Journal of Computational Mathematics
Vol.34, No.6, 2016, 626–647.

http://www.global-sci.org/jcm
doi:10.4208/jcm.1607-m2016-0537

REDUCING STAIRCASING ARTIFACTS IN SPECT RECONSTRUCTION BY AN INFIMAL CONVOLUTION REGULARIZATION*

Zhifeng Wu
School of Mathematics, Guangdong Provincial Key Lab of Computational Science
Sun Yat-sen University, Guangzhou 510275, P. R. China
E-mail: whutxxz@163.com

Si Li
School of Data and Computer Science, Guangdong Provincial Key Lab of Computational Science
Sun Yat-sen University, Guangzhou 510275, P. R. China
E-mail: reesiloveu@163.com

Xueying Zeng
School of Mathematical Sciences, Ocean University of China, Qingdao 266100, P. R. China
E-mail: zxying@ouc.edu.cn

Yuesheng Xu
School of Data and Computer Science, Guangdong Provincial Key Lab of Computational Science
Sun Yat-sen University, Guangzhou 510275, P. R. China
Department of Mathematics, Syracuse University, Syracuse, NY 13244, USA
E-mail: yxu06@syr.edu

Andrzej Krol
Department of Radiology and Department of Pharmacology, SUNY Upstate Medical University
Syracuse, NY 13210, USA

Abstract

The purpose of this paper is to investigate the ability of the infimal convolution regularization in curing the staircasing artifacts of the TV model in the SPECT reconstruction. We formulate the problem of SPECT reconstruction with the infimal convolution regularization as a convex three-block optimization problem and characterize its solution by a system of fixed-point equations in terms of the proximity operator of the functions involved in its objective function. We then develop a novel fixed-point proximity algorithm based on the fixed-point equations. Moreover, we introduce a preconditioning matrix motivated by the classical MLEM (maximum-likelihood expectation maximization) algorithm. We prove convergence of the proposed algorithm. The numerical results are included to show that the infimal convolution regularization is capable of effectively reducing the staircasing artifacts, while maintaining comparable image quality in terms of the signal-to-noise ratio and coefficient recovery contrast.

Mathematics subject classification: 68U10, 49Q20.

Key words: SPECT, Infimal Convolution Regularization, Staircasing Artifacts, Fixed-point Proximity Algorithm.

* Received February 11, 2016 / Revised version received June 30, 2016 / Accepted July 1, 2016 / Published online December 16, 2016 /

1. Introduction

The goal of this article is to investigate the ability of the infimal convolution of functionals with the first- and second-order derivatives, denoted by ICTV, to cure staircasing artifacts in SPECT reconstruction. We develop a novel fixed-point proximity algorithm to solve the resulting three-block optimization problem, motivated by the fixed-point framework studied in [16, 17, 21, 22].

Single photon emission computerized tomography (SPECT) is an important medical imaging tool for studying functional characteristics of human bodies. In contrast, computed tomography (CT) or magnetic resonance imaging (MRI) provides only structure information of human bodies. For this reason, SPECT has become one of the most important techniques for detection and evaluation of coronary artery diseases. Other clinical applications of SPECT include detecting, staging and monitoring response to cancer therapy, pulmonary ventilation/perfusion scans, renal scans, and bone scans [32]. Clinical SPECT data are often severely corrupted by Poisson noise due to low tracer dosage and short acquisition time. A widely used algorithm today for clinical SPECT image reconstruction is the ordered-subset expectation maximization algorithm (OSEM) [13], which has greatly improved the convergence speed compared to the original maximum-likelihood expectation maximization algorithm (MLEM) [15, 29]. However, one disadvantage of OSEM is that the noise in the reconstructed image increases with iterations, and the resulting noise level often makes the reconstructed image clinically unacceptable. Therefore, it is stopped prematurely and post-filtering should be applied [20].

As an alternative to OSEM, maximum *a posteriori* (MAP) in a Bayesian framework was proposed [11], and has gained wide interest since then. MAP incorporates a regularization term into the objective function, which assembles the *a priori* knowledge of the image. The total variation (TV) regularization is capable of suppressing noise effectively, and at the same time capturing sharp edges [26]. Hence the TV semi-norm has been widely used as a regularizer in SPECT reconstruction [3, 14, 23, 25, 27]. The TV method works quite well when the true object contains only piecewise constant regions. However, when the true object is smooth, the use of the TV regularization often results in patchy, cartoon-like images, known as staircasing artifacts [31]. In the past two decades, considerable research interest was devoted to overcoming this side effect of the TV regularization. Several variants of the TV regularization were proposed. Chan et al. introduced in [8] a fourth-order non-linear filter. A generalized total variation model, known as TGV, was proposed by Bredies et al. in [4]. A recent work [19] studied effective noise-suppressed and artifact-reduced reconstruction of SPECT data using a higher-order TV regularization and the resulting optimization problem was solved by using a preconditioned alternating project algorithm.

The first issue that we study in this paper is to investigate the suitability of using the ICTV regularization in SPECT reconstruction. The ICTV regularization, suggested by Chambolle and Lions in [7], combining the total variation of the image with the total variation of its gradient, showed a possibility of reducing staircasing artifacts in mathematical imaging. Whether it is suitable for SPECT reconstruction remains unclear. Studying this issue is the first objective of this paper. The reason that we believe that ICTV is able to reduce staircasing artifacts in SPECT reconstruction may be explained in two ways. The null space of the ICTV functional is much larger than that of the TV functional. We expect that the ICTV regularization would be more likely to produce smoother reconstruction than the TV regularization, and the staircasing artifacts would be alleviated. We can also view the ICTV functional as

decomposing a bounded variation function into two components, one having the appearance of TV-regularized reconstructions (with sharp edges and piecewise constant regions) and the other resembling higher-order derivative reconstructed images (with smoother estimated radiotracer distribution). As a result, the ICTV regularized reconstruction, as a linear combination of the above two components, would be smoother than the TV reconstructed image and the staircasing artifacts would likely to be reduced.

The second issue is to develop a mathematically sound and computationally efficient algorithm to solve the resulting non-smooth three-block optimization problem. A direct extension of the alternating direction method of multipliers (ADMM) may give a shot to solve the problem. However, it has been pointed out in [9, 18] that this direct extension of ADMM is not necessarily convergent. In contrast, we shall develop an iterative algorithm based on the fixed-point equation that characterizes the solution of the optimization problem and show that the proposed algorithm is guaranteed to converge. It is shown in [14] that, the preconditioning matrix adopted from the MLEM algorithm can significantly improve the convergence speed. Following this idea, we shall propose an EM-type preconditioning matrix for the proposed algorithm. We shall demonstrate by numerical experiments that the use of the ICTV regularization can greatly reduce staircasing artifacts appearing in the conventional TV regularization.

We organize this paper in seven sections. In Section 2, we describe the ICTV model of SPECT reconstruction. Section 3 presents the fixed-point characterization of the solution of the optimization problem which models SPECT reconstruction. Based on the characterization we propose in Section 4 a fixed-point proximity algorithm to solve the ICTV model. Convergence of the proposed algorithm is analyzed in Section 5. Numerical results are presented in Section 6 to demonstrate the capability of the infimal convolution regularization in effectively reducing the staircasing artifacts, with maintaining comparable image quality in terms of the signal-to-noise ratio and coefficient recovery contrast. Finally, we draw conclusions in Section 7.

2. Mathematical Model of SPECT

In a SPECT system, a patient is injected with a radiopharmaceutical, which concentrates in the organ of interest. The gamma camera rotates around the patient, collecting photons emitted by the radioactive isotopes in the radiopharmaceutical. The goal of SPECT reconstruction is to obtain an estimate of the three-dimensional distribution of radiopharmaceutical from the detected array of photon counts. It is well established that the gamma photon emission rate and subsequent photon detection at each detector element follow the temporal Poisson distribution. Therefore, the emission data and the unobservable expected radioactivity distribution can be described by a Poisson model.

We now describe the Poisson model for the SPECT system. Let $\mathbb{N}_p := \{1, 2, \ldots, p\}$ and denote by $A \in \mathbb{R}^{p\times d}$ the SPECT system matrix with its (ij)th entry equal to the probability of detection of the photon emitted from voxel j of an image by the ith detector element, with $i \in \mathbb{N}_p$ and $j \in \mathbb{N}_d$. The system matrix A summarizes detector geometry, attenuation and other physical degradations. Since photons emitted from each voxel of an image are recorded by detector elements in the SPECT imaging system, it is reasonable to assume that each column of A is a nonzero vector in $\mathbb{R}^p_+ := \{x : x \in \mathbb{R}^p \text{ and } x \geq 0\}$. Furthermore, we let f represent the expected radioactivity distribution within a phantom and let $g \in \mathbb{R}^p$ denote a vector with the ith component g_i being the number of single photons originated from the radiotracer and recorded by the ith detector element, for $i \in \mathbb{N}_p$. Under these assumptions, the observed

emission data g and the unknown radioactivity distribution $f \in \mathbb{R}^d$ can be modelled as

$$g = Af + \gamma,$$

with the positive number γ representing the mean number of background counts recorded by each detector element. Here, for simplicity, we have assumed that the mean number of background counts recorded by each detector element takes the same value. The summation of a vector and a scalar will be defined later in this section. The above linear equation is highly ill-posed and is improper to be solved directly. It requires a proper regularization.

We now formulate the reconstruction problem from a probability viewpoint. Since the photon detections follow the Poisson distribution, we may consider the number of detected counts g_i as a realization of a Poisson distributed random variable X_i, $i \in \mathbb{N}_p$. The expected value of X_i is given by $(Af)_i + \gamma$, $i \in \mathbb{N}_p$. Applying the maximum *a posteriori* criterion (MAP), a reasonable estimate of the radioactivity distribution f is determined as the maximizer of the conditional probability $P(f|g)$, the probability that f occurs when g is observed. This probability may in turn be computed using the Bayes law:

$$P(f|g) = \frac{P(g|f)P(f)}{P(g)}, \tag{2.1}$$

where $P(f)$ is the *a priori* distribution of f. Since $P(g)$ does not depend on f, MAP is equivalent to maximizing $P(g|f)P(f)$. Recalling the Poissonian nature of the emission data, we have that

$$P(g|f) = \prod_{i=1}^{p} \frac{((Af)_i + \gamma)^{g_i}}{g_i!} e^{-((Af)_i + \gamma)}.$$

As for $P(f)$, the Gibbs *priors* are most commonly advocated:

$$P(f) = e^{-\lambda U(f)}, \tag{2.2}$$

where $\lambda > 0$ is a regularization weight called the hyper-parameter and U is a convex energy functional. Since logarithmic operation does not change the monotonicity of a function, we in turn minimize $-\ln(P(g|f)P(f))$. As a result, we arrive at the following minimization problem

$$\min \left\{ \sum_{i=1}^{p} ((Af)_i + \gamma - g_i \ln((Af)_i + \gamma)) + \lambda U(f) : f \in \mathbb{R}^d, f \geq 0 \right\}. \tag{2.3}$$

We next write (2.3) in a compact form. To this end, we introduce some notation. We define the component-wise multiplication and division, respectively, by

$$x \odot y := (x_i y_i : i \in \mathbb{N}_d) \quad \text{and} \quad \frac{x}{y} := \left(\frac{x_i}{y_i} : i \in \mathbb{N}_d \text{ and } y_i \neq 0 \right),$$

for any vectors x, $y \in \mathbb{R}^d$. The logarithmic function of $x \in \mathbb{R}^d$ is given by

$$\ln x := (\ln x_i : i \in \mathbb{N}_d).$$

For a vector x and a scalar $\theta \in \mathbb{R}$, $x + \theta$ means,

$$x + \theta := (x_i + \theta : i \in \mathbb{N}_d).$$

The inner product of x, $y \in \mathbb{R}^d$ is defined as

$$\langle x, y \rangle := \sum_{i=1}^{d} x_i y_i.$$

We use '1' to represent both the real number 1 and the vector with all components equal to 1. As a result, we have that

$$\sum_{i=1}^{p}(Af)_i = \langle Af, 1\rangle \quad \text{and} \quad \sum_{i=1}^{p} g_i \ln\left((Af)_i + \gamma\right) = \langle \ln\left(Af+\gamma\right), g\rangle.$$

With these notations, if we choose the energy functional U in model (2.3) to be the TV semi-norm, we obtain a TV regularized optimization model for SPECT reconstruction,

$$\min\Big\{\langle Af, 1\rangle - \langle \ln(Af+\gamma), g\rangle + \lambda|\nabla f| : f \in \mathbb{R}^d_+\Big\}. \tag{2.4}$$

However, it is well-known that the TV regularizer may create staircasing artifacts [31]. Typically, such artifacts can be reduced by incorporating higher-order derivatives into the regularization function [19]. The infimal convolution of functionals with the first- and second-order derivatives as a regularizer, per discussed in the introduction, is an appealing approach in this direction. With this motivation, we decompose the image f into the sum of components f_1 and f_2, where f_1 has the appearance of TV-regularized reconstructions with sharp edges and piecewise constant regions and f_2 resembles second-order derivative reconstructed images with a smoother distribution. As a result, we propose the infimal convolution regularization model (ICTV) for SPECT reconstruction,

$$\begin{aligned}\min\Big\{&\langle A(f_1+f_2), 1\rangle - \langle \ln(A(f_1+f_2)+\gamma), g\rangle + \lambda_1|\nabla f_1| \\ &+ \lambda_2|\nabla\cdot(\nabla f_2)| : f_1+f_2 \in \mathbb{R}^d_+\Big\}.\end{aligned} \tag{2.5}$$

Adopting the notation used in [19, 21], we have that

$$|\nabla f_1| = \varphi_1(B_1 f_1) \quad \text{and} \quad |\nabla\cdot(\nabla f_2)| = \varphi_2(B_2 f_2),$$

where $B_1 \in \mathbb{R}^{m_1\times d}$ and $B_2 \in \mathbb{R}^{m_2\times d}$ are, respectively, the first- and second-order difference matrices and φ_1, φ_2 are both d-sum of the ℓ^2-norms for the isotropic TV, except that their ℓ^2-norms are the Euclidean norm with different dimensions. Concrete expressions of B_1, B_2, φ_1 and φ_2 are given in Section 4. The ICTV model (2.5) can then be rewritten as

$$\begin{aligned}\min\Big\{&\langle A(f_1+f_2), 1\rangle - \langle \ln(A(f_1+f_2)+\gamma), g\rangle + \lambda_1\varphi_1(B_1 f_1) \\ &+ \lambda_2\varphi_2(B_2 f_2) : f_1+f_2 \in \mathbb{R}^d_+\Big\}.\end{aligned} \tag{2.6}$$

Since each column of $A \in \mathbb{R}^p_+$ is a nonzero vector, we have that

$$\lim_{\substack{\|f_1\|_2+\|f_2\|_2\to+\infty \\ f_1+f_2\in\mathbb{R}^d_+}} \langle A(f_1+f_2), 1\rangle - \langle \ln(A(f_1+f_2)+\gamma), g\rangle = +\infty.$$

Hence, the solution set of the ICTV model (2.6) is nonempty.

Finally, for the sake of notational simplicity and algorithm development convenience, we rewrite the ICTV model (2.6) in a compact form. Let

$$u := \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \quad C := \begin{pmatrix} I, & I \end{pmatrix} \quad \text{and} \quad B := \begin{pmatrix} B_1 & \\ & B_2 \end{pmatrix},$$

where I is the $d\times d$ identity matrix and $B \in \mathbb{R}^{m\times 2d}$ with $m := m_1 + m_2$. We further let

$$\begin{aligned}&\Psi(u) := \langle A(f_1+f_2), 1\rangle - \langle \ln(A(f_1+f_2)+\gamma), g\rangle, \\ &\Phi(\xi,\eta) := \lambda_1\varphi_1(\xi) + \lambda_2\varphi_2(\eta).\end{aligned}$$

Recalling the indicator function of the nonnegative constraint set

$$\iota_{\mathbb{R}^d_+}(x) = \begin{cases} 0, & \text{if } x \in \mathbb{R}^d_+, \\ +\infty, & \text{otherwise,} \end{cases}$$

and setting $\Upsilon := \iota_{\mathbb{R}^d_+} \circ C$, we have that

$$\min \Big\{ \Psi(u) + \Phi(Bu) + \Upsilon(u) : u \in \mathbb{R}^{2d} \Big\}. \tag{2.7}$$

We shall develop a fixed-point algorithm to solve the non-smooth optimization problem (2.7).

3. Fixed-point Characterizations

This section is devoted to presenting a characterization of solutions of the ICTV model (2.7) via a system of fixed-point equations. This characterization plays a crucial role in developing a novel fixed-point proximity algorithm and its convergence analysis.

We start with introducing the concepts of the proximity operator, the subdifferential and the conjugate function. We denote by $\mathbb{S}^n$ (resp. $\mathbb{S}^n_+$) the set of $n \times n$ symmetric positive semi-definite (resp. definite) matrices. For a matrix $T \in \mathbb{S}^n_+$, we define the weighted inner product by

$$\langle x, y \rangle_T := \langle x, Ty \rangle, \quad x,\ y \in \mathbb{R}^n.$$

The induced weighted norm is accordingly defined by

$$\|x\|_T := \sqrt{\langle x, x \rangle_T}.$$

We further denote by $\Gamma_0(\mathbb{R}^n)$ the class of all proper lower semi-continuous convex functions $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$. For a function $\varphi \in \Gamma_0(\mathbb{R}^n)$, the proximity operator of φ with respect to a given matrix $T \in \mathbb{S}^n_+$, denoted by prox^T_φ, is a mapping from $\mathbb{R}^n$ to itself, defined for a given vector $x \in \mathbb{R}^n$ by

$$\text{prox}^T_\varphi(x) := \arg\min \left\{ \varphi(u) + \frac{1}{2}\|u - x\|^2_T : u \in \mathbb{R}^n \right\}. \tag{3.1}$$

In particular, we use prox_φ for prox^I_φ. The subdifferential of φ on $\mathbb{R}^n$ at a given vector $x \in \mathbb{R}^n$ is the set defined by

$$\partial\varphi(x) := \Big\{ y :\ y \in \mathbb{R}^n \text{ and } \varphi(z) \geq \varphi(x) + \langle y, z - x \rangle, \text{ for all } z \in \mathbb{R}^n \Big\}.$$

The subdifferential and the proximity operator of a convex function are intimately related. Specifically, for any x in the domain of φ and $y \in \mathbb{R}^n$, we have that

$$Ty \in \partial\varphi(x) \quad \text{if and only if} \quad x = \text{prox}^T_\varphi(x + y), \tag{3.2}$$

(see, for example, [2, Proposition 16.34] and [21, 24]). The function $\varphi^* \in \Gamma_0(\mathbb{R}^n)$ defined by

$$\varphi^*(u) := \sup\{\langle x, u \rangle - \varphi(x) : x \in \mathbb{R}^n\}, \quad u \in \mathbb{R}^n,$$

is called the conjugate of φ.

Now, we are in a position to characterize the solution of the ICTV model (2.7) via a system of fixed-point equations.

Theorem 3.1. *If $u \in \mathbb{R}^{2d}$ is a solution of the ICTV model (2.7), then for any $P \in \mathbb{S}_+^{2d}$ and $Q \in \mathbb{S}_+^m$, there exists $b \in \mathbb{R}^m$ such that the pair $(b, u) \in \mathbb{R}^m \times \mathbb{R}^{2d}$ is a solution of the following coupled equations*

$$b = \operatorname{prox}_{\Phi^*}^{Q}\left(b + Q^{-1}Bu\right), \tag{3.3}$$

$$u = \operatorname{prox}_{\Upsilon}^{P}\left(u - P^{-1}\left(\nabla\Psi\left(u\right) + B^{\top}b\right)\right). \tag{3.4}$$

Conversely, if there exist $P \in \mathbb{S}_+^{2d}$, $Q \in \mathbb{S}_+^m$, $b \in \mathbb{R}^m$ and $u \in \mathbb{R}^{2d}$ such that (3.3) and (3.4) hold, then u is a solution of the ICTV model (2.7).

Proof. Let $u \in \mathbb{R}^{2d}$ be a solution of the ICTV model (2.7). Applying Fermat's rule, we get the following inclusion relation

$$0 \in \nabla\Psi(u) + B^{\top}\partial\Phi(Bu) + \partial\Upsilon(u). \tag{3.5}$$

Hence, for arbitrary matrix $P \in \mathbb{S}_+^{2d}$, there exist $\eta \in \mathbb{R}^{2d}$ and $b \in \mathbb{R}^m$ such that $P\eta \in \partial\Upsilon(u)$, $b \in \partial\Phi(Bu)$, and the following equation holds

$$0 = \nabla\Psi(u) + P\eta + B^{\top}b. \tag{3.6}$$

Applying relation (3.2) to the inclusion $P\eta \in \partial\Upsilon(u)$ yields that

$$u = \operatorname{prox}_{\Upsilon}^{P}(\eta + u). \tag{3.7}$$

Solving η from (3.6) and substituting it into (3.7) give (3.4). For any $x \in \operatorname{dom}(\Phi)$ and $y \in \operatorname{dom}(\Phi^*)$, according to [12, Proposition 6.1.2], we obtain that

$$y \in \partial\Phi(x) \quad \text{if and only if} \quad x \in \partial\Phi^*(y). \tag{3.8}$$

Hence, from the inclusion $b \in \partial\Phi(Bu)$, we have that $Bu \in \partial\Phi^*(b)$. Therefore, for arbitrary matrix $Q \in \mathbb{S}_+^m$, we know that $Q\left(Q^{-1}Bx\right) \in \partial\Phi^*(b)$, which together with (3.2) yields (3.3).

Conversely, if $P \in \mathbb{S}_+^{2d}$ and $Q \in \mathbb{S}_+^m$ are given matrices such that (b, u) is a solution of (3.3)-(3.4), then all the arguments discussed above are reversible. This completes the proof. □

4. Fixed-point Algorithm

In this section, we develop an iterative scheme to solve the ICTV model (2.7) based on the system of fixed-point equations described in Theorem 3.1. Besides, we present the closed forms of proximity operators $\operatorname{prox}_{\Upsilon}^{P}$ and $\operatorname{prox}_{\Phi^*}^{Q}$ required for implementing the scheme. At the end of this section, we discuss the choice of the preconditioning matrix P.

We begin with rewriting the fixed-point characterization (3.3)-(3.4) in a compact form. For this purpose, we first introduce an operator which couples the two proximity operators involved in (3.3) and (3.4). In fact, for given matrices $P \in \mathbb{S}_+^{2d}$ and $Q \in \mathbb{S}_+^m$, we define the operator $\mathcal{T} : \mathbb{R}^m \times \mathbb{R}^{2d} \to \mathbb{R}^m \times \mathbb{R}^{2d}$, at a vector $v := (b, u) \in \mathbb{R}^m \times \mathbb{R}^{2d}$, as

$$\mathcal{T}(v) := \left(\operatorname{prox}_{\Phi^*}^{Q}(b), \operatorname{prox}_{\Upsilon}^{P}(u)\right). \tag{4.1}$$

We next show that operator $\mathcal{T}$ is indeed a proximity operator of the convex function $F(v) := \Phi^*(b) + \Upsilon(u)$, for some $v := (b, u) \in \mathbb{R}^m \times \mathbb{R}^{2d}$, with respect to the block diagonal matrix $R := \operatorname{diag}(Q, P) \in \mathbb{S}_+^{m+2d}$.

Lemma 4.1. *If $\mathcal{T}$ is the operator defined by (4.1), then $\mathcal{T}$ is the proximity operator of the convex function F with respect to the matrix R, that is, $\mathcal{T} = \text{prox}_F^R$.*

Proof. Let $v := (b, u) \in \mathbb{R}^m \times \mathbb{R}^{2d}$ be given. By the definition in (3.1) of the proximity operator, we have that

$$\text{prox}_F^R(v) = \arg\min \left\{ \frac{1}{2}\|\widetilde{v} - v\|_R^2 + F(\widetilde{v}) : \widetilde{v} = (\widetilde{b}, \widetilde{u}) \in \mathbb{R}^m \times \mathbb{R}^{2d} \right\}. \tag{4.2}$$

Recalling the separable structures of convex function F and matrix R, we can rewrite the minimizer in (4.2) in the form

$$\arg\min \left\{ \frac{1}{2}\|\widetilde{u} - u\|_P^2 + \Upsilon(\widetilde{u}) + \frac{1}{2}\|\widetilde{b} - b\|_Q^2 + \Phi^*(\widetilde{b}) : \widetilde{u} \in \mathbb{R}^{2d},\, \widetilde{b} \in \mathbb{R}^m \right\},$$

which by definition is the vector $\left(\text{prox}_{\Phi^*}^Q(b), \text{prox}_{\Upsilon}^P(u)\right)$. This completes the proof. □

Recall that a non-linear operator $\mathcal{P} : \mathbb{R}^n \to \mathbb{R}^n$ is called firmly nonexpansive with respect to a matrix $S \in \mathbb{S}_+^n$ if

$$\|\mathcal{P}x - \mathcal{P}y\|_S^2 \leq \langle \mathcal{P}x - \mathcal{P}y, x - y\rangle_S, \quad \text{for all } x, y \in \mathbb{R}^n. \tag{4.3}$$

It was proved in [10] that prox_{φ} is firmly non-expansive with respect to the identity matrix for any convex function φ on $\mathbb{R}^n$. This result can be naturally generalized to guarantee the firm non-expansivity of the operator $\mathcal{T}$ with respect to the matrix R.

Given an $m \times 2d$ matrix B, we define the $(m + 2d) \times (m + 2d)$ skew symmetric matrix S_B by

$$S_B := \begin{pmatrix} 0 & B \\ -B^\top & 0 \end{pmatrix}. \tag{4.4}$$

Using (4.4), we further introduce the $(m + 2d) \times (m + 2d)$ matrix

$$E := I + R^{-1}S_B. \tag{4.5}$$

Then, we introduce an operator consisting of the gradient of function Ψ and the null operator on $\mathbb{R}^m$. Indeed, we define the operator $\mathcal{G} : \mathbb{R}^m \times \mathbb{R}^{2d} \to \mathbb{R}^m \times \mathbb{R}^{2d}$ at $v = (b, u) \in \mathbb{R}^m \times \mathbb{R}^{2d}$ as

$$\mathcal{G}(v) := (0, \nabla\Psi(u)). \tag{4.6}$$

With the above notation, (3.3)-(3.4) can be rewritten in a compact form

$$v = \mathcal{T}\left(Ev - R^{-1}\mathcal{G}(v)\right). \tag{4.7}$$

We next propose an iterative scheme to solve (4.7). To this end, we first decompose the matrix E as $E = E - R^{-1}M + R^{-1}M$ for some $(m + 2d) \times (m + 2d)$ matrix M. Based on this decomposition, the fixed-point characterization (4.7) can be rewritten as

$$v = \mathcal{T}\left((E - R^{-1}M)v + R^{-1}Mv - R^{-1}\mathcal{G}(v)\right).$$

Accordingly, an implicit iterative scheme can be developed as follows

$$v^{k+1} = \mathcal{T}\left((E - R^{-1}M)v^{k+1} + R^{-1}Mv^k - R^{-1}\mathcal{G}(v^k)\right). \tag{4.8}$$

In order that algorithm (4.8) may be practically implemented, the matrix M has to be properly chosen so that b^{k+1} (resp., u^{k+1}) is not needed when updating itself, that is, scheme (4.8)

is blockwise explicit. Besides, $\mathrm{prox}^{Q}_{\Phi^*}(b)$ and $\mathrm{prox}^{P}_{\Upsilon}(u)$ are required to have closed forms. These closed forms are convenient for numerical evaluation in the sense that no further optimization problems need to be solved. Moreover, the choise of matrix M has to ensure convergence of the resulting algorithm.

In what follows, we specify the matrix M and develop a fixed-point proximity algorithm to solve the ICTV model (2.7) based on the iterative scheme (4.8). In particular, we set

$$M := \begin{pmatrix} Q & B \\ B^\top & P \end{pmatrix}, \tag{4.9}$$

where

$$Q = \begin{pmatrix} \mu_1^{-1} I_1 & \\ & \mu_2^{-1} I_2 \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} \beta_1^{-1} S_1^{-1} & \\ & \beta_2^{-1} S_2^{-1} \end{pmatrix}, \tag{4.10}$$

with positive parameters μ_1, μ_2, β_1 and β_2, and positive definite diagonal matrices S_1 and S_2. Thus, we obtain the following specific iterative scheme

$$\begin{cases} b^{k+1} = \mathrm{prox}^{Q}_{\Phi^*}\left(b^k + Q^{-1} B u^k\right) \\ u^{k+1} = \mathrm{prox}^{P}_{\Upsilon}\big(u^k - P^{-1}\left(\nabla\Psi(u^k) + B^\top\left(2b^{k+1} - b^k\right)\right)\big). \end{cases} \tag{4.11}$$

Next, we present the closed form of $\mathrm{prox}^{P}_{\Upsilon}$. For this purpose, we need two lemmas. The first lemma presents a characterization of the subdifferential of Υ.

Lemma 4.2. *Let $\xi \in \mathbb{R}^{2d}$. The subdifferential $\partial\Upsilon(\xi)$ is nonempty if and only if for $i \in \mathbb{N}_d$, $\xi_i + \xi_{i+d} \geq 0$. Moreover, for nonempty $\partial\Upsilon(\xi)$, $\eta \in \partial\Upsilon(\xi)$ if and only if it holds that $\eta_i = \eta_{i+d}$ for $i \in \mathbb{N}_d$, and*

$$\begin{cases} \eta_i = 0, & \text{if} \quad \xi_i + \xi_{i+d} > 0, \\ \eta_i \leq 0, & \text{if} \quad \xi_i + \xi_{i+d} = 0. \end{cases}$$

Proof. For the first statement, we only prove its necessary condition by contradiction. The sufficient condition can be obtained in a similar way.

Let $\partial\Upsilon(\xi)$ be nonempty. Suppose there exists an index $i \in \mathbb{N}_d$ such that $\xi_i + \xi_{i+d} < 0$. In this case, $\Upsilon(\xi) = +\infty$, which in turn yields

$$\Upsilon(\xi) + \langle \eta, \varepsilon - \xi \rangle = +\infty, \text{ for all } \varepsilon \in \mathbb{R}^{2d}_{+} \quad \text{and} \quad \eta \in \mathbb{R}^{2d}.$$

Note that $\Upsilon(\varepsilon) = 0$, and we have that $\Upsilon(\xi) + \langle \eta, \varepsilon - \xi \rangle > \Upsilon(\varepsilon)$. Recalling the definition of subdifferential, the above inequality contradicts the nonemptiness of $\partial\Upsilon(\xi)$.

Next, we characterize the subdifferential of Υ. For nonempty $\partial\Upsilon(\xi)$, let $\eta \in \partial\Upsilon(\xi)$. Again by the definition of subdifferential, we observe that

$$\Upsilon(\varepsilon) \geq \langle \eta, \varepsilon - \xi \rangle + \Upsilon(\xi), \text{ for all } \varepsilon \in \mathbb{R}^{2d}. \tag{4.12}$$

In what follows, we restrict ε to the domain of Υ. In this case, $\varepsilon_i + \varepsilon_{i+d} \geq 0$, for $i \in \mathbb{N}_d$, and $\Upsilon(\varepsilon) = 0$. (4.12) then reduces to the following component form

$$0 \geq \sum_{i=1}^{d} \Big(\eta_i \left(\varepsilon_i - \xi_i\right) + \eta_{i+d} \left(\varepsilon_{i+d} - \xi_{i+d}\right) \Big). \tag{4.13}$$

(4.13) holds true for each term, that is, for $i \in \mathbb{N}_d$, we have that

$$0 \geq \eta_i(\varepsilon_i - \xi_i) + \eta_{i+d}(\varepsilon_{i+d} - \xi_{i+d}). \tag{4.14}$$

Otherwise, suppose there exists an index j such that $\eta_j(\varepsilon_j - \xi_j) + \eta_{j+d}(\varepsilon_{j+d} - \xi_{j+d})$ is greater than 0. Under these circumstances, we let $\widetilde{\varepsilon}_i = \xi_i$ for $i \neq j$, and $\widetilde{\varepsilon}_j = \varepsilon_j$. Then $\widetilde{\varepsilon}$ does not satisfy (4.13), which is a contradiction.

For each index $i \in \mathbb{N}_d$, we next explore (4.14) for two cases. When $\xi_i + \xi_{i+d} = 0$, we may rewrite (4.14) as

$$0 \geq \eta_i(\varepsilon_i + \varepsilon_{i+d}) + (\eta_{i+d} - \eta_i)\,(\varepsilon_{i+d} - \xi_{i+d}). \tag{4.15}$$

Estimate (4.15) holds true for all possible ε if and only if $\eta_i = \eta_{i+d}$ and $\eta_i \leq 0$.

When $\xi_i + \xi_{i+d} > 0$, (4.14) reads

$$0 \geq \eta_i\left((\varepsilon_i + \varepsilon_{i+d}) - (\xi_i + \xi_{i+d})\right) + (\eta_{i+d} - \eta_i)\,(\varepsilon_{i+d} - \xi_{i+d})\,, \tag{4.16}$$

which is equivalent to the fact that $\eta_i = \eta_{i+d} = 0$.

All the arguments discussed above are reversible. This completes the proof. □

As a result of Lemma 4.2, we next calculate the closed form of $\mathrm{prox}_{\Upsilon}^{P}$ using the relationship between the subdifferential and the proximity operator given in (3.2).

Lemma 4.3. *If $\xi,\ \eta \in \mathbb{R}^{2d}$ is given such that $\eta = \mathrm{prox}_{\Upsilon}(\xi)$, then for $i \in \mathbb{N}_d$,*

$$\eta_i = \begin{cases} \xi_i, & \text{if} \quad \xi_i + \xi_{i+d} > 0, \\ \frac{\xi_i - \xi_{i+d}}{2}, & \text{if} \quad \xi_i + \xi_{i+d} \leq 0, \end{cases} \quad \text{and} \quad \eta_{i+d} = \begin{cases} \xi_{i+d}, & \text{if} \quad \xi_i + \xi_{i+d} > 0, \\ \frac{\xi_{i+d} - \xi_i}{2}, & \text{if} \quad \xi_i + \xi_{i+d} \leq 0. \end{cases} \tag{4.17}$$

Proof. By the definition of the proximity operator, we have that

$$\eta = \arg\min\left\{\frac{1}{2}\|\varepsilon - \xi\|_2^2 + \Upsilon(\varepsilon) : \varepsilon \in \mathbb{R}^{2d}\right\}. \tag{4.18}$$

Applying Fermat's rule to model (4.18), we obtain the relation

$$\xi - \eta \in \partial\Upsilon\,(\eta)\,. \tag{4.19}$$

Invoking Lemma 4.2, we know that $\eta \in \mathbb{R}^{2d}$ satisfies inclusion (4.19) if and only if for $i \in \mathbb{N}_d$, the following relations hold

$$\begin{cases} \xi_i - \eta_i = \xi_{i+d} - \eta_{i+d} = 0, & \text{if} \quad \eta_i + \eta_{i+d} > 0, \\ \xi_i - \eta_i = \xi_{i+d} - \eta_{i+d} \leq 0, & \text{if} \quad \eta_i + \eta_{i+d} = 0. \end{cases} \tag{4.20}$$

Next, we explore (4.20) for two cases. If $\eta_i + \eta_{i+d} > 0$, we have that $\xi_i = \eta_i$ and $\xi_{i+d} = \eta_{i+d}$. In this case, $\xi_i + \xi_{i+d}$ is also guaranteed to be positive. In the case that $\eta_i + \eta_{i+d} = 0$, we can easily verify that

$$\eta_i = \frac{\xi_i - \xi_{i+d}}{2} \quad \text{and} \quad \eta_{i+d} = \frac{\xi_{i+d} - \xi_i}{2}.$$

Under these circumstances, $\xi_i + \xi_{i+d}$ is nonpositive.

In summary, the above arguments together with the uniqueness of the solution of model (4.18) complete the proof. □

Proposition 4.1. *Let $P := diag\{\lambda_j\}$ be a $2d \times 2d$ diagonal matrix with positive diagonal entries. If $\xi, \eta \in \mathbb{R}^{2d}$ satisfy $\eta = \mathrm{prox}_{\Upsilon}^{P}(\xi)$, then for all $i \in \mathbb{N}_d$*

$$\eta_i = \begin{cases} \xi_i, & \text{if} \quad \xi_i + \xi_{i+d} > 0, \\ \frac{\lambda_i\xi_i - \lambda_{i+d}\xi_{i+d}}{\lambda_i + \lambda_{i+d}}, & \text{if} \quad \xi_i + \xi_{i+d} \leq 0, \end{cases} \quad \text{and} \quad \eta_{i+d} = \begin{cases} \xi_{i+d}, & \text{if} \quad \xi_i + \xi_{i+d} > 0, \\ \frac{\lambda_{i+d}\xi_{i+d} - \lambda_i\xi_i}{\lambda_i + \lambda_{i+d}}, & \text{if} \quad \xi_i + \xi_{i+d} \leq 0. \end{cases}$$

Proof. This result can be obtained by a similar computation performed in Lemma 4.3. □

Next, we present specific expressions of B_1, B_2, φ_1 and φ_2, and describe the closed forms of proximity operators $\mathrm{prox}_{\varphi_1}$, $\mathrm{prox}_{\varphi_2}$ and $\mathrm{prox}^{Q}_{\Phi^*}$.

We assume that the dimension of a three-dimensional image considered in this article is $s \times s \times t$. This image is reshaped into a vector in $\mathbb{R}^{s^2t}$ in such a way that the (ijk)th pixel of the image, where $i, j \in \mathbb{N}_s$ and $k \in \mathbb{N}_t$, corresponds to the $(i + (j-1)s + (k-1)s^2)$th component of the vector in $\mathbb{R}^{s^2t}$. Since the radioactivity distribution f is a vector in $\mathbb{R}^d$, we have that $d = s^2t$. For a positive integer α, we define an $\alpha \times \alpha$ first-order difference matrix D_α by

$$D_\alpha := \begin{pmatrix} 0 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix}.$$

Employing the Kronecker product $\otimes$, we define the $m_1 \times d$ matrix B_1 and $m_2 \times d$ matrix B_2 by

$$B_1 := \begin{pmatrix} I_t \otimes I_s \otimes D_s \\ I_t \otimes D_s \otimes I_s \\ D_t \otimes I_s \otimes I_s \end{pmatrix},$$

$$B_2 := \begin{bmatrix} D_{xx}^\top & D_{xy}^\top & D_{xz}^\top & D_{yx}^\top & D_{yy}^\top & D_{yz}^\top & D_{zx}^\top & D_{zy}^\top & D_{zz}^\top \end{bmatrix}^\top,$$

where

$$\begin{aligned} &D_{xx} = I_t \otimes I_s \otimes -D_s^\top D_s, \quad D_{xy} = I_t \otimes -D_s^\top \otimes D_s, \quad D_{xz} = -D_t^\top \otimes I_s \otimes D_s, \\ &D_{yx} = I_t \otimes D_s \otimes -D_s^\top, \quad D_{yy} = I_t \otimes -D_s^\top D_s \otimes I_s, \quad D_{yz} = -D_t^\top \otimes D_s \otimes I_s, \\ &D_{zx} = D_t \otimes I_s \otimes -D_s^\top, \quad D_{zy} = D_t \otimes -D_s^\top \otimes I_s, \quad D_{zz} = -D_t^\top D_t \otimes I_s \otimes I_s. \end{aligned}$$

From the definitions of B_1 and B_2, we know that $m_1 = 3d$ and $m_2 = 9d$. The convex functions $\varphi_1 : \mathbb{R}^{m_1} \to \mathbb{R}$ and $\varphi_2 : \mathbb{R}^{m_2} \to \mathbb{R}$ are defined at $z \in \mathbb{R}^{m_1}$ and $\xi \in \mathbb{R}^{m_2}$ as

$$\varphi_1(z) := \sum_{i=1}^{d} \sqrt{z_i^2 + z_{i+d}^2 + z_{i+2d}^2},$$

$$\varphi_2(\xi) := \sum_{i=1}^{d} \left(\sum_{j=0}^{8} \xi_{i+jd}^2 \right)^{\frac{1}{2}}.$$

The proximity operators of φ_1 and φ_2 can be calculated explicitly. For a positive number μ, vectors $z \in \mathbb{R}^{m_1}$ and $\xi \in \mathbb{R}^{m_2}$, $y := \mathrm{prox}_{\mu\varphi_1}(z)$ and $\eta := \mathrm{prox}_{\mu\varphi_2}(\xi)$ can be computed component-wise by, for $i \in \mathbb{N}_d$,

$$[y_i, y_{i+d}, y_{i+2d}]^\top = \max\left\{ \|[z_i, z_{i+d}, z_{i+2d}]^\top\| - \mu, 0 \right\} \frac{[z_i, z_{i+d}, z_{i+2d}]^\top}{\|[z_i, z_{i+d}, z_{i+2d}]^\top\|},$$

$$[\eta_i, \eta_{i+2d}, \cdots, \eta_{i+8d}]^\top = \max\left\{ \|[\xi_i, \xi_{i+2d}, \cdots, \xi_{i+8d}]^\top\| - \mu, 0 \right\} \frac{[\xi_i, \xi_{i+2d}, \cdots, \xi_{i+8d}]^\top}{\|[\xi_i, \xi_{i+2d}, \cdots, \xi_{i+8d}]^\top\|}.$$

Because $\Phi(z, \xi)$ is separable for z and ξ, $\Phi^*(z, \xi)$ is also separable for z and ξ. Recalling the definition (4.10) of Q, we have, for any $z \in \mathbb{R}^{m_1}$ and $\xi \in \mathbb{R}^{m_2}$ that

$$\mathrm{prox}^{Q}_{\Phi^*}(z, \xi) = \left(\mathrm{prox}_{\mu_1(\lambda_1\varphi_1)^*}(z), \mathrm{prox}_{\mu_2(\lambda_2\varphi_2)^*}(\xi) \right),$$

where, by Moreau's decomposition theorem [2, Theorem 14.3],

$$\mathrm{prox}_{\mu_i(\lambda_i\varphi_i)^*}(\cdot) = \mu_i\left(I - \mathrm{prox}_{\frac{\lambda_i}{\mu_i}\varphi_i}\right)(\mu_i^{-1}\cdot), \quad i = 1, 2. \tag{4.21}$$

Finally, we consider the choice of the preconditioning matrix P. In [14], it has been shown that the EM-type preconditioning matrix is capable of significantly accelerating the convergence rate. Hence, we propose the following algorithm (Algorithm 4.1) for solving the ICTV model (2.7).

Algorithm 4.1

Initialization: u^0, b^0.
repeat
Step 1: $S_1^k \leftarrow diag(\frac{u_1^k}{A^\top\cdot 1}), S_2^k \leftarrow diag(\frac{u_2^k}{A^\top\cdot 1})$
$P^k \leftarrow \begin{pmatrix} S_1^k & \\ & S_2^k \end{pmatrix}$
Step 2: $b^{k+1} \leftarrow \mathrm{prox}_{\Phi^*}^{Q}\left(b^k + Q^{-1}Bu^k\right)$
Step 3: $u^{k+1} \leftarrow \mathrm{prox}_{\Upsilon}^{P^k}\left(u^k - \left(P^k\right)^{-1}\left(\nabla\Psi(u^k) + B^\top\left(2b^{k+1} - b^k\right)\right)\right)$
until "convergence"
Write the output of u^k from the above loop as u^∞ and compute $u_1^\infty + u_2^\infty$ as the reconstructed image

In Algorithm 4.1, the preconditioning matrix P is updated at every iteration. However, as we shall see in Section 6, the values of the spectral norms of the preconditioning matrices S_1 and S_2 almost remain constant after some iteration. Therefore, we may fix the preconditioning matrices S_1 and S_2 after some iteration, and then obtain the semi-dynamic version of Algorithm 4.1 (Algorithm 4.2). As a result, the convergence theorem given in Section 5 can be applied.

Algorithm 4.2

Preparation: The parameter l is a positive integer.
Initialization: u^0, b^0.
Run the loop in Algorithm 4.1 until $k > l$
Set $S_1^l \leftarrow diag(\frac{u_1^l}{A^\top\cdot 1}), S_2^l \leftarrow diag(\frac{u_2^l}{A^\top\cdot 1})$
$P \leftarrow \begin{pmatrix} S_1^l & \\ & S_2^l \end{pmatrix}$
repeat
Step 1: $b^{k+1} \leftarrow \mathrm{prox}_{\Phi^*}^{Q}\left(b^k + Q^{-1}Bu^k\right)$
Step 2: $u^{k+1} \leftarrow \mathrm{prox}_{\Upsilon}^{P}\left(u^k - P^{-1}\left(\nabla\Psi(u^k) + B^\top\left(2b^{k+1} - b^k\right)\right)\right)$
until "convergence"
Write the output of u^k from the above loop as u^∞ and compute $u_1^\infty + u_2^\infty$ as the reconstructed image

5. Convergence Analysis

In this section, we establish the convergence result of the iterative scheme (4.8) for solving the ICTV model (2.7).

We shall prove that for the chosen matrix M (4.9), the sequence generated by the iterative scheme (4.8) converges to a fixed-point of the operator $\mathcal{T}_M$ defined below, and hence to a solution of the the ICTV model (2.7). As a key ingredient in the convergence analysis, we shall establish the following estimate

$$\|\nu^N - \nu\|_M^2 + \sum_{k=0}^{N-1} \|\nu^{k+1} - \nu^k\|_{\widetilde{M}}^2 \leq \|\nu^0 - \nu\|_M^2,$$

where $\{\nu^k : k \in \mathbb{N}\}$ is the sequence generated by the iterative scheme (4.8) and the expression of matrix $\widetilde{M}$ will be clarified later.

First, we introduce a set-valued operator from scheme (4.8). Suppose that for any $w \in \mathbb{R}^{m+2d}$, there exists a vector $\nu \in \mathbb{R}^{m+2d}$ satisfying

$$\nu = \mathcal{T}\left((E - R^{-1}M)\nu + R^{-1}Mw - R^{-1}\mathcal{G}(w)\right). \tag{5.1}$$

(5.1) implicitly defines a non-linear operator mapping from w to ν. Specifically, we denote by $2^{\mathbb{R}^{m+2d}}$ the family of all subsets of $\mathbb{R}^{m+2d}$. Then we define the set-valued operator $\mathcal{T}_M : \mathbb{R}^{m+2d} \to 2^{\mathbb{R}^{m+2d}}$ at $w \in \mathbb{R}^{m+2d}$ as

$$\mathcal{T}_M(w) := \Big\{\nu \in \mathbb{R}^{m+2d} : (w, \nu) \text{ satisfies (5.1)}\Big\}. \tag{5.2}$$

Next, we define the continuity of the general set-valued operator. For an operator $\mathcal{H} : \mathbb{R}^{m+2d} \to 2^{\mathbb{R}^{m+2d}}$, we denote by $\mathrm{gra}(\mathcal{H})$ the graph of $\mathcal{H}$, defined by

$$\mathrm{gra}(\mathcal{H}) := \Big\{(w, \nu) \in \mathbb{R}^{m+2d} \times \mathbb{R}^{m+2d} : \nu \in \mathcal{H}(w)\Big\}. \tag{5.3}$$

We say that operator $\mathcal{H}$ is continuous if its graph is a closed set [2]. Moreover, given a vector $\nu \in \mathbb{R}^{m+2d}$ satisfying the inclusion $\nu \in \mathcal{H}(\nu)$, we say that ν is a fixed-point of $\mathcal{H}$. This is a generalization of the traditional concept of fixed-points.

Let $L_{\nabla\Psi} := \|AC\|_2\|g\|_\infty/\gamma^2$. It is easy to verify the following result.

Lemma 5.1. *The function Ψ in the ICTV model (2.7) is differentiable in $\mathbb{R}_+^{2d}$ with a Lipschitz continuous gradient. Furthermore, for any $\xi, \eta \in \mathbb{R}_+^{2d}$, we have that*

$$\|\nabla\Psi(\xi) - \nabla\Psi(\eta)\|_2 \leq L_{\nabla\Psi}\|\xi - \eta\|_2.$$

We now discuss two properties of $\mathcal{T}_M$.

Lemma 5.2. *Let $P \in \mathbb{S}_+^{2d}$ and $Q \in \mathbb{S}_+^m$. If $\mathcal{T}_M$ is defined by (5.2) and the vector pairs $(w_i, \nu_i) \in \mathrm{gra}(\mathcal{T}_M)$ for $i = 1, 2$, then*

$$\langle \nu_2 - \nu_1, M(\nu_2 - \nu_1)\rangle \leq \langle \nu_2 - \nu_1, M(w_2 - w_1)\rangle - \langle \nu_2 - \nu_1, \mathcal{G}(w_2) - \mathcal{G}(w_1)\rangle. \tag{5.4}$$

Moreover, $\mathcal{T}_M$ is continuous.

Proof. We first prove (5.4). Recalling the definition of operator $\mathcal{T}_M$, we know that for $i = 1, 2$, there holds

$$\nu_i = \mathcal{T}\left((E - R^{-1}M)\nu_i + R^{-1}Mw_i - R^{-1}\mathcal{G}(w_i)\right).$$

By the firm nonexpansivity of operator $\mathcal{T}$ with respect to matrix R and the expression of matrix E in (4.5), we further have that

$$\begin{aligned}
&\langle \nu_2 - \nu_1, R(\nu_2 - \nu_1)\rangle \\
\leq\ &\langle \nu_2 - \nu_1, R(\nu_2 - \nu_1)\rangle + \langle \nu_2 - \nu_1, S_B(\nu_2 - \nu_1)\rangle - \langle \nu_2 - \nu_1, M(\nu_2 - \nu_1)\rangle \\
&+\langle \nu_2 - \nu_1, M(w_2 - w_1)\rangle - \langle \nu_2 - \nu_1, \mathcal{G}(w_2) - \mathcal{G}(w_1)\rangle.
\end{aligned} \tag{5.5}$$

Since matrix S_B is skew symmetric, we can verify that $\langle \nu_2 - \nu_1, S_B(\nu_2 - \nu_1)\rangle = 0$. Substituting this equation into (5.5) yields (5.4).

Next, we prove the continuity of $\mathcal{T}_M$. Let $\{(w^n, \nu^n) : n \in \mathbb{N}\}$ be an arbitrary convergent sequence in $\mathrm{gra}(\mathcal{T}_M)$, and let $(w, \nu) \in \mathbb{R}^{m+2d} \times \mathbb{R}^{m+2d}$ be its limit. Then we have that

$$\nu^n = \mathcal{T}\left((E - R^{-1}M)\nu^n + R^{-1}Mw^n - R^{-1}\mathcal{G}(w^n)\right).$$

Set $\widetilde{\nu} := \mathcal{T}\left((E - R^{-1}M)\nu + R^{-1}Mw - R^{-1}\mathcal{G}(w)\right)$. Then the firm nonexpansivity of $\mathcal{T}$ with respect to R ensures that

$$\|\nu^n - \widetilde{\nu}\|_R^2 \le \langle \nu^n - \widetilde{\nu}, (RE - M)(\nu^n - \nu) + M(w^n - w) - (\mathcal{G}(w^n) - \mathcal{G}(w))\rangle. \tag{5.6}$$

Recalling the definition of operator $\mathcal{G}$ in (4.6), the Lipschitz continuity of the gradient of Ψ ensures the strong convergence of sequence $\{\mathcal{G}(w^n) : n \in \mathbb{N}\}$ to $\mathcal{G}(w)$. This together with (5.6) implies that $\widetilde{\nu}$ is also a limit of sequence $\{\nu^n : n \in \mathbb{N}\}$. Therefore, we have $\nu = \widetilde{\nu}$, which indicates that $(w, \nu) \in \mathrm{gra}(\mathcal{T}_M)$. This completes the proof. □

Next, we present a property regarding the gradient of Ψ, which is crucial in the convergence analysis.

Lemma 5.3. *For any ξ, η, $\nu \in dom(\Psi)$, there holds*

$$-\langle \xi - \eta, \nabla\Psi(\nu) - \nabla\Psi(\eta)\rangle \le \langle \xi - \nu, \nabla\Psi(\xi) - \nabla\Psi(\nu)\rangle. \tag{5.7}$$

Proof. By spitting the left hand side of (5.7), we have that

$$\begin{aligned}
&-\langle \xi - \eta, \nabla\Psi(\nu) - \nabla\Psi(\eta)\rangle \\
&= \langle \xi - \nu, \nabla\Psi(\xi) - \nabla\Psi(\nu)\rangle - \langle \xi - \nu, \nabla\Psi(\xi) - \nabla\Psi(\eta)\rangle - \langle \nu - \eta, \nabla\Psi(\nu) - \nabla\Psi(\eta)\rangle \\
&= \langle \xi - \nu, \nabla\Psi(\xi) - \nabla\Psi(\nu)\rangle + \langle \nu - \xi, \nabla\Psi(\xi)\rangle + \langle \eta - \nu, \nabla\Psi(\nu)\rangle + \langle \xi - \eta, \nabla\Psi(\eta)\rangle.
\end{aligned} \tag{5.8}$$

We further employ the convexity of Ψ to obtain

$$\begin{aligned}
\langle \nu - \xi, \nabla\Psi(\xi)\rangle &\le \Psi(\nu) - \Psi(\xi), \\
\langle \eta - \nu, \nabla\Psi(\nu)\rangle &\le \Psi(\eta) - \Psi(\nu), \\
\langle \xi - \eta, \nabla\Psi(\eta)\rangle &\le \Psi(\xi) - \Psi(\eta).
\end{aligned}$$

Substituting the above estimates into (5.8) leads to the desired result (5.7). □

Directly combining Lemma 5.3 with the fact that the gradient of Ψ is $L_{\nabla\Psi}$-Lipschitz continuous, we get the following estimate.

Lemma 5.4. *If $\xi, \eta, \nu \in \mathbb{R}^{2d}$, then*

$$-\langle \xi - \eta, \nabla\Psi(\nu) - \nabla\Psi(\eta)\rangle \le L_{\nabla\Psi}\langle \xi - \nu, \xi - \nu\rangle.$$

We define a matrix $\widetilde{M}$, which will be used later

$$\widetilde{M} := M - \begin{pmatrix} 0 & \\ & 2L_{\nabla\Psi} I \end{pmatrix}. \tag{5.9}$$

Now, we give a key estimate of the sequence $\{\nu^k : k \in \mathbb{N}\}$ generated by (4.8). It is crucial for the convergence analysis.

Lemma 5.5. *If both M and $\widetilde{M}$ are positive definite and symmetric matrices, and ν is a solution of (4.7), then for any initial value $\nu^0 \in \mathbb{R}^{2d}$, the sequence $\{\nu^k : k \in \mathbb{N}\}$ generated by (4.8) satisfies,*

$$\|\nu^N - \nu\|_M^2 + \sum_{k=0}^{N-1} \|\nu^{k+1} - \nu^k\|_{\widetilde{M}}^2 \leq \|\nu^0 - \nu\|_M^2, \quad \text{for every } N \in \mathbb{N}. \tag{5.10}$$

Proof. Since ν is a solution of (4.7), ν is a fixed-point of $\mathcal{T}_M$. Let k be a positive integer. Identifying ν_1, ν_2, w_1 and w_2 in (5.4), respectively, with ν, ν_{k+1}, ν and ν_k, and employing Lemma 5.2, we observe that

$$\langle \nu^{k+1} - \nu, M(\nu^{k+1} - \nu)\rangle \leq \langle \nu^{k+1} - \nu, M(\nu^k - \nu)\rangle - \langle \nu^{k+1} - \nu, \mathcal{G}(\nu^k) - \mathcal{G}(\nu)\rangle.$$

Multiplying the above inequality by 2 and writing the term $\langle \nu^{k+1} - \nu, M(\nu^k - \nu)\rangle$ as $\langle \nu^{k+1} - \nu, M(\nu^k - \nu^{k+1})\rangle + \langle \nu^{k+1} - \nu, M(\nu^{k+1} - \nu)\rangle$, we obtain that

$$\begin{aligned} &2\langle \nu^{k+1} - \nu, M(\nu^{k+1} - \nu)\rangle \\ &\leq \langle \nu^{k+1} - \nu, M(\nu^{k+1} - \nu)\rangle + \langle \nu^{k+1} - \nu, M(\nu^k - \nu^{k+1})\rangle \\ &\quad + \langle \nu^{k+1} - \nu, M(\nu^k - \nu)\rangle - 2\langle \nu^{k+1} - \nu, \mathcal{G}(\nu^k) - \mathcal{G}(\nu)\rangle. \end{aligned} \tag{5.11}$$

We further split the inner product $\langle \nu^{k+1} - \nu, M(\nu^k - \nu^{k+1})\rangle$ at the right hand side of (5.11) as $\langle \nu^{k+1} - \nu^k, M(\nu^k - \nu^{k+1})\rangle + \langle \nu^k - \nu, M(\nu^k - \nu^{k+1})\rangle$, and recall the symmetry of matrix M to conclude that

$$\begin{aligned} &\langle \nu^{k+1} - \nu, M(\nu^{k+1} - \nu)\rangle \\ &\leq \langle \nu^k - \nu, M(\nu^k - \nu)\rangle - \langle \nu^{k+1} - \nu^k, M(\nu^{k+1} - \nu^k)\rangle - 2\langle \nu^{k+1} - \nu, \mathcal{G}(\nu^k) - \mathcal{G}(\nu)\rangle. \end{aligned} \tag{5.12}$$

Recalling the definition of $\mathcal{G}$ and applying Lemma 5.4, we have the estimate

$$-2\langle \nu^{k+1} - \nu, \mathcal{G}(\nu^k) - \mathcal{G}(\nu)\rangle \leq \langle u^{k+1} - u^k, 2L_{\nabla\Psi}(u^{k+1} - u^k)\rangle. \tag{5.13}$$

Combining (5.12) with estimate (5.13) leads to

$$\langle \nu^{k+1} - \nu, M(\nu^{k+1} - \nu)\rangle \leq \langle \nu^k - \nu, M(\nu^k - \nu)\rangle - \langle \nu^{k+1} - \nu^k, \widetilde{M}(\nu^{k+1} - \nu^k)\rangle. \tag{5.14}$$

Finally, summing (5.14) for k running from 0 to $N-1$ and recalling the positive definiteness of matrices M and $\widetilde{M}$, we obtain the desired estimate (5.10). □

Now, we are ready to give the main convergence result.

Theorem 5.1. *If M and $\widetilde{M}$ are both positive definite, then for any initial value $\nu^0 \in \mathbb{R}^{2d}$, the sequence $\{\nu^k : k \in \mathbb{N}\}$ generated by (4.8) converges to a solution of the ICTV model (2.7).*

Proof. Since N in (5.10) is arbitrary, we have that sequence $\{\nu^k : k \in \mathbb{N}\}$ is bounded and $\sum_{k=0}^{\infty} \|\nu^{k+1} - \nu^k\|_{\widetilde{M}}^2$ converges, ensuring that $\{\nu^k : k \in \mathbb{N}\}$ converges. Due to the continuity of $\mathcal{T}_M$, the limit point of $\{\nu^k : k \in \mathbb{N}\}$ is a fixed-point of $\mathcal{T}_M$, which means that $\{\nu^k : k \in \mathbb{N}\}$ converges to a solution of the ICTV model (2.7). □

It remains to investigate the conditions that M and $\widetilde{M}$ are positive definite.

Lemma 5.6. *If for some $\varepsilon \in (0,1)$ and $i = 1, 2$, μ_i and β_i satisfy*

$$0 < \beta_i < \frac{(1-\varepsilon)\gamma^2}{2\|AC\|_2\|S_i\|_2\|g\|_\infty}, \tag{5.15}$$

$$0 < \mu_i\beta_i < \frac{\varepsilon}{\|B_i\|_2^2\|S_i\|_2}, \tag{5.16}$$

then matrices M and $\widetilde{M}$ are positive definite.

Proof. By the definition of matrix $\widetilde{M}$, we know that the positive definiteness of $\widetilde{M}$ implies that of M. Hence, it suffices to verify that matrix $\widetilde{M}$ is positive definite.

Let $G := Q^{-1/2}B(P-2L_{\nabla\Psi}I)^{-1/2}$. Matrix $\widetilde{M}$ can be decomposed as

$$\begin{pmatrix} Q^{1/2} & \\ & (P-2L_{\nabla\Psi}I)^{1/2} \end{pmatrix} \begin{pmatrix} I & -G \\ & I \end{pmatrix}^{-1} \begin{pmatrix} I-GG^\top & \\ & I \end{pmatrix} \begin{pmatrix} I & \\ -G^\top & I \end{pmatrix}^{-1} \begin{pmatrix} Q^{1/2} & \\ & (P-2L_{\nabla\Psi}I)^{1/2} \end{pmatrix}^\top.$$

Therefore, it suffices to show that matrix $P-2L_{\nabla\Psi}I$ is positive definite and $\|G\|_2 < 1$.

Noting that

$$P-2L_{\nabla\Psi}I = P^{1/2}\left(I-2L_{\nabla\Psi}P^{-1}\right)\left(P^{1/2}\right)^\top,$$

we consider matrix $I-2L_{\nabla\Psi}P^{-1}$. Recalling the definitions of $L_{\nabla\Psi}$ and preconditioning matrix P, we obtain that all the eigenvalues of matrix $I-2L_{\nabla\Psi}P^{-1}$ are larger than ε if β_1 and β_2 satisfy (5.15). As a result, matrix $P-2L_{\nabla\Psi}I$ is positive definite.

Next, we verify that $\|G\|_2 < 1$. Recalling the block forms of Q, B and P, we have that

$$\begin{aligned} G &= \begin{pmatrix} \mu_1^{1/2}I_1 & \\ & \mu_2^{1/2}I_2 \end{pmatrix} \begin{pmatrix} B_1 & \\ & B_2 \end{pmatrix} \begin{pmatrix} \beta_1^{-1}S_1^{-1}-2L_{\nabla\Psi}I_1 & \\ & \beta_2^{-1}S_2^{-1}-2L_{\nabla\Psi}I_2 \end{pmatrix}^{-1/2} \\ &= \begin{pmatrix} \mu_1^{1/2}B_1(\beta_1^{-1}S_1^{-1}-2L_{\nabla\Psi}I_1)^{-1/2} & \\ & \mu_2^{1/2}B_2(\beta_2^{-1}S_2^{-1}-2L_{\nabla\Psi}I_2)^{-1/2} \end{pmatrix}. \end{aligned}$$

Noting that

$$\mu_i^{1/2}B_i(\beta_i^{-1}S_i^{-1}-2L_{\nabla\Psi}I_i)^{-1/2} = (\mu_i\beta_i)^{1/2}\,B_i(I-2L_{\nabla\Psi}\beta_iS_i)^{-1/2}S_i^{1/2},$$

it suffices to show

$$\|\,(\mu_i\beta_i)^{1/2}\,B_i(I-2L_{\nabla\Psi}\beta_iS_i)^{-1/2}S_i^{1/2}\|_2 < 1, \quad i=1,2. \tag{5.17}$$

Since all the eigenvalues of $I-2L_{\nabla\Psi}P^{-1}$ are larger than ε, we know that

$$\|(I-2L_{\nabla\Psi}\beta_iS_i)^{-1}\|_2 < \frac{1}{\varepsilon}, \quad i=1,2. \tag{5.18}$$

This together with condition (5.16) yields the desired result. □

The culmination of our effort is the following convergence result.

Theorem 5.2. *Let λ_1 and λ_2 be positive numbers. If μ_1, μ_2, β_1 and β_2 are chosen to satisfy*

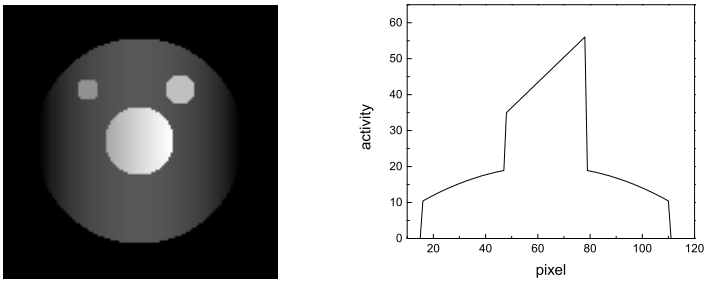
$$0 < \beta_i < \frac{(1-\varepsilon)\gamma^2}{2\|AC\|_2\|S_i\|_2\|g\|_\infty} \quad \text{and} \quad 0 < \mu_i\beta_i < \frac{\varepsilon}{\|B_i\|_2^2\|S_i\|_2}, \quad i=1,2,$$

for some $\varepsilon \in (0,1)$, then for any initial value $u^0 \in \mathbb{R}^{2d}$, the sequence $\{u^k : k \in \mathbb{N}\}$ generated by Algorithm 4.2 converges to a solution of the ICTV model (2.7).

Proof. This result follows immediately from the combination of Theorem 5.1 and Lemma 5.6. □

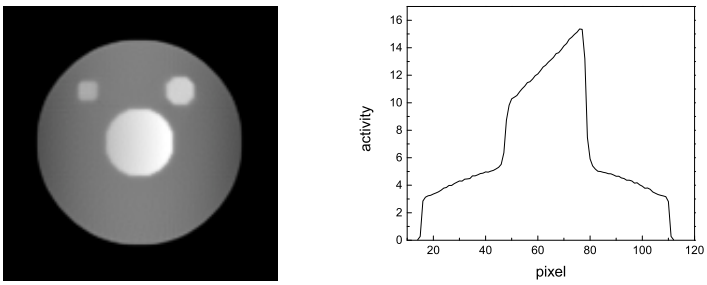
6. Numerical Experiments

In this section, we report numerical results for the TV model (2.4) and the ICTV model (2.7). We first compare staircasing artifacts of the constructed images obtained from these



(a) the phantom (b) central horizontal line profile

Fig. 6.1. (a) slice 26 of the phantom; (b) its central horizontal line profile.



(a) figure (b) central horizontal line profile

Fig. 6.2. (a) slice 26 of the reconstruction from noiseless projection using the MLEM algorithm with 40 iterations ; (b) its central horizontal line profile.

models, and then compare their image quality using signal-to-noise ratio (SNR) and contrast recovery coefficient (CRC).

We conducted studies over a computer generated phantom, to investigate the properties of the TV model (2.4) and the ICTV model (2.7). The phantom array is $128 \times 128 \times 64$ and the array of each slice 128×128. In each slice of the phantom between slices 12 and 30, there are three small discs contained in a large disc. The large disc, used as background, has curved intensity. The intensity of the central disc changes linearly along the horizontal axis, with the slope being 0.7. The remaining two small discs located in the upper part of the background, have uniform intensities. In Fig. 6.1 we show a representative slice (slice 26) and the central horizontal line profile of the phantom.

The parallel-collimator SPECT projection data had 120 views in 256×128 matrix with pixel size 1.78mm. The projection data were generated using analytical pixel-wise discretized projector with 20 rays per detector bin [30]. Neither attenuation nor scatter was modelled and the detector was set to be ideal. We created noisy images corrupted by Poisson noise by using the Matlab's *imnoise* function. The total number of counts in 120 views was 9.5×10^6. Each image in these projection sets was then down-sampled to a 128×64 matrix with the pixel size 3.56mm.

We reconstructed the noiseless projections using the MLEM algorithm [15] with 40 iterations. The image obtained is used as a standard to be compared with other reconstructions. Slice 26 of this image is shown in Fig. 6.2, together with the central horizontal line profile. The mean number of background counts γ in the TV model (2.4) and the ICTV model (2.7) was set to be 0.01. The TV model was solved using the PAPA algorithm with 400 iterations [14].

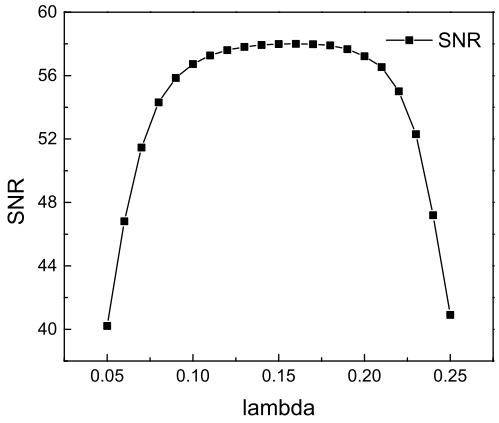


Fig. 6.3. SNR versus weighting parameter λ curve for the TV model.

In the TV model and the ICTV model, the results are sensitive to the regularization weighting parameter(s). In our study, the weighting parameter λ in the TV model (2.4) varies from 0.01 to 1.5. When λ was small ($\lambda \leq 0.07$), the resulting images contained more high frequency noise; when λ was large ($\lambda \geq 0.3$), the hot discs were blurred and contrast reduced. We then used signal-to-noise ratio (SNR), as proposed in [6], to select the optimal λ within the reasonable range $[0.05, 0.25]$, where SNR is defined by

$$\text{SNR} := 20 \ln\left(\frac{\|\tilde{f}\|_2}{\|\tilde{f} - f\|_2}\right),$$

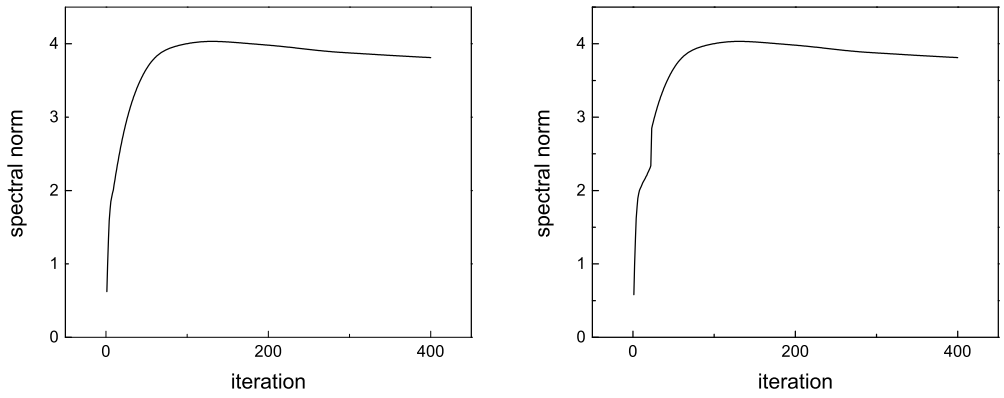
with $\tilde{f}$ and f being the original image and reconstructed image respectively. In this work, $\tilde{f}$ is chosen to be slice 26 of reconstructed image from noiseless projection data using the MLEM algorithm with 40 iterations. And f is chosen to be slice 26 of reconstructed image. The larger the SNR is, the more accurate the reconstructed image is. Fig. 6.3 shows the curve of SNR versus λ for the TV model. When λ was 0.16, SNR achieved its maximum in the range $[0.05, 0.25]$. Hence, the optimal value of λ was chosen to be 0.16.

Solving the ICTV model (2.7) with Algorithm 4.1 required estimation of six parameters: λ_1, λ_2, β_1, β_2, μ_1 and μ_2. We set $\beta_1 = \beta_2 = 1$, and chose μ_1 and μ_2 to be the upper bounds suggested by Theorem 5.2. Specifically, for $i = 1, 2$, we set $\mu_i = 1/\left(\|B_i\|_2^2 \|S_i^k\|_2\right)$ in the kth iteration. The optimal values of weighting parameters λ_1 and λ_2 were obtained by performing sets of trial reconstructions with λ_1 ranging from 0.01 to 1.5 and λ_2 ranging from 10^{-4} to 1.5. We qualitatively (visually) balanced the trade-off between SNR and the severity of staircasing artifacts. This was performed by increasing λ_1 and λ_2 until the artifacts were greatly reduced while avoiding unacceptable SNR loss. As a result, the optimal values of λ_1 and λ_2 were chosen to be 0.09 and 0.08, respectively. The iteration number of Algorithm 4.1 was set to be 400.

We first investigate numerically how the preconditioning matrices S_1 and S_2 (4.10) change in iterations in terms of their spectral norms. Fig. 6.4 displays the curves of the spectral norm versus iterations for S_1 and S_2. It can be seen that the change of the spectral norms of S_1 and S_2 can be neglected after the first few iterations. Hence, S_1 and S_2 may be fixed after some iteration. Consequently, the convergence result in Section 5 can be applied.

In the following, we only use slice 26 for all figures to be shown and all quantities to be calculated.

We now evaluate the quality of the reconstructed images.



(a) (b)

Fig. 6.4. Curves of the spectral norm versus iteration numbers for (a) S_1 and (b) S_2. Reconstructions are obtained using Algorithm 4.1 with $\lambda_1 = 0.09$ and $\lambda_2 = 0.08$.

First, we investigate the ability of the ICTV model to reduce staircasing artifacts. For this purpose, in Fig. 6.5, we plot the central horizontal line profile and draw the contour of the central disc using Matlab's *mesh* function. Fig. 6.5 shows that there are several large steps in the central part of the line profile of the TV model, which are created by staircasing artifacts. While in the corresponding line profile of the ICTV model, these steps disappear and the line profile is much straighter. This means that the ICTV model has greatly reduced staircasing artifacts. This fact is further verified by the contours in Fig. 6.5. The contour of the ICTV model is much smoother and flatter than that of the TV model.

Next, we use SNR to estimate the accuracy of reconstructions. Table 6.1 tells us that the SNR of TV model is slightly higher than that of the ICTV model.

Table 6.1: Results of CRC for the upper right sphere and SNR.

Method	TV	ICTV
SNR	58.00	56.25
CRC	0.87	0.96

Finally, we compare the contrast recovery coefficient (CRC) [1], which serves to measure quantitatively the accuracy of reconstructions. The target ROI should match the cross-sections of the upper right uniform disc, meanwhile avoiding the edges. The background ROIs should be the same size as targets, and we should use 2 of them distributed around the upper right disc and not overlapping the centre disc. The CRC for a hot disc is defined as,

$$\mathrm{CRC} := \frac{m_{obj}/m_{bkg} - 1}{R - 1},$$

where m_{obj} and m_{bkg} are the mean intensities in target and background ROIs respectively, and R is the true disc-to-background intensity concentration ratio. The higher CRC value indicates that the reconstruction for the chosen target ROI is more accurate. As shown in Table 1, the ICTV model has larger CRC compared to the TV model. Therefore, the contrast of upper right disc has been improved by adopting the ICTV model.

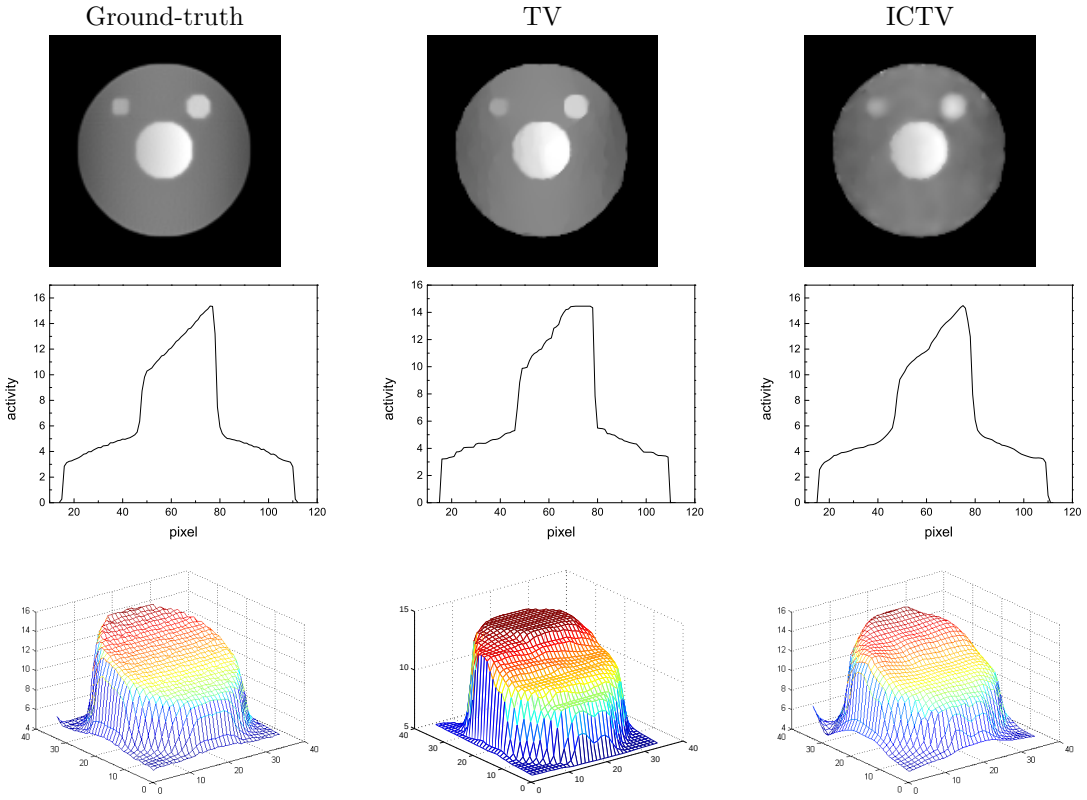


Fig. 6.5. Reconstructions, central horizontal line profiles, and contours of the central hot disc for MLEM from noiseless projection with 40 iterations in Column Ground-truth; the TV model using the PAPA algorithm with 400 iterations and $\lambda = 0.16$ in Column TV; the ICTV model using Algorithm 4.1 with 400 iterations and $\lambda_1 = 0.09$ and $\lambda_2 = 0.08$ in Column ICTV.

7. Conclusions

The TV semi-norm has been widely used in the penalized likelihood reconstruction for SPECT imaging. However, it suffers from staircasing artifacts, which result in cartoon-like images and create faulty edges. To reduce this undesired phenomenon, we propose to use the ICTV model instead of the TV model and develop a novel fixed-point proximity algorithm to solve the underlying model. We prove convergence of the proposed algorithm. We also introduce a preconditioning matrix motivated by the MLEM algorithm to accelerate convergence of the proposed algorithm. Numerical experiments performed in this work show that the ICTV model can effectively reduce staircasing artifacts. Possible future research directions are adopting TGV as a penalty function and developing efficient fixed-point proximity algorithms to solve the resulting optimization problem.

Acknowledgments. This research was supported in part by the Guangdong Provincial Government of China through the "Computational Science Innovative Research Team" program, by the Natural Science Foundation of China under Grants 11471013, 91530117 and 11371333, by the Promotive Research Fund for Excellent Young Scientists of Shandong Province under Grant BS2014DX003, by the Fundamental Research Funds of the Central Universities of O-

cean University of China 201562012, and by the US National Science Foundation under grant DMS-1522332.

References

[1] S. Ahn, S. Ross, E. Asma, J. Miao, X. Jin, L. Cheng, S. Wollenweber and R. Manjeshwar, Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET. *Phys. Med. Biol.*, **60** (2015), 5733-5751.

[2] H.L. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. AMS Books in Mathematics. Springer, New York, 2011.

[3] S. Bonettini and V. Ruggiero, An alternating extragradient method for total variation-based image restoration from poisson data. *Inverse Problems*, **27** (2011), 1-26.

[4] K. Bredies, K. Kunisch and T. Pock, Total generalized variation. *SIAM J. Imaging. Sci.*, **3** (2010), 492-526.

[5] M. Benning, C. Brune, Martin. Burger and J. Muller, Higher-order TV methods-enhancement via Bregman iteration. *J. Sci. Comput.*, **54** (2013), 269-310.

[6] M. Burger, J. Muller, E. Papoutsellis and C. B. Schonlieb, Total variation regularization in measurement and image space for PET reconstruction. *Inverse Problems*, **30** (2014), 1-42.

[7] A. Chambolle and P.-L. Lions, Image recovery via total variation minimization and relaed problems. *Numer. Math.*, **76** (1997), 167-188.

[8] T. Chan, A. Marquina, and P. Mulet, High-order total variation-based image restoration. *SIAM J. Sci. Comput.*, **22** (2000), 503-516.

[9] C. Chen, B. He, Y. Ye and X. Yuan, The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent, *Math. Program.*, **155** (2014), 57-79.

[10] P. Combettes and V. Wajs, Signal recovery by proximal forward-backward splitting. *Multiscale. Model. Sim.*, **4** (2005), 1168-1200.

[11] P. Green, Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE T. Med. Imaging.*, **9** (1990), 84-93.

[12] J.-B. Hiriart-Urruty and C. Lemarechal, *Convex analysis and minimization algorithms I*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1993.

[13] H. Hudson and R. Larkin, Accelerated image reconstruction using ordered subsets of projection data. *IEEE T. Med. Imaging.*, **13** (1994), 601-609.

[14] A. Krol, S. Li, L. Shen, and Y. Xu, Preconditioned alternating projection algorithms for maximum *a posteriori* ECT reconstruction. *Inverse Problems*, **28** (2012), 1-34.

[15] K. Lange and R. Carson, EM reconstruction algorithms for emission and transmission tomography. *J. Comput. Assist. Tomo.*, **8** (1984), 306-316.

[16] Q. Li, C.A. Micchelli, L. Shen and Y. Xu, A proximity algorithm accelerated by Gauss-Seidel iterations for L1/TV denoising models. *Inverse Problems*, **28** (2012), 1-20.

[17] Q. Li, L. Shen, Y. Xu and N. Zhang, Multi-step fixed-point proximity algorithms for solving a class of optimization problems arising from image processing. *Adv. Comput. Math.*, **41** (2015), 387-422.

[18] Q. Li, Y. Xu and N. Zhang, Two-step Fixed-point proximity algorithms for multi-block separable convex problems. *Optimization and Control*, arXiv:1601.01852, 2016.

[19] S. Li, J. Zhang, A. Krol, C. R. Schmidtlein, L. Vogelsang, L. Shen, E. Lipson, D. Feiglin and Y. Xu, Effective noise-suppressed and artifact-reduced reconstruction of SPECT data using a preconditioned alternating projection algorithm. *Med. Phys.*, **42** (2015), 4872-4887.

[20] J. Liow and S. Strother, Practical Tradoffs Between Noise, Quantitation, and Number of Iterations for Maximum Likelihood-Based Reconstructions. *IEEE T. Med. Imaging.*, **10** (1991), 563-571.

[21] C.A. Micchelli, L. Shen, and Y. Xu, Proximity algorithms for image models: Denoising. *Inverse Problems*, **27** (2011), 1-34.

[22] C.A. Micchelli, L. Shen, Y. Xu, and X. Zeng, Proximity algorithms for the L1/TV image denoising model. *Adv. Comput. Math.*, **38** (2013), 401-426.

[23] V. Panin, G. Zeng, and G. Gullberg, Total variation regulated EM algorithm. *IEEE T. Nucl. Sci.*, **46** (1999), 2202-2210.

[24] T. Pock and A. Chambolle, Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *2011 IEEE International Conference on Computer Vision*, IEEE, 2011, pp. 1762-1769.

[25] W. Qi, X. Niu and Y. Yang, Effects of piecewise smoothing on cardiac SPECT reconstruction. In *2011 IEEE International Conference on Image processing*, IEEE, 2011, pp. 457-460.

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

[27] A. Sawatzky, C. Brune, F. Wubbeling, T. Kosters, K. Schafers, and M. Burger, Accurate em-tv algorithm in pet with low snr. In *IEEE Nuclear Science Symposium Conference Record*, IEEE, 2008, pp. 5133-5137.

[28] S. Setzer, G. Steidl and T. Teuber, Infimal convolution regularizations with discrete l1-type functionals. *Commun. Math. Sci.*, **9** (2011), 797-827.

[29] L. Shepp and Y. Vardi, Maximum likelihood reconstruction for emission tomography. *IEEE T. Med. Imaging.*, **1** (1982), 113-122.

[30] L. Vogelsang, *Development of SPECT and CT tomographic image reconstruction.* PhD thesis, Syracuse University, 2011.

[31] G. Zeng. On few-view tomography and staircase artifacts, *IEEE T. Nucl. Sci.*, **62** (2015), 851-858.

[32] H. Ziessman, J. O'Malley, and J. Thrall, *Nuclear Medicine: the Requisites. Third Edition.* Mosby, 2006.