverse Problems,
23* (2007), 645–667.

[9] L. A. Shepp and Y. Vardi, *Maximum likelihood restoration for emission tomography*, *IEEE
Transactions on Medical Imaging*, 1 (1982), 113–122.

[10] H. M. Hudson and R. S. Larkin, *Accelerated image reconstruction using ordered subsets of
projection data*, *IEEE Transactions on Medical Imaging*, 13 (1994), 601–609.

[11] I. Haiao and H. Huang, *An accelerated ordered subsets reconstruction algorithm using an
accelerating power factor for emission tomography*, *Physics in Medicine and Biology*, 55
(2010), 599–614.

[12] Y. Vardi, L. A. Shepp and L. Kaufman, *A statistical model for positron emission tomography,
Journal of the American Statistical Association*, 80 (1985), 8–20.

[13] R. Gordon, R. Bender and G. T. Herman, *Algebraic reconstruction techniques (ART) for
dimensional electron microscopy and X-ray photography*, *Journal of Theoretical Biol-
yology*, 29 (1970), 471–482.

[14] Y. Censor, T. Elfving and G. T. Herman, *Averaging strings of sequential iterations for convex
feasibility problems*, in *Inherently Parallel Algorithms in Feasibility and Optimization and
their Applications*, (2001), 101–113.

[15] E. Y. Sidky, C.-M. Kao and X. Pan, *Accurate image reconstruction from few-views and
limited-angle data in divergent-beam CT*, *X-Ray Science and Technology*, 14 (2006), 119–
139.

[16] S. J. LaRoque, E. Y. Sidky and X. Pan, *Accurate image reconstruction from few-view and
limited-angle data in diffraction tomography*, *Journal of the Optical Society of America*, 25
(2008), 1772–1782.

[17] X. Pan, E. Y. Sidky and M. Vannier, *Why do commercial CT scanners still employ traditional,
filtered backprojection for image reconstruction?*, *Inverse Problems*, 25 (2009), 123009.

[18] M. Defrise, C. Vanhove and X. Liu, *An algorithm for total variation regularization in high-
dimensional linear problems*, *Inverse Problems*, 27 (2011), 065002.

[19] E. Y. Sidky, J. H. Jogensen 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), 3065–3091.

[20] B. Dong, J. Li and Z. Shen, *X-ray CT image reconstruction via wavelet frame based regular-
ization and Radon domain inpainting*, *Journal of Scientific Computing*, 54 (2013), 333–349.
[21] D. Butnariu, R. Davidi, G. T. Herman and I. G. Kazantsev, Stable convergence behavior under summable perturbations of a class of projection methods for convex feasibility and optimization problem, IEEE Journal of Selected Topics Signal Processing, 1 (2007), 540–547.
[22] G. T. Herman and R. Davidi, Image reconstruction from a small number of projections, Inverse Problems, 24 (2008), 045011.
[23] E. Garduno, G. T. Herman and R. Davidi, Reconstruction from a few projections by $l^1$ minimization of the Haar transform, Inverse Problems, 24 (2011), 055006.
[24] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problem with applications to imaging, Journal of Mathematical Imaging and Vision, 40 (2011), 120–145.
[25] Y. Censor, R. Davidi and G. T. Herman, Perturbation resilience and superiorization of iteration algorithm, Inverse Problems, 25 (2010), 065008.
[26] R. Davidi, G. T. Herman and Y. Censor, Perturbation-resilient block-iterative projection methods with application to image reconstruction from projection, International Transactions in Operational Research, 16 (2009), 505–524.
[27] T. Nikazad, R. Davidi and G. T. Herman, Accelerated perturbation-resilient block-iterative projection methods with application to image reconstruction, Inverse Problems, 28 (2012), 035005.
[28] W. Jin, Y. Censor and M. Jiang, A heuristic superiorization-like approach to bioluminescence tomography, World Congress on Medical Physics and Biomedical Engineering in Proceedings of the International Federation for Medical and Biological Engineering (IFMBE), 39 (2013), 1026-1029.
[29] L. I. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D., 60 (1992), 259–268.
[30] E. Candes, J. Romberg and T. Tao, The dantzig selector: statistical estimation when $p$ is much larger than $n$, the Annals of Statistics, 35 (2007), 2313–2351.
[31] D. L. Snyder, T. J. Schulz and J. A. O. Sullivan, Deblurring subject to nonnegativity constraints, IEEE Transactions on Signal Processing, 40 (1992), 1143–1150.
[32] E. Candes, J. Romberg and T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Transactions on Information Theory, 52 (2006), 489–509.
[33] H. Yu and G. Wang, Compressed sensing based interior tomography, Physics in Medicine and Biology, 54 (2009), 2791–2805.
[34] J. A. Fessler, Matlab code for emission tomograph, Available from: http://www.eecs.umich.edu/~fessler/.

Received December 2012; revised May 2013.

E-mail address: kxylss@163.com
E-mail address: tzhou@math.pku.edu.cn