Preconditioned Alternating Projection Algorithms for Maximum *a Posteriori* ECT Reconstruction

Andrzej Krol[*] Si Li[†] Lixin Shen[‡†] Yuesheng Xu[†‡§]

Abstract

We propose a preconditioned alternating projection algorithm (PAPA) for solving the maximum *a posteriori* (MAP) emission computed tomography (ECT) reconstruction problem. Specifically, we formulate the reconstruction problem as a constrained convex optimization problem with the total variation (TV) regularization. We then characterize the solution of the constrained convex optimization problem and show that it satisfies a system of fixed-point equations defined in terms of two proximity operators raised from the convex functions that define the TV-norm and the constrain involved in the problem. The characterization (of the solution) via the proximity operators that define two projection operators naturally leads to an alternating projection algorithm for finding the solution. For efficient numerical computation, we introduce to the alternating projection algorithm a preconditioning matrix (the EM-preconditioner) for the dense system matrix involved in the optimization problem. We prove theoretically convergence of the preconditioned alternating projection algorithm. In numerical experiments, performance of our algorithms, with an appropriately selected preconditioning matrix, is compared with performance of the conventional MAP expectation-maximization (MAP-EM) algorithm with TV regularizer (EM-TV) and that of the recently developed nested EM-TV algorithm for ECT reconstruction. Based on the numerical experiments performed in this work, we observe that the alternating projection algorithm with the EM-preconditioner outperforms significantly the EM-TV in all aspects including the convergence speed, the noise in the reconstructed images and the image quality. It also outperforms the nested EM-TV in the convergence speed while providing comparable image quality.

1 Introduction

Emission computed tomography (ECT) is a noninvasive molecular imaging method that requires administration of radioactive tracers to patients. It comprises of two branches: positron emission tomography (PET) and single-photon emission computed tomography (SPECT) [49, 60]. It might provide estimate of the spatial and/or temporal distribution of radioactive tracer inside a patient body through tomographic reconstruction from the detected emission events (typically in the form of projection images for SPECT and sinograms or list data for PET). ECT provides a three-dimensional (3D) functional rather than structural information provided by computed tomography (CT) or magnetic resonance imaging (MRI). In a SPECT imaging system, the detectors (detector units) record the number of single events due to gamma or x-ray photons emitted by SPECT

---

[*] Department of Radiology, SUNY Upstate Medical University, Syracuse, NY 13210, USA. Email: krola@upstate.edu.

[†] Guangdong Province Key Lab of Computational Science, School of Mathematics and Computational Sciences, Sun Yat-sen University, Guangzhou 510275, P. R. China. E-mail: sili.sysu@gmail.com.

[‡] Department of Mathematics, Syracuse University, Syracuse, NY 13244, USA. Email: lshen03@syr.edu.

[§] Author to whom all correspondence should be sent. E-mail: `yxu06@syr.edu`.

radioactive tracer distributed inside a patient body. The photons are detected only if they travel along directions well defined by a collimator. The collection of such emission data can be sorted into a set of 2D projection images [60]. In PET systems, the pairs of detectors record the number of coincidence events due to two 511 keV gamma photons emitted in positron annihilation events. Positrons are emitted by PET radioactive tracer distributed inside a patient body. They travel certain randomly distributed distances in the tissue before they annihilate electrons and produce pairs of 511 keV photons emitted in random and approximately opposite directions. The collections of such emission data form list data and can be sorted into a set of sinograms [49]. Clinical applications of ECT include detection, staging and monitoring response to cancer therapy, detection and risk stratification of cardiovascular diseases, mapping of regional blood flow in the brain, bone scans, pulmonary ventilation/perfusion scans, and renal scans [63].

The aim of the reconstruction process is to obtain accurate estimation of the radiotracer distribution in a patient or a phantom from the detected emission photons. The most commonly used probability distribution for a description of raw ECT data is the Poisson model [35, 55, 60]. It states that the vector of the number of events recorded by the detector units during ECT scan is a Poisson distributed random vector with a mean equal to the sum of the system matrix multiplied by the mean radiotracer activity distribution vector within an object of interest and by the mean "background" counts vector. Sources of background counts include cosmic rays and terrestrial radioactive background. They are assumed to follow the Poisson distribution. For a given realization of the detected ECT data and for the known expected background counts, this model allows one to estimate the mean radiotracer activity distribution.

Finding numerical solutions of the model is a long-standing research problem. The iterative expectation-maximization (EM) algorithm is a frequently used reconstruction method for ECT imaging [35, 58]. Many reconstruction algorithms for the model are based on optimizing an objective function deduced in part from the statistical model of detection realization data. For example, the maximum-likelihood (ML) method is based on minimizing the negative log-likelihood of observed emission data conditional on radiotracer distribution. Under the Bayesian framework, the maximum *a posteriori* (MAP) estimator seeks minimizing the sum of the negative log-likelihood of observed emission data conditional on radiotracer distribution and a regularizing penalty function, which penalizes solutions that have low probability. With various considerations such as computational efficiency of the algorithm to be developed and the spatial resolution of the reconstructed images, many different types of penalty functions were proposed. The total-variation based penalty function introduced in [28, 48] is particularly interesting in the field of ECT image reconstruction because it preserves the high spatial frequencies components of the reconstructed radiopharmacuetical distribution including discontinuities and steep gradients. Many efficient algorithms for this model were proposed, including EM-based methods [28, 48, 54], projected quasi-Newton methods [1, 2, 3], and forward-backward approaches [7, 13, 56]. In particular, a nested EM-TV iterative scheme was proposed recently in [54] for reconstruction of PET data with low signal-to-noise ratio, while an alternating extragradient method was proposed in [7] for solving the primal-dual formulation of the Poisson ECT data model with the TV regularization.

In the present paper, we study the numerical solution of the model that optimizes the sum of the negative log-likelihood of observed emission data conditional on radiotracer distribution (i.e. the Kullbach-Leibler divergence) and a total-variation regularization term. The ECT detector physics requires that the solution of this model be nonnegative. Difficulties with the numerical solution of such a model stem from the nonlinearity due to the use of the Kullbach-Leibler divergence and the TV regularization, from the nonnegativity constraint on the solution, and from the dense and large sized system matrix necessary for realistic ECT models. Motivated by our previous work [39, 42, 43], we characterize the solutions of the model in terms of a fixed-point of the proximity operators with a precondition matrix. Numerical algorithms are then developed based on such a characterization. More precisely, by identifying the total-variation as a composition of the convex function, which defines the 'ℓ^1-norm' or the 'ℓ^2-norm', and the first order difference operator, we formulate a characterization of the exact *positive* solutions to the optimization problem in terms of a system of fixed-point equations via the proximity operators of the convex function that defines the 'ℓ^1-norm' or the 'ℓ^2-norm' and the one that defines the first quadrant of an Euclidean space. The proximity operators of these simple functions have close forms, which provide great computational advantages. The nonlinearity is expressed in terms of a system of fixed-point equations, which naturally leads to an alternating projection algorithm. The dense system matrix of a large size is treated by preconditioning. Appropriate choices of the preconditioning matrices lead to efficient computational algorithms for solving the model.

This paper is divided into seven sections. In Section 2 we outline the Maximum *a Posteriori* ECT image reconstruction model in terms of a somewhat general constrained convex optimization problem. In Section 3, characterizations of the solution of the optimization problem are presented in terms of a system of fixed-point equations via proximity operators, and the iterative algorithms for finding the solution are developed based on the characterizations. Section 4 is devoted to convergence analysis of the algorithm. In Section 5, we specialize the general algorithm to the TV-regularized ECT image reconstruction problem. Numerical experiments are presented in Section 6 to test the approximation accuracy and computational efficiency of the proposed algorithm. The numerical results demonstrate that the alternating projection algorithm with the EM-preconditioner outperforms significantly the EM-TV in both the convergence speed and the image quality. We also observe that our algorithm performs favorably in comparison to the nested EM-TV in both the convergence speed and the image quality. We make concluding remarks in Section 7.

2 Maximum *a Posteriori* Estimation for ECT Reconstruction

In this section, we first present a mathematical model of a realistic ECT imaging system. We then find the unknown radioactive tracer distribution by maximizing the *posterior* probability distribution (the object function) using the observed emission data, known probability density function of the unknown radioactive tracer distribution, and the Bayes law. This approach is called the maximum *a posteriori* expectation-maximization (MAP-EM). Finally, the existence of a solution of the resulting variational problem is discussed.

We begin by introducing the notation to be used throughout this paper. Let $\mathbb{N}_k := \{1, 2, \ldots, k\}$. We use the same notation "1" to represent both the scalar number 1 and the vector with all components equal to one. They can be distinguished by the context of their use. For $x \in \mathbb{R}^k$, the expression $x \geq 0$ means that all components of x are no less than 0, and in this case, we say that x is a nonnegative vector. We denote by $\mathbb{R}^k_+$ the set $\{x : x \in \mathbb{R}^k \text{ and } x \geq 0\}$. For any vectors x and y in $\mathbb{R}^k$, we define

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

respectively, as the componentwise multiplication of x and y, and componentwise division of x by y. The logarithmic function at $x \in \mathbb{R}^k$ is defined as

$$\ln x := (\ln x_i : i \in \mathbb{N}_k),$$

while the expression "$x + \gamma$", the sum of the vector x with a scalar $\gamma \in \mathbb{R}$, is understood as the vector

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

We use $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively, for the inner product and the corresponding ℓ^2-norm in an Euclidean space, while we use $\| \cdot \|_1$ and $\| \cdot \|_\infty$, respectively, for the ℓ^1-norm and ℓ^∞-norm.

It is well accepted that gamma and x-ray photons, as well as positrons emitted in radioactive decay follow a Poisson distribution. Assuming that the detection of photons by detector units are events independent of one another and can be described by a Bernoulli process, consequently the projection images are a collection of Poisson random variables. The Poisson distribution approximates the photon detection process only if one can neglect the detector dead-timed i.e. count losses in the real detector systems, and no corrections are applied to the raw data. In addition to photons emitted from the patient, the detector bins (units) detect events due to ubiquitous radioactive terrestrial background and cosmic rays. It is assumed that they also follow the Poisson distribution.

Let f represent the expected radiotracer distribution within a patient or a phantom. Let g denote a vector with the i-th component being the number of single photons for SPECT or numbers of pairs of annihilation photons for PET originated from the radiotracer and recorded by the i-th detector unit or detector unit pair during the SPECT or PET scan, respectively. The dimension of the vector g is the number of detector units or detector unit pairs in the SPECT or PET imaging system, respectively. We assume that γ is a vector of the same size as g, with its i-th component being the mean number of "background" counts recorded by the i-th bin in SPECT or bin pair in PET. Under these assumptions, the observed emission data vector g related to the unknown radiotracer distribution f can be approximated by the following model [35, 55]

$$g = \text{Poisson}\left(Af + \gamma\right), \tag{1}$$

where Poisson(α) denotes a Poisson distributed random vector with mean α and A is the ECT system matrix with its (ij)-th element equal to the probability of detection of the photon emitted from voxel j of image f by the i-th detector bin in SPECT or detector bin pair in PET.

The maximum *a posteriori* probability (MAP) EM estimate has been proven useful in ECT for estimating the unobservable radiotracer distribution f when prior knowledge on probability distribution function of f is available, especially when the observed emission data g are noisy or incomplete [10, 19, 20, 21, 22, 30, 45]. Specifically, we assume that g in (1) is a given random vector in $\mathbb{R}^m$ and f is a random vector in $\mathbb{R}^d$. The MAP estimate $f_\star$ is obtained by maximizing the conditional *a posteriori* probability $p(f|g)$, the probability that f occurs when g is observed. This probability may be computed using the Bayes law

$$p(f|g) \propto p(g|f)p(f), \tag{2}$$

where $\alpha \propto \beta$ means that the scalar α is proportional to the scalar β. By taking the logarithm of the both sides of equation (2), the MAP estimate can then be calculated using the formula

$$f_\star = \arg\max \left\{ \ln p(g|f) + \ln p(f) : f \in \mathbb{R}^d \right\}. \tag{3}$$

In other words, the MAP estimate $f_\star$ is obtained by maximizing expectation, that is, by minimizing the sum of a negative log-likelihood of the observed emission data conditional on radiotracer distribution and a positive logarithm of the prior probability distribution. The first term can be considered as a fidelity term, a measure of the discrepancy between the estimated and the observed data. The second term is a regularization function, which penalizes solutions that have low probability. The Gibbs priors are commonly used in ECT reconstruction [18, 30, 36] in both convex [17, 20, 45, 47] and nonconvex [21, 31, 62] forms

$$p(f) \propto \exp(-\lambda U(f)) \tag{4}$$

with Gibbs real-valued energy function $U(f)$ defined on $\mathbb{R}^d$, and a positive regularization parameter λ called hyperparameter.

We formulate our maximum *a posteriori* ECT reconstruction model from (3) and (4) by computing the likelihood objective $p(g|f)$ using equation (1) and specifying the energy function U. According to equation (1), g follows the Poisson distribution having $Af + \gamma$ as its mean. As a result, the probability density function $p(g|f)$ of g conditioned on f can be computed by using the formula

$$p(g|f) = \prod_{i \in \mathbb{N}_m} \frac{((Af)_i + \gamma_i)^{g_i} \exp(-((Af)_i + \gamma_i))}{g_i!}. \tag{5}$$

We choose the energy function U in (4) as the total variation semi-norm. Accordingly, following the notation used in [42], we have that

$$U(f) := (\varphi \circ B)(f), \tag{6}$$

where B is an $n \times d$ first-order difference matrix and φ the ℓ^1-norm for the anisotropic total-variation or the ℓ^2-norm for the isotropic total-variation on $\mathbb{R}^n$. We shall provide more details of the total variation later in Section 5. Taking the logarithm of the both sides of equation (5) and incorporating identities

$$\sum_{i \in \mathbb{N}_m} g_i \ln((Af)_i + \gamma_i) = \langle \ln(Af + \gamma), g \rangle$$

and

$$\sum_{i \in \mathbb{N}_m} (Af)_i = \langle Af, 1 \rangle$$

in the resulting equation, we obtain that

$$\ln p(g|f) = \langle \ln(Af + \gamma), g \rangle - \langle Af, 1 \rangle + \text{const}_1,$$

where const_1 is a constant independent of f. Moreover, from (4) and (6) we have that

$$\ln p(f) = -\lambda(\varphi \circ B)(f) + \text{const}_2,$$

where const_2 is again a constant independent of f. Substituting the above two equations into (3) leads to the following variational problem

$$\min \left\{ \langle Af, 1 \rangle - \langle \ln(Af + \gamma), g \rangle + \lambda(\varphi \circ B)(f) : f \in \mathbb{R}^d_+ \right\}. \tag{7}$$

Although the variational problem (7) is derived in the specific context of the ECT image reconstruction, in this paper, we shall consider solving the problem in the following somewhat more general setting. We assume that A is a matrix in $\mathbb{R}^{m \times d}$, g is a given vector in $\mathbb{R}^m$, γ is a positive vector in $\mathbb{R}^m$, λ is a positive number, φ is a convex nonnegative function on $\mathbb{R}^n$, and B is an $n \times d$ matrix. The ECT image reconstruction model is a special case of model (7) when $\varphi \circ B$ is chosen as the total-variation semi-norm. For this reason, we call the above variational problem the Poisson-TV model. In the remaining part of this section and Sections 3 and 4, we consider model (7) in the general setting described above while in Sections 5 and 6 we specify it for the ECT image reconstruction model. In model (7), the requirement of $f \in \mathbb{R}^d_+$ is feasible due to the fact that f represents the mean radiotracer activity distribution in the object in ECT.

In the remaining part of this section, we establish the existence of the solutions to the variational problem (7). Recall that for a lower semicontinuous and convex function F defined over $\mathbb{R}^d$, a sufficient condition for F to have a minimizer over a closed convex set C is that the intersection of C and a lower level set of F at some height $\xi \in \mathbb{R}$ defined by

$$\mathrm{lev}_{\leq \xi} F := \left\{ f : f \in \mathbb{R}^d, F(f) \leq \xi \right\},$$

is nonempty and bounded (see, for example, [5]). We require some conditions on the system matrix A, motivated from the ECT imaging system. We denote by $\mathbb{A}$ the collection of $m \times d$ matrices, each of whose columns is a nonzero vector in $\mathbb{R}^m_+$. Since photons emitted from each voxel of an object are detected by detector bins in the ECT imaging system, it is reasonable to assume that the general system matrix A in (1) is in $\mathbb{A}$.

Proposition 2.1. *If $A \in \mathbb{A}$, $g \in \mathbb{R}^m_+$, γ is a positive vector in $\mathbb{R}^m$, λ is a positive number, φ is a convex nonnegative function on $\mathbb{R}^n$, and B is an $n \times d$ matrix, then the solution set of model* (7) *is nonempty.*

Proof. Let $F : \mathbb{R}^d_+ \to \mathbb{R}$ be defined at any $f \in \mathbb{R}^d_+$ as

$$F(f) := \langle Af, 1\rangle - \langle \ln(Af + \gamma), g\rangle + \lambda(\varphi \circ B)(f).$$

Clearly, F is the objective function of the variational problem (7). It suffices to prove that F is convex and coercive on the unbounded convex domain $\mathbb{R}^d_+$.

Note that φ is a convex function. The convexity of F on $\mathbb{R}^d_+$ is equivalent to the convexity of the function

$$F_1 := \langle A\cdot, 1\rangle - \langle \ln(A \cdot + \gamma), g\rangle$$

on $\mathbb{R}^d_+$. Since F_1 is twice continuously differentiable and the Hessian of F_1 at $f \in \mathbb{R}^d_+$

$$\nabla^2 F_1(f) := A^\top \mathrm{diag}\left(\frac{g}{(Af + \gamma)^2}\right) A,$$

is positive semi-definite due to $g \in \mathbb{R}^m_+$, the convexity of F follows.

Since φ is bounded below by 0, for any $\xi \in \mathbb{R}$ we have that $\mathrm{lev}_{\leq \xi} F \subset \mathrm{lev}_{\leq \xi} F_1$. Hence, by Theorem 11.9 in [5], the existence of the minimizer of the objective function F over R^d_+ follows from the boundedness of the lower level set $\mathrm{lev}_{\leq \xi} F_1$ for some $\xi \in \mathbb{R}$. Recalling the exercise 14 of Section 1.2 in [8] and the conditions imposed on A, g and γ, we know that the function F_1 has compact lower level sets, which in turn completes the proof. ☐

3 Characterizations and Algorithms

In this section, we first characterize solutions of the optimization problem (7) via the proximity operator. We then develop an alternating projection algorithm for solving the optimization problem based on the characterizations.

Let $\mathbb{H}$ denote an Euclidean space. For a proper convex function $\psi : \mathbb{H} \to \mathbb{R} \cup \{+\infty\}$, having a nonempty domain (the set on which ψ is finite), the proximity operator of ψ, denoted by prox_ψ, is a mapping from $\mathbb{H}$ to itself, defined for a given vector $x \in \mathbb{H}$ by

$$\mathrm{prox}_\psi(x) := \arg\min \left\{ \psi(u) + \frac{1}{2}\|u - x\|^2 : u \in \mathbb{H} \right\}. \tag{8}$$

The subdifferential of a proper function ψ on $\mathbb{H}$ at a given vector $x \in \mathbb{H}$ is the set defined by

$$\partial\psi(x) := \{y : \ y \in \mathbb{H} \quad \text{and} \quad \psi(z) \geq \psi(x) + \langle y, z - x\rangle \quad \text{for all} \quad z \in \mathbb{H}\}.$$

The subdifferential and the proximity operator of the function ψ are intimately related. Specifically, for x in the domain of ψ and $y \in \mathbb{H}$ we have that

$$y \in \partial\psi(x) \quad \text{if and only if} \quad x = \operatorname{prox}_{\psi}(x + y). \tag{9}$$

For a discussion of this relation, see, for example, [5, Proposition 16.34] and [42, 50].

The indicator function of a closed convex set C in $\mathbb{H}$ is defined as

$$\iota_C(u) := \begin{cases} 0, & \text{if } u \in C, \\ +\infty, & \text{otherwise.} \end{cases}$$

It can be observed that the proximity operator for the indicator function of a closed convex subset C in $\mathbb{H}$ is the projection operator onto C. For notational simplicity, when the set C is $\mathbb{R}_+^d$, we let

$$\Upsilon := \iota_{\mathbb{R}_+^d}.$$

The subdifferential of Υ can be explicitly given. To this end, for a given vector $x = (x_i : i \in \mathbb{N}_d)$ in $\mathbb{R}_+^d$, we define a projection matrix P associated with x by:

$$P(x) := \operatorname{diag}\left(\delta(x_i) : i \in \mathbb{N}_d\right),$$

where $\delta(x_i)$ equals to 1 if $x_i = 0$, and 0 otherwise. Then, by the definition of subdifferential, for a vector x in $\mathbb{R}_+^d$, we have that

$$\partial\Upsilon(x) = -P(x)\mathbb{R}_+^d. \tag{10}$$

Define

$$\mathbb{S} := \{S : S \text{ is a } d \times d \text{ diagonal matrix with positive diagonal entries}\}.$$

For any $S \in \mathbb{S}$ and $x \in \mathbb{R}_+^d$, from (10) together with the identity $S\mathbb{R}_+^d = \mathbb{R}_+^d$, we have that

$$S\partial\Upsilon(x) = -SP(x)\mathbb{R}_+^d = -P(x)S\mathbb{R}_+^d = -P(x)\mathbb{R}_+^d = \partial\Upsilon(x).$$

Thus, we observe that

$$S \circ \partial\Upsilon = \partial\Upsilon. \tag{11}$$

For given positive numbers λ and μ, a vector g in $\mathbb{R}_+^m$, an $m \times d$ matrix $A \in \mathbb{A}$, an $n \times d$ matrix B, we define $H : \mathbb{R}_+^d \times \mathbb{R}^n \to \mathbb{R}$, at $(f, b) \in \mathbb{R}_+^d \times \mathbb{R}^n$, as

$$H(f, b) := \langle Af, 1\rangle - \langle \ln(Af + \gamma), g\rangle + \lambda\mu\langle Bf, b\rangle. \tag{12}$$

We use $\nabla_f H$ and $\nabla_b H$ to denote the gradient of H with respect to its first and the second variables, respectively. More precisely, we have for any point $(f, b) \in \mathbb{R}_+^d \times \mathbb{R}^n$ that

$$\begin{aligned} \nabla_f H(f, b) &= A^\top \left(1 - \frac{g}{Af + \gamma}\right) + \lambda\mu B^\top b, & (13) \\ \nabla_b H(f, b) &= \lambda\mu B f. & (14) \end{aligned}$$

Set

$$\tau := \frac{\beta}{\lambda} \quad \text{and} \quad \sigma := \frac{1}{\lambda\mu}. \tag{15}$$

With this preparation, we present below a characterization of the solutions of model (7) via a coupled fixed-point equations.

Theorem 3.1. *Let* $A \in \mathbb{A}$, $g \in \mathbb{R}^m_+$, γ *and* λ *be two positive numbers,* φ *be a proper convex nonnegative function defined on* $\mathbb{R}^n$, *and* B *be an* $n \times d$ *matrix. If* $f \in \mathbb{R}^d_+$ *is a solution of the minimization problem* (7), *then for any* $\beta, \mu > 0$, *and* $S \in \mathbb{S}$, *there exists* $b \in \mathbb{R}^n$ *such that the pair* $(f, b) \in \mathbb{R}^d_+ \times \mathbb{R}^n$ *is a solution of the following coupled equations*

$$\begin{aligned} b &= \left(\mathcal{I} - \mathrm{prox}_{\mu^{-1}\varphi}\right)\left(b + \sigma \nabla_b H(f, b)\right), && (16) \\ f &= \mathrm{prox}_{\Upsilon}\left(f - \tau S \nabla_f H(f, b)\right). && (17) \end{aligned}$$

Conversely, if there exist $\beta, \mu > 0$, $S \in \mathbb{S}$, $b \in \mathbb{R}^n$ *and* $f \in \mathbb{R}^d_+$ *such that the above equations hold, then* f *is a solution of the minimization problem* (7).

Proof. Let f be a solution of the minimization problem (7). Applying Fermat's rule to (7), we get the relation

$$0 \in A^\top \left(1 - \frac{g}{Af + \gamma}\right) + \lambda B^\top \partial\varphi(Bf) + \partial\Upsilon(f). \tag{18}$$

Hence, for arbitrary positive numbers β and μ, there exist $u \in \frac{\beta}{\lambda}\partial\Upsilon(f)$ and $b \in \frac{1}{\mu}\partial\varphi(Bf)$ such that from (13) and (18), we have that

$$0 = \nabla_f H(f, b) + \frac{\lambda}{\beta} u. \tag{19}$$

By using relation (9), equation (16) is a direct consequence of $b \in \frac{1}{\mu}\partial\varphi(Bf)$ and (14). Multiplying the inclusion $u \in \frac{\beta}{\lambda}\partial\Upsilon(f)$ by the diagonal matrix S and using (11) yield the relation $Su \in \partial\Upsilon(f)$. Applying (9) to the above inclusion, we have that

$$f = \mathrm{prox}_{\Upsilon}(f + Su). \tag{20}$$

Solving u from (19) and substituting it into (20) yield (17).

Conversely, if β, μ are given positive numbers and S is a matrix in $\mathbb{S}$ such that (f, b) is a solution of (16)-(17), then all the arguments discussed above are reversible. This completes the proof. ☐

The characterization of the solutions to the minimization problem (7) in Theorem 3.1 is essential for deriving other equivalent ones with the aim of developing efficient algorithms for finding the solutions of the variational problem. As an example, an alternative formulation for the minimization problem (7) based on Theorem 3.1 is presented below.

Proposition 3.2. *Let* $A \in \mathbb{A}$, $g \in \mathbb{R}^m_+$, γ *and* λ *be two positive numbers,* φ *be a proper convex nonnegative function defined on* $\mathbb{R}^n$, *and* B *be an* $n \times d$ *matrix. If* $f \in \mathbb{R}^d_+$ *is a solution of the minimization problem* (7), *then for any* $\beta, \mu > 0$, *and* $S \in \mathbb{S}$, *there exists* $b \in \mathbb{R}^n$ *such that the pair* $(f, b) \in \mathbb{R}^d_+ \times \mathbb{R}^n$ *is a solution of the coupled equations*

$$\begin{aligned} b &= \left(\mathcal{I} - \mathrm{prox}_{\mu^{-1}\varphi}\right)\left(b + \sigma \nabla_b H\left(\mathrm{prox}_{\Upsilon}\left(f - \tau S \nabla_f H(f, b)\right), b\right)\right), && (21) \\ f &= \mathrm{prox}_{\Upsilon}\left(f - \tau S \nabla_f H(f, b)\right). && (22) \end{aligned}$$

Conversely, if there exist $\beta, \mu > 0$, $S \in \mathbb{S}$, $b \in \mathbb{R}^n$ *and* $f \in \mathbb{R}^d_+$ *such that the above equations hold, then* f *is a solution of the minimization problem* (7).

From the given assumptions, the existence of a fixed point $(f_\star, b_\star)$ of the coupled equations (21)-(22) is a direct consequence of Proposition 2.1 and Theorem 3.1 together with Proposition 3.2.

Based on the above characterization, the following Picard iteration is adopted to find a solution of the equations (21)-(22): Given any initial $(f^{(0)}, b^{(0)}) \in \mathbb{R}_+^d \times \mathbb{R}^n$, for any $k = 0, 1, \ldots$, we compute

$$\begin{cases} b^{(k+1)} = \left(\mathcal{I} - \operatorname{prox}_{\mu^{-1}\varphi}\right)\left(b^{(k)} + \sigma \nabla_b H\left(\operatorname{prox}_{\Upsilon}\left(f^{(k)} - \tau S \nabla_f H(f^{(k)}, b^{(k)})\right), b^{(k)}\right)\right) \\ f^{(k+1)} = \operatorname{prox}_{\Upsilon}\left(f^{(k)} - \tau S \nabla_f H(f^{(k)}, b^{(k+1)})\right). \end{cases} \tag{23}$$

We show next that both $\mathcal{I} - \operatorname{prox}_{\mu^{-1}\varphi}$ and $\operatorname{prox}_{\Upsilon}$ are projections. Since the iterative scheme (23) applies these two projections alternatingly, we shall call scheme (23) an alternating projection algorithm. It is of particularly interest that S in the characterization of Proposition 3.2 can be viewed as a preconditioner in our developed iterative scheme (23), therefore, it makes the algorithm practically tractable. For this reason, we call this matrix the preconditioning matrix.

To better prepare the convergence analysis of this algorithm which will be our focus of the next section, we present an equivalent form of (23). This new form is as follows:

$$\begin{cases} h^{(k)} := \operatorname{prox}_{\Upsilon}\left(f^{(k)} - \tau S \nabla_f H(f^{(k)}, b^{(k)})\right), \\ b^{(k+1)} = \left(\mathcal{I} - \operatorname{prox}_{\mu^{-1}\varphi}\right)\left(b^{(k)} + \sigma \nabla_b H\left(h^{(k)}, b^{(k)}\right)\right), \\ f^{(k+1)} = \operatorname{prox}_{\Upsilon}\left(f^{(k)} - \tau S \nabla_f H(f^{(k)}, b^{(k+1)})\right). \end{cases} \tag{24}$$

We now derive a technical lemma to show that the operator $\mathcal{I} - \operatorname{prox}_{\mu^{-1}\varphi}$ in (24) is a projection operator on a closed convex set of $\mathbb{R}^n$. To this end, we need to review the concept of the conjugate function and some related results. Let $\psi : \mathbb{H} \to \mathbb{R} \cup \{+\infty\}$ be a proper function. The function $\psi^* : \mathbb{H} \to \mathbb{R} \cup \{+\infty\}$ defined, at $u \in \mathbb{H}$, by

$$\psi^*(u) := \sup\{\langle x, u\rangle - \psi(x) : x \in \mathbb{H}\}$$

is called the conjugate of ψ at u. For a proper lower semicontinuous convex function ψ on $\mathbb{H}$ and any positive number α, the proximity operator of ψ and that of its conjugate ψ^* satisfy the following relation ([5, Theorem 14.3 (ii)] and [44, Proposition 4.a])

$$\mathcal{I} = \operatorname{prox}_{\alpha\psi} + \alpha \operatorname{prox}_{\alpha^{-1}\psi^*} \circ \alpha^{-1}\mathcal{I}. \tag{25}$$

Next, we recall the notion of the positive homogeneous function. A function $\psi : \mathbb{H} \to \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$ is positive homogeneous if for any $x \in \mathbb{H}$ and any positive number α,

$$\psi(\alpha x) = \alpha\psi(x). \tag{26}$$

Clearly, $\psi(0) = 0$. Furthermore, if ψ is a positive homogeneous and proper lower semicontinuous convex function on $\mathbb{H}$ then the conjugate function ψ^* is the indicator function on the set $\partial\psi(0)$ (see, e.g., [5, 52]), that is,

$$\psi^* = \iota_{\partial\psi(0)}. \tag{27}$$

Lemma 3.3. *If ψ is a positive homogeneous and proper lower semicontinuous convex function on $\mathbb{H}$ and α is a positive number, then $\mathcal{I} - \operatorname{prox}_{\alpha\psi}$ is the projection operator on the set $\alpha\partial\psi(0)$, that is,*

$$\mathcal{I} - \operatorname{prox}_{\alpha\psi} = \operatorname{prox}_{\iota_{\alpha\partial\psi(0)}}. \tag{28}$$

Proof. From (27), we know that $\alpha^{-1}\psi^* = \alpha^{-1}\iota_{\partial\psi(0)} = \iota_{\partial\psi(0)}$. Combining this relation with (25), we obtain that

$$\mathcal{I} - \operatorname{prox}_{\alpha\psi} = \alpha \operatorname{prox}_{\iota_{\partial\psi(0)}} \circ \alpha^{-1}\mathcal{I} = \operatorname{prox}_{\iota_{\alpha\partial\psi(0)}},$$

proving the desired formula. □

By Lemma 3.3, we conclude that $\mathcal{I} - \operatorname{prox}_{\mu^{-1}\varphi}$ in (24) is the same as $\operatorname{prox}_{\iota_{\mu^{-1}\partial\varphi(0)}}$. Hence, each equation in the iterative algorithm (24) involves an operator of the form $\operatorname{prox}_{\iota_C}(\cdot - T(\nabla G)(\cdot))$ for a convex set C in $\mathbb{H}$, a function G defined on $\mathbb{H}$ and a symmetric positive definite matrix T mapping $\mathbb{H}$ to itself.

We shall end this section by providing a more general characterization of the solutions of the minimization problem (7). To this end, for a proper convex nonnegative function φ on $\mathbb{R}^n$, an $n \times d$ matrix B, a given $h \in \mathbb{R}^d$, a positive number μ, and a positive integer r, we define $Q_h^r : \mathbb{R}^n \to \mathbb{R}^n$, with

$$Q_h^0 := \mathcal{I} \quad \text{and} \quad Q_h := \left(\mathcal{I} - \operatorname{prox}_{\mu^{-1}\varphi}\right)(\cdot + Bh),$$

recursively, by:

$$Q_h^r := Q_h \circ Q_h^{r-1}. \tag{29}$$

By using equation (14) and the proof of Proposition 3.2, we may establish the following result.

Proposition 3.4. *Let $A \in \mathbb{A}$, $g \in \mathbb{R}_+^m$, γ and λ be two positive numbers, φ be a proper convex nonnegative function defined on $\mathbb{R}^n$, and B be an $n \times d$ matrix. If $f \in \mathbb{R}_+^d$ is a solution of the minimization problem (7), then for any $\beta, \mu > 0$, and $S \in \mathbb{S}$, there exists $b \in \mathbb{R}^n$ such that the pair $(f, b) \in \mathbb{R}_+^d \times \mathbb{R}^n$ is a solution of the coupled equations*

$$b = Q_f^r(b) \tag{30}$$

$$f = \operatorname{prox}_{\Upsilon}\left(f - \tau S \nabla_f H(f, b)\right). \tag{31}$$

Conversely, if there exist $\beta, \mu > 0$, $S \in \mathbb{S}$, $b \in \mathbb{R}^n$ and $f \in \mathbb{R}_+^d$ such that the above equations hold, then f is a solution of the minimization problem (7).

Essentially, both Theorem 3.1 and Proposition 3.2 can be viewed as special cases of Proposition 3.4 with $r = 1$. The purpose of introducing the operator Q_f^r with r greater than 1 is mainly from consideration of developing efficient algorithms in Sections 5 and 6.

4 Convergence Analysis

We analyze in this section convergence of the preconditioned alternating projection algorithm described in the last section.

The convergence consideration of the algorithm requires introducing the notion of the weighted norm. For a symmetric and positive definite matrix $T : \mathbb{H} \to \mathbb{H}$, we define the weighted inner product by

$$\langle x, y \rangle_T := \langle T^{-1}x, y \rangle, \quad x, y \in \mathbb{H}.$$

The induced weighted norm is accordingly defined by

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

Note that a symmetric positive definite matrix has precisely one symmetric positive definite square root. Hence, we can rewrite $\langle x, x \rangle_T = \langle T^{-\frac{1}{2}}x, T^{-\frac{1}{2}}x \rangle$, which in turn implies that the above weighted norm is a norm on $\mathbb{H}$.

Next we present a lemma which provides a tool for the proof of the convergence of the sequence generated by the iterative scheme (23). For any initial pair $(f^{(0)}, b^{(0)}) \in \mathbb{R}_+^d \times \mathbb{R}^n$, we let

$$\mathcal{U} := \{(f^{(k)}, b^{(k)}) : k \in \mathbb{N}\}, \quad \text{where} \quad \mathbb{N} := \{1, 2, \ldots\}$$

be the sequence generated by the iterative scheme (23).

Lemma 4.1. *Let σ and τ be the two positive numbers defined by* (15)*, H be a function on $\mathbb{R}^d \times \mathbb{R}^n$ defined by* (12)*, $S \in \mathbb{S}$ and φ be a positive homogeneous and convex function on $\mathbb{R}^n$. If the following conditions hold*

(i) the sequence $\mathcal{U}$ is bounded,

(ii) $\lim_{k\to\infty} \|f^{(k+1)} - f^{(k)}\|_S = \lim_{k\to\infty} \|b^{(k+1)} - b^{(k)}\| = 0$,

then there exists a subsequence of $\mathcal{U}$ that converges to a solution of the coupled equations (21)-(22).

Proof. Since by hypothesis $\mathcal{U}$ is a bounded sequence in $\mathbb{R}^d_+ \times \mathbb{R}^n$, there exists a subsequence $\{(f^{(k_i)}, b^{(k_i)}) : i \in \mathbb{N}\}$ that converges to a point $(\widehat{f}, \widehat{b})$ in $\mathbb{R}^d_+ \times \mathbb{R}^n$. This together with Condition (ii) ensures that the subsequence $\{(f^{(k_i+1)}, b^{(k_i+1)}) : i \in \mathbb{N}\}$ converges to the same point. Therefore, in (23) by choosing $k := k_i$ and letting $i \to \infty$, we conclude that the point $(\widehat{f}, \widehat{b})$ satisfies the coupled equations (21)-(22). ☐

The use of Lemma 4.1 in the proof of the convergence result requires that we verify the hypotheses of the lemma. This is fulfilled by establishing the following estimate on the quantities defined by

$$e^{(k)} := \frac{1}{2\tau}\|f^{(k)} - f_\star\|_S^2 + \frac{1}{2\sigma}\|b^{(k)} - b_\star\|^2$$

for a solution $(f_\star, b_\star)$ of equations (21)-(22) that

$$e^{(k)} + \frac{1}{2\tau}\sum_{j=0}^{k-1}\|f^{(j+1)} - h^{(j)}\|_S^2 + \frac{1}{2\tau}\sum_{j=0}^{k-1} C_j\|h^{(j)} - f^{(j)}\|_S^2 + \frac{1}{2\sigma}\sum_{j=0}^{k-1} D_j\|b^{(j+1)} - b^{(j)}\|^2 \le e^{(0)}, \quad (32)$$

where $h^{(j)}$ is defined as in (24), and C_j, D_j are some positive constants.

We now establish several results needed for proving estimate (32). For a proper convex function $\psi : \mathbb{H} \to \mathbb{R} \cup \{+\infty\}$ and a symmetric positive definite matrix T, the proximity operator of ψ with respect to T, denoted by prox_ψ^T, is defined for a given vector $x \in \mathbb{H}$ by

$$\mathrm{prox}_\psi^T(x) := \arg\min\left\{\psi(u) + \frac{1}{2}\|u - x\|_T^2 : u \in \mathbb{H}\right\}.$$

Clearly, we have that $\mathrm{prox}_\psi = \mathrm{prox}_\psi^I$. Moreover, for x in the domain of ψ and $y \in \mathbb{H}$, we have the following generalization of (9)

$$y \in T\partial\psi(x) \quad \text{if and only if} \quad x = \mathrm{prox}_\psi^T(x + y). \quad (33)$$

We next present a reformulation of the proximity operator prox_Υ.

Proposition 4.2. *If $S \in \mathbb{S}$, then*

$$\mathrm{prox}_\Upsilon = \mathrm{prox}_\Upsilon^S. \quad (34)$$

Proof. Let $x \in \mathbb{R}^d$ and $p = \mathrm{prox}_\Upsilon(x)$. By relation (9), we have the inclusion $x - p \in \partial\Upsilon(p)$. Since $p \in \mathbb{R}^d_+$, using equation (11), we get that $x - p \in S\partial\Upsilon(p)$. By relation (33), it yields that $p = \mathrm{prox}_\Upsilon^S(x)$, which completes the proof. ☐

With Proposition 4.2 in mind, we next discuss several properties of the operator $\mathrm{prox}_{\iota_C}^T(\cdot - T(\nabla G)(\cdot))$, which are crucial in our convergence analysis.

Lemma 4.3. *Let C be a nonempty closed convex subset of $\mathbb{H}$, T be a symmetric positive definite matrix mapping $\mathbb{H}$ to itself and G be a proper convex differentiable function on $\mathbb{H}$. Set*

$$p := \operatorname{prox}_{\iota_C}^{T}(x - T\nabla G(x)). \tag{35}$$

Then the following statements hold:

(i) For any $y \in C$,

$$\|p - y\|_T^2 \leq \|x - y\|_T^2 - \|p - x\|_T^2 - 2\langle \nabla G(x), p - y\rangle \tag{36}$$

and

$$\|p - y\|_T^2 \leq \|x - y\|_T^2 - \|p - x\|_T^2 - 2\langle \nabla G(x), p - x\rangle + 2\left[G(y) - G(x)\right]. \tag{37}$$

(ii) Furthermore, if G is also differentiable at p, then for any $y \in C$,

$$\|p - y\|_T^2 \leq \|x - y\|_T^2 - \|p - x\|_T^2 - 2\langle \nabla G(x) - \nabla G(p), p - x\rangle + 2\left[G(y) - G(p)\right]. \tag{38}$$

Proof. We first prove the inequalities in Item (i). Note that for any $y \in \mathbb{H}$, we have that

$$\|p - y\|_T^2 = \|x - y\|_T^2 - \|p - x\|_T^2 + 2\langle T^{-1}(p - x), p - y\rangle. \tag{39}$$

By equations (35) and (9), we have the inclusion relation

$$x - T\nabla G(x) - p \in T\partial\iota_C(p).$$

Multiplying the above inclusion by the symmetric positive definite matrix T^{-1} and recalling the definition of the subdifferential, we have for all $y \in C$ that

$$\langle T^{-1}(x - T\nabla G(x) - p), y - p\rangle \leq \iota_C(y) - \iota_C(p) = 0.$$

By splitting the term $\langle T^{-1}(p-x), p-y\rangle$ as the sum of $\langle T^{-1}(x-T\nabla G(x)-p), y-p\rangle$ and $\langle -\nabla G(x), p-y\rangle$, and using the above inequality, from (39) we conclude that inequality (36) holds. Inequality (37) follows from (36) together with

$$\langle \nabla G(x), p - y\rangle = \langle \nabla G(x), p - x\rangle + \langle \nabla G(x), x - y\rangle$$

and the inequality

$$\langle \nabla G(x), x - y\rangle \geq G(x) - G(y)$$

ensured by the convexity of G.

Finally, since G is differentiable at p, again by the convexity of G, we have that

$$\langle \nabla G(p), p - x\rangle \geq G(p) - G(x).$$

This together with (37) leads to (38). □

As a consequence of Lemma 4.3 with G being identical to zero, using (36) twice we derive for all $x, y \in \mathbb{H}$ that

$$\|\operatorname{prox}_{\iota_C}^{T}(x) - \operatorname{prox}_{\iota_C}^{T}(y)\|_T \leq \|x - y\|_T. \tag{40}$$

We next show that a solution of the coupled equations (16)-(17) is a saddle point of the function H. This result is necessary for establishing inequality (32).

Lemma 4.4. *Let H be a function on $\mathbb{R}^d \times \mathbb{R}^n$ defined by* (12)*. If $(f_\star, b_\star) \in \mathbb{R}_+^d \times \mathbb{R}^n$ is a solution of the coupled equations* (16)-(17)*, then inequality*

$$H(f_\star, b) \leq H(f_\star, b_\star) \leq H(f, b_\star) \tag{41}$$

holds for any point $(f, b) \in \mathbb{R}_+^d \times \left(\mu^{-1}\partial\varphi(0)\right)$.

Proof. We prove this lemma by showing the following two inequalities

$$H(f_\star, b) \leq H(f_\star, b_\star), \quad \text{for all } \ b \in \left(\mu^{-1}\partial\varphi(0)\right) \tag{42}$$

and

$$H(f_\star, b_\star) \leq H(f, b_\star), \quad \text{for all } \ f \in \mathbb{R}_+^d. \tag{43}$$

We first prove inequality (42). By the definition of $b_\star$ and (28), we have that

$$b_\star = \operatorname{prox}_{\iota_{\mu^{-1}\partial\varphi(0)}}(b_\star + Bf_\star).$$

Employing the characterization of the projection, we observe that

$$\langle Bf_\star, b - b_\star\rangle = \langle (b_\star + Bf_\star) - b_\star, b - b_\star\rangle \leq 0, \quad \text{for all } \ b \in \left(\mu^{-1}\partial\varphi(0)\right),$$

which is equivalent to inequality (42).

It remains to show inequality (43). Using the relations (9) and (34), from (17) we get for any $S \in \mathbb{S}$ that

$$0 \in S\nabla_f H(f_\star, b_\star) + S\partial\iota_{\mathbb{R}_+^d}(f_\star).$$

Multiplying the above inclusion by the matrix S^{-1} guarantees that $H(\cdot, b_\star)$ achieves its minimum value at the point $f_\star$. Thus, inequality (43) is valid. $\square$

We next prove estimate (32) by employing Lemma 4.3 and Lemma 4.4. To this end, for an $S \in \mathbb{S}$, we define the quantities

$$A_j := \begin{cases} \frac{\|\nabla_f H(h^{(j)}, b^{(j+1)}) - \nabla_f H(f^{(j)}, b^{(j+1)})\|_{S^{-1}}}{\|h^{(j)} - f^{(j)}\|_S}, & \text{if } \|h^{(j)} - f^{(j)}\|_S \neq 0, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$B_j := \begin{cases} \frac{\|\nabla_f H(f^{(j)}, b^{(j+1)}) - \nabla_f H(f^{(j)}, b^{(j)})\|}{\|b^{(j+1)} - b^{(j)}\|}, & \text{if } \|b^{(j+1)} - b^{(j)}\| \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Lemma 4.5. *If $(f_\star, b_\star)$ is a solution of the coupled equations* (21)-(22)*, then the estimate (32) holds with $C_j := 1 - 2\tau A_j$ and $D_j := 1 - 2\tau\|B\|_2\|S\|_2 B_j$.*

Proof. Let j be a positive integer. Identifying x, y, p, T and G in (38), respectively, with $b^{(j)}$, $b_\star$, $b^{(j+1)}$, I and $-\sigma H(h^{(j)}, \cdot)$, and recalling $b^{(j+1)}$ being defined by the projection (24), we have that

$$\frac{1}{2\sigma}\|b^{(j+1)} - b_\star\|^2 \leq \frac{1}{2\sigma}\|b^{(j)} - b_\star\|^2 - \frac{1}{2\sigma}\|b^{(j+1)} - b^{(j)}\|^2 + H(h^{(j)}, b^{(j+1)}) - H(h^{(j)}, b_\star), \tag{44}$$

where we have used the fact that $\nabla_b H(f, \cdot)$ is a constant. Likewise, identifying x, y, p, T and G in (37), respectively, with $f^{(j)}$, $f_\star$, $f^{(j+1)}$, S and $\tau H(\cdot, b^{(j+1)})$, recalling $f^{(j+1)}$ being defined by the projection (23) and applying Proposition 4.2, we observe that

$$\begin{aligned} \frac{1}{2\tau}\|f^{(j+1)} - f_\star\|_S^2 \ \leq \ & \frac{1}{2\tau}\|f^{(j)} - f_\star\|_S^2 - \frac{1}{2\tau}\|f^{(j+1)} - f^{(j)}\|_S^2 + \langle \nabla_f H(f^{(j)}, b^{(j+1)}), f^{(j)} - f^{(j+1)}\rangle \\ & + H(f_\star, b^{(j+1)}) - H(f^{(j)}, b^{(j+1)}). \end{aligned} \tag{45}$$

Using the inequality

$$H(f_\star, b^{(j+1)}) \leq H(h^{(j)}, b_\star),$$

which is ensured by Lemma 4.4, in the sum of the last two terms of (44) and (45) yields

$$H(h^{(j)}, b^{(j+1)}) - H(f^{(j)}, b^{(j+1)}) + H(f_\star, b^{(j+1)}) - H(h^{(j)}, b_\star) \leq H(h^{(j)}, b^{(j+1)}) - H(f^{(j)}, b^{(j+1)}). \quad (46)$$

The convexity of $H(\cdot, b^{(j+1)})$ ensures that

$$\langle \nabla_f H(h^{(j)}, b^{(j+1)}), f^{(j)} - h^{(j)} \rangle \leq H(f^{(j)}, b^{(j+1)}) - H(h^{(j)}, b^{(j+1)}),$$

which combined with (46) gives

$$H(h^{(j)}, b^{(j+1)}) - H(f^{(j)}, b^{(j+1)}) + H(f_\star, b^{(j+1)}) - H(h^{(j)}, b_\star) \leq \langle \nabla_f H(h^{(j)}, b^{(j+1)}), h^{(j)} - f^{(j)} \rangle. \quad (47)$$

Moreover, identifying x, y, p, T and G in (36), respectively, with $f^{(j)}$, $f^{(j+1)}$, $h^{(j)}$, S and $\tau H(\cdot, b^{(j)})$, and recalling the definition (24) of $h^{(j)}$ in terms of the projection, we obtain that

$$\frac{1}{2\tau}\|f^{(j+1)} - f^{(j)}\|_S^2 \geq \frac{1}{2\tau}\|f^{(j+1)} - h^{(j)}\|_S^2 + \frac{1}{2\tau}\|h^{(j)} - f^{(j)}\|_S^2 + \langle \nabla_f H(f^{(j)}, b^{(j)}), h^{(j)} - f^{(j+1)} \rangle. \quad (48)$$

Recalling the definition of $e^{(j)}$, summing (44) and (45), and using (48), (47) in the resulting sum yield

$$e^{(j+1)} - e^{(j)} \leq -\frac{1}{2\tau}\|f^{(j+1)} - h^{(j)}\|_S^2 - \frac{1}{2\tau}\|h^{(j)} - f^{(j)}\|_S^2 - \frac{1}{2\sigma}\|b^{(j+1)} - b^{(j)}\|^2 + I_1 + I_2, \quad (49)$$

where

$$\begin{aligned} I_1 &:= \langle \nabla_f H(f^{(j)}, b^{(j+1)}) - \nabla_f H(h^{(j)}, b^{(j+1)}), f^{(j)} - h^{(j)} \rangle, \\ I_2 &:= \langle \nabla_f H(f^{(j)}, b^{(j)}) - \nabla_f H(f^{(j)}, b^{(j+1)}), f^{(j+1)} - h^{(j)} \rangle. \end{aligned}$$

Next we further estimate the last two inner products in (49). Note that S and S^{-1} are diagonal and positive definite matrices, and hence they have unique diagonal positive definite square roots $S^{\frac{1}{2}}$ and $S^{-\frac{1}{2}}$, respectively. We then have that

$$\begin{aligned} I_1 &= \left\langle S^{\frac{1}{2}}\left(\nabla_f H(f^{(j)}, b^{(j+1)}) - \nabla_f H(h^{(j)}, b^{(j+1)})\right), S^{-\frac{1}{2}}\left(f^{(j)} - h^{(j)}\right)\right\rangle \\ &\leq \|\nabla_f H(f^{(j)}, b^{(j+1)}) - \nabla_f H(h^{(j)}, b^{(j+1)})\|_{S^{-1}} \cdot \|f^{(j)} - h^{(j)}\|_S. \end{aligned} \quad (50)$$

The last inequality follows from using the Cauchy-Schwartz inequality together with the definition of the weighted norm. Likewise, we can also get that

$$I_2 \leq \|\nabla_f H(f^{(j)}, b^{(j)}) - \nabla_f H(f^{(j)}, b^{(j+1)})\|_{S^{-1}} \cdot \|f^{(j+1)} - h^{(j)}\|_S. \quad (51)$$

From the definitions of $f^{(j+1)}$ in (23) and $h^{(j)}$ in (24) and inequality (40), we get

$$\begin{aligned} \|f^{(j+1)} - h^{(j)}\|_S &\leq \tau\left\|S\left(\nabla_f H(f^{(j)}, b^{(j+1)}) - \nabla_f H(f^{(j)}, b^{(j)})\right)\right\|_S \\ &= \tau\left\|S^{\frac{1}{2}}\left(\nabla_f H(f^{(j)}, b^{(j+1)}) - \nabla_f H(f^{(j)}, b^{(j)})\right)\right\|. \end{aligned} \quad (52)$$

On the other hand, since

$$\nabla_f H(f^{(j)}, b^{(j)}) - \nabla_f H(f^{(j)}, b^{(j+1)}) = \lambda\mu B^\top(b^{(j)} - b^{(j+1)}),$$

this equation with (52) yields that

$$\|f^{(j+1)} - h^{(j)}\|_S \leq \tau\lambda\mu\|B\|_2\|S^{\frac{1}{2}}\|_2\|b^{(j)} - b^{(j+1)}\|,$$

which together with the estimate (51) implies that

$$I_2 \leq \lambda\mu\tau B_j\|B\|_2\|S\|_2\|b^{(j+1)} - b^{(j)}\|^2. \tag{53}$$

Combining inequalities (49), (50) and (53) together with using the definition of A_j gives the estimate

$$e^{(j+1)} - e^{(j)} \leq -\frac{1}{2\tau}\|f^{(j+1)} - h^{(j)}\|_S^2 - \frac{C_j}{2\tau}\|h^{(j)} - f^{(j)}\|_S^2 - \frac{D_j}{2\sigma}\|b^{(j+1)} - b^{(j)}\|^2. \tag{54}$$

Summing the above inequality (54) for j running from 0 to k yields estimate (32). $\square$

In order to use the last lemma to show the validity of the hypotheses of Lemma 4.1, we bound the constants A_j and B_j that appear in the last lemma. To this end, we need the following lemma that pertains to the Lipschitz continuity of the gradient of H with respect to each of its variables.

Lemma 4.6. *If H is a function on $\mathbb{R}^d \times \mathbb{R}^n$ defined by* (12) *and $S \in \mathbb{S}$, then for a fixed vector b in $\mathbb{R}^n$, $\nabla_f H(\cdot, b)$ is Lipschitz continuous with constant $\frac{\|g\|_\infty\|A\|_2^2\|S\|_2}{\gamma^2}$ while for a fixed vector f in $\mathbb{R}^d$, $\nabla_f H(f, \cdot)$ is Lipschitz continuous with constant $\lambda\mu\|B\|_2$. That is, for any $f_1, f_2 \in \mathbb{R}_+^d$ and $b \in \mathbb{R}^n$*

$$\|\nabla_f H(f_1, b) - \nabla_f H(f_2, b)\|_{S^{-1}} \leq \frac{\|g\|_\infty\|A\|_2^2\|S\|_2}{\gamma^2}\|f_1 - f_2\|_S \tag{55}$$

and for any $b_1, b_2 \in \mathbb{R}^n$ and $f \in \mathbb{R}_+^d$

$$\|\nabla_f H(f, b_1) - \nabla_f H(f, b_2)\| \leq \lambda\mu\|B\|_2\|b_1 - b_2\|. \tag{56}$$

Proof. Using (13) and the definition of the weighted norm, for any $f_1, f_2 \in \mathbb{R}_+^d$ and $b \in \mathbb{R}^n$, we have that

$$\|\nabla_f H(f_1, b) - \nabla_f H(f_2, b)\|_{S^{-1}} = \left\| S^{\frac{1}{2}} A^\top \left(\frac{(AS^{\frac{1}{2}}S^{-\frac{1}{2}}(f_2 - f_1)) \odot g}{(Af_1 + \gamma) \odot (Af_2 + \gamma)} \right) \right\|.$$

It can be verified that

$$\|(Af_1 + \gamma) \odot (Af_2 + \gamma)\|_\infty \geq \gamma^2 \quad \text{and} \quad \|(AS^{\frac{1}{2}}S^{-\frac{1}{2}}(f_2 - f_1)) \odot g\| \leq \|A\|_2\|S^{\frac{1}{2}}\|_2\|g\|_\infty\|f_1 - f_2\|_S,$$

from which inequality (55) follows.

Inequality (56) follows directly from the expression of $\nabla_f H$ given in (13). $\square$

Finally, we show the convergence of the sequence $\mathcal{U} := \{(f^{(k)}, b^{(k)}) : k \in \mathbb{N}\}$ generated by the preconditioned alternating projection algorithm (23).

Theorem 4.7. *Let λ, μ, γ be positive numbers, σ and τ be the numbers defined by* (15)*, $S \in \mathbb{S}$, φ be a positive homogeneous convex function on $\mathbb{R}^n$, and H be the function on $\mathbb{R}^d \times \mathbb{R}^n$ defined by* (12)*. If β and μ are chosen to satisfy the conditions*

$$0 < \beta \leq \frac{(1-\varepsilon)\lambda\gamma^2}{2\|g\|_\infty\|A\|_2^2\|S\|_2}, \quad 0 < \beta\mu \leq \frac{1-\varepsilon}{2\|B\|_2^2\|S\|_2}$$

for some $\varepsilon \in (0, 1)$, then for any initial pair $(f^{(0)}, b^{(0)}) \in \mathbb{R}_+^d \times \mathbb{R}^n$, the sequence $\mathcal{U}$ converges to a solution of the coupled equations (21)*-*(22)*.*

Proof. We prove this theorem by employing Lemma 4.1. The conditions imposed on β and μ together with Lemma 4.6 lead to that both $1-2\tau A_j$ and $1-2\tau\|B\|_2\|S\|_2 B_j$ are bounded below by ϵ for all j. Hence, Lemma 4.5 implies that the sequence $\{(f^{(j)}, b^{(j)}) : j \in \mathbb{N}\}$ is bounded and sequence $\{\|b^{(j+1)} - b^{(j)}\| : j \in \mathbb{N}\}$ converges to 0 and moreover both the sequences $\{\|h^{(j)} - f^{(j)}\|_S : j \in \mathbb{N}\}$, $\{\|f^{(j+1)} - h^{(j)}\|_S : j \in \mathbb{N}\}$ converge to 0, which implies the convergence of the sequence $\{\|f^{(k+1)} - f^{(k)}\|_S : k \in \mathbb{N}\}$ to 0. Therefore, by Lemma 4.1, there exists a subsequence $\{(f^{(k_i)}, b^{(k_i)}) : i \in \mathbb{N}\}$ which converges to $(\widehat{f}, \widehat{b}) \in \mathbb{R}^d_+ \times (\mu^{-1}\partial\varphi(0))$, a solution of the equations (21)-(22). It remains to show that the sequence $\mathcal{U}$ also converges to the same point $(\widehat{f}, \widehat{b})$. Indeed, inequality (54) ensures that

$$\frac{1}{2\tau}\|f^{(j+1)} - f_\star\|_S^2 + \frac{1}{2\sigma}\|b^{(j+1)} - b_\star\|^2 \le \frac{1}{2\tau}\|f^{(j)} - f_\star\|_S^2 + \frac{1}{2\sigma}\|b^{(j)} - b_\star\|^2. \tag{57}$$

Replacing the point $(f_\star, b_\star)$ in inequality (57) by the point $(\widehat{f}, \widehat{b})$ and summing the resulting inequality for j from k_i to $k-1$ with $k > k_i$ lead to the inequality

$$\frac{1}{2\tau}\|f^{(k)} - \widehat{f}\|_S^2 + \frac{1}{2\sigma}\|b^{(k)} - \widehat{b}\|^2 \le \frac{1}{2\tau}\|f^{(k_i)} - \widehat{f}\|_S^2 + \frac{1}{2\sigma}\|b^{(k_i)} - \widehat{b}\|^2.$$

This clearly guarantees that the sequence $\mathcal{U}$ converges to the point $(\widehat{f}, \widehat{b})$. □

We remark that the conditions stated in the last theorem imposed on the parameter β and μ are rather restricted. We shall demonstrate in Section 6 by numerical examples that choices of significantly larger β and μ than those allowed by the theorem lead to convergence of the iteration.

5 TV-Regularized MAP ECT Reconstruction

In this section, we specialize the general preconditioned alternating projection algorithm developed in Section 3 to the MAP ECT reconstruction by specifying the function φ, the matrices B and the preconditioning matrix S in (23). In particular, we present explicit formulas of the proximity operators for the two special convex functions involved in the algorithm. We also discuss the importance of the inner iteration for efficient computation in the context of MAP ECT reconstruction. Finally, we compare differences and advantages of our proposed algorithm with those of several existing algorithms for the Poisson-TV model (7).

We first present explicit expressions of φ and B according to the definition of the total-variation [53]. The concrete expressions of φ and B depend on how three dimensional images are vectorized. A three-dimensional image is assembled by a stack of two-dimensional images. For convenience of exposition, we assume that an image considered in this paper has a size of $p \times p \times q$. The image is treated as a vector in $\mathbb{R}^{p^2 q}$ in such a way that the ijk-th voxel of the image, where $i, j \in \mathbb{N}_p$ and $k \in \mathbb{N}_q$, corresponds to the $(i + (j-1)p + (k-1)p^2)$-th element of the vector in $\mathbb{R}^{p^2 q}$. In the current section, we set $d := p^2 q$. To define the matrix B, we define an $\alpha \times \alpha$ difference matrix D_α by

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

In terms of the notion of the matrix Kronecker product $\otimes$, we define the $3d \times d$ matrix B by

$$B := \begin{bmatrix} I_q \otimes I_p \otimes D_p \\ I_q \otimes D_p \otimes I_p \\ D_q \otimes I_p \otimes I_p \end{bmatrix}.$$

The convex function $\varphi : \mathbb{R}^{3d} \to \mathbb{R}$ is defined at $z \in \mathbb{R}^{3d}$ as

$$\varphi(z) := \sum_{i=1}^{d} \left\| [z_i, z_{d+i}, z_{2d+i}]^\top \right\|. \tag{58}$$

The isotropic total-variation of a vector f is then expressed as $\varphi(Bf)$.

Execution of the iterative scheme (23) requires the availability of explicit formulas for the proximity operators of functions φ defined in (58) and the indicator function $\Upsilon := \iota_{\mathbb{R}^d_+}$. For a positive number μ and a vector $z \in \mathbb{R}^{3d}$, the components of the vector

$$y := \operatorname{prox}_{\mu^{-1}\varphi}(z)$$

can be computed by using the formula

$$[y_i, y_{d+i}, y_{2d+i}]^\top = \max\left\{ \|[z_i, z_{d+i}, z_{2d+i}]^\top\| - \frac{1}{\mu}, 0 \right\} \frac{[z_i, z_{d+i}, z_{2d+i}]^\top}{\|[z_i, z_{d+i}, z_{2d+i}]^\top\|}, \quad i \in \mathbb{N}_d.$$

With this formula, for any positive integer r and a vector $h \in \mathbb{R}^d$, the operator Q_h^r defined by (29) can be explicitly computed. The proximity operator of the indicator function Υ (the projection operator onto the first octant $\mathbb{R}^d_+$) also has an explicit expression. That is, for $x \in \mathbb{R}^d_+$,

$$(\operatorname{prox}_\Upsilon(x))_i = \max\{x_i, 0\}, \quad i \in \mathbb{N}_d.$$

Thus, both $\operatorname{prox}_{\mu^{-1}\varphi}$ and $\operatorname{prox}_\Upsilon$ have close forms in the current context. These close forms are convenient for numerical evaluation of the proximity operators of the two specific functions in the sense that no further optimization problems are required to solve. These specific examples can be evaluated numerically within the machine precision. While in general, computing the proximity operator of a convex function requires solving an optimization problem by its definition. Even in the situation when close forms of the proximity operators are available, some round-off errors may be introduced during computation. In such a case, one needs to consider the stability issue of computing the proximity operator.

Based on Proposition 3.4, we propose the following algorithm (Algorithm 1) for the MAP ECT reconstruction.

Algorithm 1 (Alternating Projection Algorithm for MAP ECT Reconstruction with a Fixed Matrix S)

```
Preparation: $\nabla_f H$, $\nabla_b H$, $\tau$ are defined in (13), (14) and (15), respectively. The parameter $r$ is a positive integer.
Initialization: $f^{(0)} = 1$, $b^{(0)} = 0$.
repeat
    Step 1: $h^{(k)} \leftarrow \operatorname{prox}_\Upsilon(f^{(k)} - \tau S \nabla_f H(f^{(k)}, b^{(k)}))$
    Step 2: $b^{(k+1)} \leftarrow Q^r_{h^{(k)}}(b^{(k)})$
    Step 3: $f^{(k+1)} \leftarrow \operatorname{prox}_\Upsilon\left(f^{(k)} - \tau S \nabla_f H(f^{(k)}, b^{(k+1)})\right)$
until "convergence"
```

The parameter r in Algorithm 1 is the iteration number for the inner iteration. When r is chosen to be 1, three steps in Algorithm 1 correspond to the three equations in (24), respectively. In Algorithm 1, both Steps 1 and 3 involve the matrices A and $A^\top$ while Step 2 involves matrix B. In the context of the MAP ECT reconstruction, matrix A is a dense matrix of a large size and matrix B is a sparse matrix of a relatively small size. As a result, computation with the matrix A or $A^\top$ is more costly than computation with the matrix B. To reduce the overall computational cost, we suggest that the inner iteration be carried out with an appropriate choice of iteration number r. By paying less computational effort with an appropriate r we could obtain a more accurate estimate $f^{(k+1)}$ at the k-th outer iterate. We shall study in the next section by numerical examples the choices of the iteration number r.

The preconditioning matrix S is not specified in Algorithm 1. The choice of the preconditioning matrix is crucial in designing computationally efficient algorithms. One may choose the preconditioning matrix S as the identity matrix, which corresponds to the trivial case without preconditioning. More interesting cases are the nontrivial choices. The choice of the preconditioning matrix S may be motivated from different ways with the same purpose of speeding up the convergence of the algorithm. A possible nontrivial choice is motivated by the idea of the projected Newton method. For more details of the projected Newton method, see [3, 6]. In passing, we point it out that preconditioning techniques were used in the context of emission computed tomography for other algorithms (see, for example, [15, 32, 34, 45]).

We propose a choice of the preconditioner S based on the classical expectation-maximization (EM). Recall that EM is an iterative scheme for computing the maximum likelihood estimate. According to [35], when the EM algorithm is applied to (1), we have for any $f^{(0)} \in \mathbb{R}^d_+$ that

$$f^{(k+1)} = E^{(k)} A^\top \left(\frac{g}{A f^{(k)} + \gamma} \right), \tag{59}$$

where $E^{(k)}$ is a diagonal matrix defined by

$$E^{(k)} := \operatorname{diag}\left(\frac{f^{(k)}}{A^\top 1} \right). \tag{60}$$

In the EM algorithm, the matrix $E^{(k)}$ determines the direction for the next step of a search for the minimizer, for the purpose of finding the maximum likelihood estimate. By comparing (59) with the form of $\nabla_f H$ given in (13), motivated by the matrix $E^{(k)}$ having the form (60) we suggest that we choose the matrix S in Algorithm 1 as the diagonal matrix $E^{(k)}$ at the k-th iteration. This choice of the preconditioning matrix allows the search of the minimizer to follow the direction of the search in the classical EM algorithm for finding the maximum likelihood estimate while preserving the advantage of the alternating projection nature in the proposed algorithm. In this way, the preconditioning matrix is updated at every iterate step when a new value $f^{(k)}$ is available. This leads to the following algorithm (Algorithm 2) for the MAP ECT reconstruction.

Since the choice of S is motivated from the EM algorithm (59), we shall call Algorithm 2 the EM preconditioned alternating projection algorithm (PAPA) for MAP ECT reconstruction and call the matrix E the EM-preconditioner. A numerical comparison of the proposed Algorithm 1 and Algorithm 2 will be presented in the next section. The numerical study shows that the EM-preconditioner speeds up significantly the convergence of the alternating projection algorithm.

We further comment on the dynamics of the EM-preconditioner. The PAPA algorithm which we described in Section 3 and for which we proved convergence in Section 4 has a *fixed* preconditioning matrix S. The preconditioner in Algorithm 2 changes dynamically from step to step. In the next section, we shall study the dynamics of the EM-preconditioner numerically and shall see that after some iteration steps, the change in the EM-preconditioner is so small that it can be neglected. In other wards, the EM-preconditioner tends to a fixed preconditioning matrix as the iteration number increases. Therefore, in practise, we may *fix* the preconditioning matrix after some iteration steps

Algorithm 2 (Preconditioned Alternating Projection Algorithm for MAP ECT Reconstruction)

```
Preparation: $\nabla_f H$, $\nabla_b H$, $\tau$ are defined in (13), (14) and (15), respectively. The parameter $r$ is a positive integer.
Initialization: $f^{(0)} = 1$, $b^{(0)} = 0$.
repeat
    Step 1: $S^{(k)} \leftarrow \mathrm{diag}\left(\frac{f^{(k)}}{A^\top 1}\right)$
    Step 2: $h^{(k)} \leftarrow P_{\mathbb{R}_+^d}(f^{(k)} - \tau S^{(k)} \nabla_f H(f^{(k)}, b^{(k)}))$
    Step 3: $b^{(k+1)} \leftarrow Q_{h^{(k)}}^r (b^{(k)})$
    Step 4: $f^{(k+1)} \leftarrow P_{\mathbb{R}_+^d}\left(f^{(k)} - \tau S^{(k)} \nabla_f H(f^{(k)}, b^{(k+1)})\right)$
until "convergence"
```

Algorithm 3 (Semi-Dynamic PAPA for MAP ECT Reconstruction)

```
Preparation: $\nabla_f H$, $\nabla_b H$, $\tau$ are defined in (13), (14) and (15), respectively. The parameters $r$ and $l$ are positive integers.
Initialization: $f^{(0)} = 1$, $b^{(0)} = 0$.
Run Algorithm 2 until $k > l$.
Set $S^{(l)} = \mathrm{diag}\left(\frac{f^{(l)}}{A^\top 1}\right)$.
repeat
    Step 1: $h^{(k)} \leftarrow P_{\mathbb{R}_+^d}(f^{(k)} - \tau S^{(l)} \nabla_f H(f^{(k)}, b^{(k)}))$
    Step 2: $b^{(k+1)} \leftarrow Q_{h^{(k)}}^r (b^{(k)})$
    Step 3: $f^{(k+1)} \leftarrow P_{\mathbb{R}_+^d}\left(f^{(k)} - \tau S^{(l)} \nabla_f H(f^{(k)}, b^{(k+1)})\right)$
until "convergence"
```

and propose the following semi-dynamic PAPA (Algorithm 3). In this way, the convergence theorem established in Section 4 is still applicable.

To close this section we compare PAPA with four existing algorithms, namely, the alternating extragradient method (AEM) [7], the nested EM-TV algorithm [54], the nested iterative algorithm for convex constrained problem [13], and the preconditioned primal-dual algorithm (P-PD) [50, 56].

The AEM algorithm, which is a variant of the extragradient method, was developed to solve the saddle-point formulation of the Poisson-TV model (7). We point out that the AEM appears to be a special case of the proposed Algorithm 1 with the trivial preconditioner $S = I$ and inner iteration number $r = 1$, but it was developed based on a different principle. The introduction of inner iteration number r and a nontrivial preconditioner S in PAPA allows us to develop more efficient reconstruction algorithms.

We now compare PAPA with the nested EM-TV algorithm. Actually, the nested EM-TV requires to solve an optimization problem of the form

$$\operatorname{argmin}\left\{\frac{1}{2}\left\langle f - f^{(k+\frac{1}{2})}, \frac{A^\top 1}{f^{(k)}} \odot (f - f^{(k+\frac{1}{2})})\right\rangle + \lambda\varphi(Bf) : f \in \mathbb{R}^d_+\right\}$$

exactly in each outer iteration. While PAPA does not need to do this and it leads to a more efficient algorithm. This will be demonstrated by numerical examples in the next section.

Next we compare PAPA with the nested iterative algorithm proposed in [13] for solving convex constrained problems. The paper combined the forward-backward and the Douglas-Rachford iterations together to minimize the sum of two functions over a convex set. The validity of the resulting algorithm requires at least one of the two functions differentiable. While PAPA is developed based on a fixed-point characterization (in terms of the proximity operator) of the solutions of the Poisson-TV model (7), which does not necessarily require any term of the objective function to be differentiable.

Finally, we compare PAPA with the P-PD algorithm developed in [56], which prototyped several convex optimization problems for computed tomography (CT) image reconstruction with the primal-dual (PD) algorithm, proposed earlier in [12]. Within each complete iterate step, the PD algorithm introduces an extrapolation step based on the current and previous iterates, which can be seen as an approximate extragradient step. In contrast, PAPA computes the primal leading point $h^{(k)}$ by taking an extragradient step based on the current iterate only. Besides, PAPA made good use of the sparsity of the difference matrix B by carrying out the inner iteration to reduce the overall computational cost, while reference [56] did not. Indeed, matrix A is much denser than matrix B, and hence computation with A or $A^\top$ is more costly than that with B. As we have pointed out earlier, by carrying out the inner iteration with an appropriate iteration number r we could greatly accelerate the whole iterative scheme. Moreover, following the idea in [50], paper [56] developed a P-PD algorithm for model (7) using fixed diagonal preconditioners. While in PAPA, we proposed a dynamic preconditioner motivated from the EM algorithm, which proves to be more efficient and converging faster in the numerical experiments presented in the next section.

6 Numerical Experiments

In this section, we report numerical results obtained from computational experiments for the proposed algorithms. We compare our algorithms with the conventional EM-TV algorithm [48] and the nested EM-TV algorithm [54] in terms of the reconstruction quality and computational performance. We also compare the convergence speed of our proposed algorithms with that of the P-PD algorithm developed in [56].

6.1 Simulated SPECT Projection Data

We created a digital cylindrical emission phantom with uniform mean background activity distribution (that is, with uniform mean number of nuclear disintegrations per time unit and per unit volume) and sets of 7 hot spheres and 7 cold spheres embedded in the cylinder. The hot and cold spheres simulate hyperperfused and hypoperfused defects, respectively. Such defects are of interest in nuclear medicine and one of the main tasks of ECT is detection of such defects. Mean activities across all spheres are uniform. The mean activity ratios of hot:background:cold areas are 40:10:1, respectively. The pixel size used is 0.172 cm. The phantom dimensions are: base radius 84 pixels and length 128 pixels. The spheres radii are 3, 4, 5, 6, 7, 9 and 14 pixels. Their centers are in slices 33 and 97. The locations of spheres in transaxial and sagittal planes are shown in Figure 1. The spheres are separated by a uniform region located between slices 47 and 82. The mean activity distribution in the phantom represents the mean radiotracer distribution, that is, the image f in (1). The parallel-collimator SPECT projection data for our experiments consist of 120 views in 256×128 matrix with pixel size 1.78 mm and were generated using analytical pixel-wise discretized projector A in (1) with 20 rays per detector bin [59]. The generated data follow Poisson probability distribution created by a random number generator and the total number of detector counts in 120 views equal to 1.79×10^6 and 1.947×10^7 corresponding to approximately $10 : 1$ total activities ratio. Neither attenuation nor scatter was modeled and an ideal detector was assumed. Each image in these projection sets was then downsampled to a 128×64 matrix with pixel size 3.56 mm. The coefficient of variation (CV) defined as the ratio of the standard deviation to the mean in the projection image of a sufficiently large uniform region away from edges is used as a measure of Poisson noise in the noisy projection data. Using such a definition, our projection sets are characterized at 56% and 15% noise levels, respectively (Figure 2). For the sake of convenience, the phantom with higher (less noisy) and lower total activity (more noisy) is called"low-noise phantom" and "high-noise phantom", respectively. The low-noise phantom corresponds to clinically realistic SPECT data, while the high-noise phantom SPECT data would result in clinically unacceptable high noise in the reconstructed images if treated with conventional reconstruction techniques.

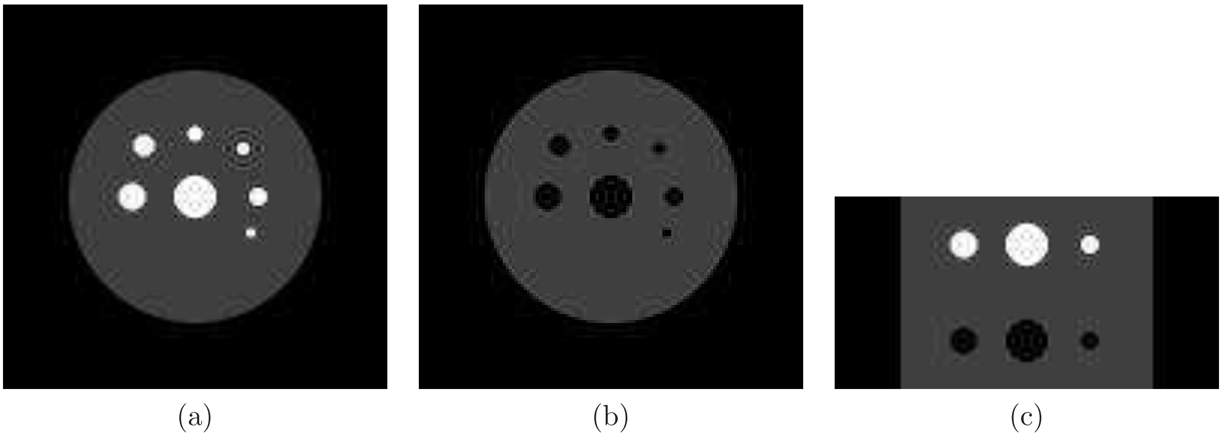
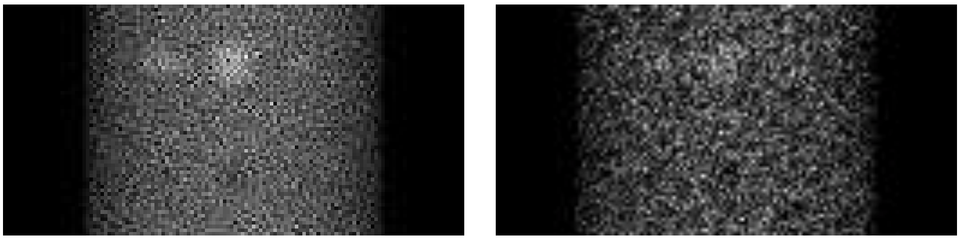

(a) (b) (c)

Figure 1: Morphology of numerical phantom used: (a) Transaxial cross-section (slice 17) through the centers of all hot spheres; (b)Transaxial cross-section (slice 49) through the centers of all cold spheres; (c) Sagittal cross-section through the centers of the two largest spheres. The mean activity ratios of hot:background:cold areas are 40:10:1, respectively. The spheres radii are counterclockwise: 3, 6, 4, 5, 7, 9 and 14 (center) pixels, respectively. The pixel size used is 0.172 cm.

Examples of one noise realization for one projection view are shown in Figure 2.



(a) (b)

Figure 2: Example of one parallel-beam collimator SPECT projection view, out of 120 views in the projection set, simulated for one noise realization for a digital phantom shown in Figure 1: (a) 15% Poisson noise, 1.62×10^5 counts per view; (b) 56% Poisson noise, 1.49×10^4 counts per view.

6.2 Numerical Studies of the Proposed Algorithms

In this subsection, we assess the numerical performance of the proposed algorithms: Algorithms 1, 2 and 3 in image reconstruction for the SPECT projection data generated from the low-noise and high-noise phantoms. Specifically, we consider three issues related to the proposed algorithms: the necessity of the preconditioning, the convergence of the proposed preconditioner (60) and the role of the multiple iteration steps in the inner iteration.

In our numerical experiments, the regularization parameter λ in the Poisson-TV model (7) is chosen by adopting the Bias-Noise curve method. Specifically, we calculated bias and the coefficient of variation (CV) in a reconstructed background region of interest for candidate regularization parameters (ranging from 1×10^{-4} to 1), and obtained the bias-CV curve. When choosing the optimal parameters, we consider the best trade-off between bias and CV. Furthermore, in order to improve the statistical accuracy, we generated three SPECT projection data sets with different noise realizations from the low-noise and the high-noise phantom, respectively, and evaluated the mean values of bias and CV with respect to the three different noise realizations. In particular, in our numerical experiments, the regularization parameter λ in the Poisson-TV model (7) is chosen to be 0.1 and 0.2 for the SPECT projection data sets generated from low-noise and high-noise phantom, respectively. The constant γ in the model is set to be 0.01 for both phantoms. Let $\{f^{(k)} : k \in \mathbb{N}\}$ be a sequence generated by either Algorithm 1, 2 or 3. For a pre-given tolerance tol, the iterative process of an algorithm is terminated if the following requirement is satisfied

$$\|f^{(k)} - f^{(k+1)}\|/\|f^{(k+1)}\| \leq \text{tol}. \tag{61}$$

The parameters in Algorithms 1, 2 and 3 are specified as follows. For Algorithm 1 without preconditioning ($S = I$) we choose β to be 10^7 times of the upper bound suggested by Theorem 4.7, and $\mu := 1/(2\beta\|B\|_2^2)$. For Algorithm 1 with the fixed diagonal preconditioner $\operatorname{diag}\left(\frac{1}{A^\top 1}\right)$, we set $\beta = \lambda$, and $\mu = 1/(2\lambda\|B\|_2^2\|\frac{1}{A^\top 1}\|_\infty)$. For Algorithms 2 and 3 we set $\beta = \lambda$, and $\mu = 1/(2\lambda\|B\|_2^2\|\frac{f^{(k)}}{A^\top 1}\|_\infty)$ in their k-th iterations.

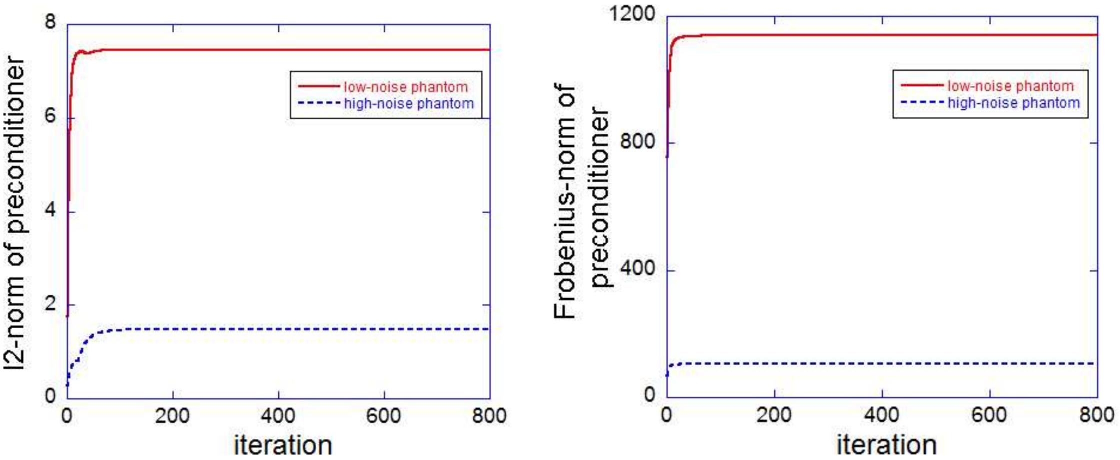
In the first experiment, we compare the performance of Algorithm 1 without preconditioning ($S = I$) with that of Algorithm 1 with the fixed diagonal preconditioner $\operatorname{diag}\left(\frac{1}{A^\top 1}\right)$ and that of Algorithms 2 and 3. The parameter r for all the algorithms is fixed at 10 and the other parameters for each algorithm are chosen as discussed above. When implementing Algorithm 3, in the first 100 iterations, we use the dynamic preconditioner and after 100 iterations, we fix the preconditioner. Table 1 gives a summary of the CPU times (process times) and numbers of the complete iterations for reconstructing images from the noisy SPECT projection data set generated from low-noise phantom. It clearly shows that Algorithms 2 and 3, which have dynamic and semi-dynamic

Table 1: The pair $(\cdot,\cdot)$ represents the CPU time and the number of the complete iterations used in Algorithms 1, 2, 3 and P-PD algorithm with different choices of tol for the noisy SPECT projection data set generated from low-noise phantom. T represents the diagonal matrix diag $\left(\frac{1}{A^\top 1}\right)$.

tol Alg.	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
Alg. 1 ($S = I$)	(5.4, 2)	(70.2, 27)	(250.9, 99)	(1359.6, 525)	(4300.9, 1653)	(10241.8, 3940)	(-, -)
Alg. 1 ($S = T$)	(5.4, 2)	(57.5, 23)	(210.5, 85)	(1088.5, 445)	(3204.9, 1301)	(7383.4, 3000)	(-, -)
Alg. 2	(12.6, 5)	(35.0, 14)	(109.6, 44)	(291.1, 117)	(768.6, 309)	(1915.6, 770)	(5435.0, 2109)
Alg. 3	(12.3, 5)	(34.8, 14)	(109.6, 44)	(305.7, 123)	(759.5, 307)	(1464.3, 592)	(3954.8, 1599)
P-PD Alg.	(6.6, 2)	(75.9, 23)	(292.7, 89)	(1569.1, 479)	(4357.6, 1331)	(9885.1, 3024)	(-, -)

EM-preconditioners respectively, converge significantly faster than Algorithm 1 with either preconditioner. We remark that Algorithm 1 with either preconditioner cannot meet the stopping criteria for the last tolerance level. This phenomenon is marked by $(-,-)$ in the table. Therefore, in the remaining part of this section, only Algorithm 3 will be used to compare with other existing algorithms for the Poisson-TV model (7). For comparison, we include in Table 1 the numerical results for the P-PD algorithm [56] which also uses a fixed preconditioner. Its performance is even worse than that of Algorithm 1 with the fixed diagonal preconditioner diag $\left(\frac{1}{A^\top 1}\right)$.

The proposed preconditioner (60) depends on the current iterate and is updated at each iterate step. The second experiment explores numerically how the preconditioner (60) changes in iterations in terms of its ℓ^2-norm and Frobenius-norm. Figure 3 presents the curves of both norms versus iteration numbers for the low-noise and high-noise phantoms. From the figure we conclude that after a few iterations, the change of the preconditioner (60) is neglectable. This suggests that except for the first few iterations, the preconditioner may be chosen to be the same. In this regard, the convergence result (Theorem 4.7) that we establish for the fixed preconditioner is applicable to the practical situation.



(a) (b)

Figure 3: Curves of (a) ℓ^2-norm and (b) Frobenius-norm of the preconditioner (60) versus iteration numbers for low-noise and high-noise phantoms.

The third experiment tests numerically the choice of the parameter r in the inner iteration for Algorithm 3 (PAPA) in terms of the CPU time that the algorithm requires to reach the stopping criteria (61) for different tolerances. We perform PAPA with r from 6 to 15 for the noisy SPECT projection data sets generated from low-noise and high-noise phantoms with the tolerance tol $= 10^{-6}$. The numerical results for the noisy SPECT projection data set generated from the low-noise phantom show that the algorithms with the parameter r ranging from 7 to 11 perform comparably and are better than those with other ranges of r. The numerical results for the noisy SPECT projection data set generated from the high-noise phantom indicate that the algorithms with the parameter r ranging from 10 to 12 perform comparably and are better than those with other ranges of r. According to this numerical observation, in the remaining part of this section, we fix the parameter r for PAPA at 10.

6.3 Comparison of PAPA with the EM-TV Algorithm

We now compare performance of the proposed PAPA with that of the conventional EM-TV [48], in terms of the CPU time, local image quality metrics: the coefficient-of-variation (CV) measured within the uniform region of interest, contrast-to-noise ratio (CNR), and a global image quality metric: the normalized mean-squared error (NMSE).

We first recall the conventional EM-TV algorithm. For a given positive number δ, the smoothed version of the total variation is a function $R : \mathbb{R}^d \to \mathbb{R}$ defined at $u \in \mathbb{R}^d$ as

$$R(u) := \sum_{i=1}^{d} \left\| [(Bu)_i, (Bu)_{d+i}, (Bu)_{2d+i}, \delta]^\top \right\|. \tag{62}$$

With this function R, the conventional EM-TV algorithm [48] is described in Algorithm 4.

Algorithm 4 (Conventional EM-TV Algorithm for MAP ECT Reconstruction)

```
Initialization: $f^{(0)} = 1$.
repeat
    $f^{(k+1)} \leftarrow \mathrm{diag}\left(\frac{f^{(k)}}{A^\top 1 + \lambda \nabla R(f^{(k)})}\right) A^\top \left(\frac{g}{Af^{(k)} + \gamma}\right)$.
until "convergence"
```

For reconstructions with the conventional EM-TV algorithm, following [48] we choose $\delta = 0.001$ which is less than 1% of the maximum value of the phantom. Moreover, we note that when λ vanishes the conventional EM-TV algorithm reduces to the EM algorithm (59).

We now evaluate the quality of the reconstructed images. First, we compare noise in the reconstructed images. We use the coefficient of variation (CV) within a uniform region of interest as a surrogate of noise measure in the reconstructed images obtained using Algorithm 2 and the conventional EM-TV. The CV of an image f^∞ reconstructed by an algorithm is defined by

$$\mathrm{CV} := \frac{\mathrm{SD}_\Omega(f^\infty)}{\mathbb{E}_\Omega(f^\infty)},$$

where $\mathrm{SD}_\Omega(\cdot)$ and $\mathbb{E}_\Omega(\cdot)$ denote the standard deviation and the mean of the reconstructed activities over a region Ω. In our case, the region Ω is a cylinder with the radius of base and the height equal to 25 and 8 pixels, respectively. This cylindrical region lies between the hot spheres and the cold spheres (please refer to Figure 1 (c)) and does not intersect with any hot or cold spheres. Under these circumstances, the means of CVs (with respect to 5 different noise realizations) for PAPA and the conventional EM-TV are 0.12% and 3.81%, respectively, for the low-noise phantom, and are 4.09% and 13.15%, respectively, for the high-noise phantom. From these numerical results, we find out that PAPA performs considerably (factor of 31 for low-noise phantom data and factor of 3 for high-noise phantom data) better than the conventional EM-TV in terms of the noise.

Next, a local quality metric, contrast-to-noise ratio (CNR), of images will be used to measure the quality of the reconstructed hot or cold spheres. Too low CNR might result in inability to detect the lesion by an observer. The contrast-to-noise ratio (CNR) for a reconstructed image f^∞ is defined as a ratio of a lesion contrast to the background noise [4, 40, 51]

$$\mathrm{CNR} := \frac{|\mathbb{E}_{\Omega_T}(f^\infty) - \mathbb{E}_{\Omega_B}(f^\infty)|}{\mathrm{SD}_{\Omega_B}(f^\infty)}.$$

Here $\mathbb{E}_{\Omega_T}(f^\infty)$ is the mean reconstructed activity in the region Ω_T where a specific hot (resp. cold) sphere is located, while Ω_B is a spherical region with the same diameter as the hot (resp. cold) sphere but located within the uniform region of the cylindrical phantom that does not intersect with any hot or cold sphere. The CNR metric is important in Radiology because it allows to asses detectability of a lesion that is one of the main tasks of ECT. It is well established by human observer studies that the lesion contrast and the background noise influence lesion detectability [40]. We note that even when the signal-to-noise ratio is high, the presence of a significant bias in the reconstructed image might result in the lesion contrast being too low for a lesion to be detected. Very high noise or correlation of the noise will contribute to low detectability of the lesion even if the lesion contrast is high [46]. The curves of the means of CNRs (with respect to 5 different noise realizations again) versus diameters of spheres are plotted in Figure 4. Again, we observe that PAPA very significantly (factor of 7-14 for high-noise phantom) outperforms the conventional EM-TV in terms of the CNR values.

We use the normalized mean-squared error (NMSE) to assess accuracy of reconstructions. The NMSE is a global image quality metric. It quantifies the difference between the activity reconstruction f^∞ and the true mean activity f in the whole object. It is defined by

$$\mathrm{NMSE} := \frac{\|f - f^\infty\|^2}{\|f\|^2}.$$

The mean values of NMSE (with respect to 5 different noise realizations) for the images reconstructed by PAPA and the conventional EM-TV are 0.485 and 0.787, respectively, for the low-noise phantom, and 2.55 and 3.93, respectively, for the high-noise phantom. Once again, PAPA significantly outperforms the conventional EM-TV in terms of the NMSE values.

6.4 Comparison of PAPA with the Nested EM-TV Algorithm

In this subsection, we compare performance of PAPA with the nested EM-TV algorithm, the algorithm recently described in [54].

In Algorithm 5, we describe the nested EM-TV algorithm for ECT reconstruction. This algorithm has two major steps: the EM step and the TV correction step, and they are performed

Algorithm 5 (Nested EM-TV Algorithm for MAP ECT Reconstruction)

```
Initialization: $f^{(0)} = 1$.
repeat
    EM Step: $f^{(k+\frac{1}{2})} \leftarrow \mathrm{diag}\left(\frac{f^{(k)}}{A^\top 1}\right) A^\top \left(\frac{g}{Af^{(k)}+\gamma}\right)$.
    TV Step: $f^{(k+1)} \leftarrow \mathrm{argmin}\left\{\frac{1}{2}\left\langle f - f^{(k+\frac{1}{2})}, \frac{A^\top 1}{f^{(k)}} \odot (f - f^{(k+\frac{1}{2})})\right\rangle + \lambda\varphi(Bf) : f \in \mathbb{R}_+^d\right\}$.
until "convergence"
```

alternatively. The EM step is identical to (59). The TV correction step is a modified version of the Rudin-Osher-Fatemi (ROF) model and was implemented by exploiting an existing scheme for the ROF model. In fact, it was stated in [54] that the TV correction step in Algorithm 5 was carried out by adopting Chambolle's method originally reported in [11]. We shall follow the suggestion made in [54] to apply Chambolle's method with 10 iterations for each TV correction step to achieve an approximate solution.

In Table 2, we list the CPU time expended and the number of the complete iterations used by PAPA, the nested EM-TV and the conventional EM-TV, for the two SPECT projection data sets. After examining the table, we conclude that under the same stopping criteria PAPA is better than the nested EM-TV, and significantly better than the conventional EM-TV in terms of the convergence speed. We remark that the conventional EM-TV algorithm cannot even meet the stopping criteria for most of the given tolerance levels.

Table 2: Comparison of performance of PAPA, the conventional EM-TV and the nested EM-TV for the test data. The pair $(\cdot, \cdot)$ represents the CPU time and the number of the complete iterations used.

tol Algorithm	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
low-noise phantom							
PAPA	(12.3, 5)	(34.8, 14)	(109.6, 44)	(305.7, 123)	(759.5, 307)	(1464.3, 592)	(3954.8, 1599)
Nested EM-TV	(14.1, 5)	(40.0, 14)	(122.8, 43)	(333.1, 117)	(872.4, 307)	(2207.3, 777)	(6184.5, 2179)
EM-TV	(7.15, 4)	(-, -)	(-, -)	(-, -)	(-, -)	(-, -)	(-, -)
high-noise phantom							
PAPA	(12.4, 5)	(35.3, 14)	(106.2, 42)	(295.3, 117)	(724.7, 289)	(1451.6, 580)	(4349.4, 1741)
Nested EM-TV	(13.2, 5)	(38.1, 14)	(115.2, 42)	(304.5, 111)	(798.8, 291)	(2024.3, 737)	(5593.6, 2045)
EM-TV	(9.01, 5)	(-, -)	(-, -)	(-, -)	(-, -)	(-, -)	(-, -)

In addition to the CPU time and the number of complete iterations, we further estimate the number of arithmetic operations for PAPA and the nested EM-TV algorithm. We list in Table 3 the number of total arithmetic operations required by PAPA and the nested EM-TV, under various stopping criteria.

Table 3: Comparison of arithmetic operations of semi-dynamic PAPA and the nested EM-TV for the test data. The number $\cdot$ represents $\cdot \times 2^{20}$ arithmetic operations required.

tol Algorithm	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
low-noise phantom							
PAPA	3380	9464	29744	83148	207532	400192	1080924
Nested EM-TV	3795	10626	32637	88803	233013	589743	1653861
high-noise phantom							
PAPA	3380	9464	28392	79092	195364	392080	1176916
Nested EM-TV	3795	10626	31878	84249	220869	559383	1552155

Furthermore, we compare the image quality metrics of the reconstructed images obtained by using PAPA and by the nested EM-TV. The means of CVs for the reconstructions by PAPA and the conventional EM-TV are reported in Table 4. The means of CNRs with respect to 5 different noise realizations versus the diameters of the hot spheres (resp. cold spheres) for PAPA, the nested EM-TV, and the conventional EM-TV are listed in Table 5 (resp. 6). Moreover, the means of

NMSEs with respect to 5 different noise realizations for the images of the low-noise phantom and the high-noise phantom are presented in Table 7. We conclude from these tables that both PAPA and the nested EM-TV outperform the conventional EM-TV in terms of the values of CV, CNR and NMSE while PAPA performs comparably with the nested EM-TV.

Table 4: Means of CVs for the reconstructions by PAPA, the nested EM-TV and the conventional EM-TV.

Phantom	low-noise phantom	high-noise phantom
PAPA	0.12%	4.09%
nested EM-TV	0.14%	4.54%
EM-TV	3.81%	13.15%

Table 5: Means of CNRs for 7 hot spheres reconstructed using PAPA, the nested EM-TV and the conventional EM-TV.

Diameters of Spheres	14	9	7	6	5	4	3
low-noise phantom							
PAPA	1915.768	1590.557	1574.134	1451.297	1280.732	1043.835	611.271
nested EM-TV	1912.141	1587.496	1571.388	1448.454	1277.969	1041.962	610.677
EM-TV	65.801	55.159	56.717	51.872	47.284	40.451	28.124
high-noise phantom							
PAPA	52.171	41.019	37.525	31.919	28.247	13.939	5.448
nested EM-TV	52.149	41.002	37.510	31.905	28.234	13.932	5.445
EM-TV	17.007	13.116	11.885	9.815	8.389	4.237	2.232

Table 6: Means of CNRs for 7 cold spheres reconstructed using PAPA, the nested EM-TV and the conventional EM-TV.

Diameters of Spheres	14	9	7	6	5	4	3
low-noise phantom							
PAPA	543.563	427.444	330.512	283.982	229.776	97.887	14.505
nested EM-TV	542.524	426.538	330.087	283.587	229.380	97.331	14.437
EM-TV	18.721	15.195	12.599	11.744	10.610	6.785	3.993
high-noise phantom							
PAPA	14.378	10.106	6.577	8.570	1.543	0.424	0.764
nested EM-TV	14.373	10.105	6.576	8.568	1.542	0.424	0.762
EM-TV	4.929	3.544	2.484	3.128	1.310	0.724	0.839

In order to qualitatively compare the image reconstruction quality, in Figures 5 and 6 we present selected transaxial cross-sections through the reconstructed images of the hot and cold spheres, respectively. Images reconstructed by PAPA, the nested EM-TV, and the conventional EM-TV for two noise levels (56% and 15%) are shown. We observe much higher background noise in the images reconstructed by the conventional EM-TV, as compared to PAPA and the nested EM-TV. The quality of images reconstructed by PAPA and the nested EM-TV is similar. The hot sphere with 4-pixel radius cannot be detected in the images reconstructed from high-noise data

Table 7: Means of NMSEs for the reconstructions by PAPA, the nested EM-TV and the conventional EM-TV.

Phantom	low-noise phantom	high-noise phantom
PAPA	0.48469	2.54701
nested EM-TV	0.48481	2.54794
EM-TV	0.78670	3.92571

by the conventional EM-TV. But, it is detectable in the reconstructions performed using PAPA and the nested EM-TV. The cold spheres with 6 and 7-pixel radii are poorly visible in the images reconstructed from high-noise data by the conventional EM-TV. However, they are easily detectable in the reconstructions performed using PAPA and the nested EM-TV.

To better access the differences between images reconstructed using PAPA and the nested EM-TV, we obtain line profiles through selected transaxial cross-sections of the reconstructed images containing the hot and cold spheres. They are shown in Figures 7 and 8 for hot and cold spheres, respectively. We observe that for hot spheres reconstructions PAPA provides images with slightly better contrast and spatial resolution, as compared to the nested EM-TV. For cold spheres reconstructions both algorithms perform similarly. The conventional EM-TV reconstructions are inferior in any case.

7 Concluding Remarks

There is a great need to reduce radiation dose to the patients undergoing ECT examinations. This could be accomplished by lowering the total amount of activity in the radiotracer administered. However, it would lead to very high Poisson noise in the raw ECT data. In turn, such very noisy data if treated by conventional techniques, such as EM-TV or OSEM, would lead to very noisy and clinically unacceptable reconstructed images. To attain good quality ECT reconstructions from low-dose ECT examinations, we propose a preconditioned alternating projection algorithm (PAPA) for solving the maximum *a posteriori* (MAP) ECT reconstruction problem. We prove that the algorithm enjoys nice theoretical convergence results in the case that the preconditioner is fixed. Motivated by the classical EM algorithm, we propose dynamic and semi-dynamic EM-preconditioners for PAPA to accelerate convergence of the original iterative scheme, which is the main contribution of this work. We demonstrate in the numerical experiments that the EM-preconditioner converges fast to a fixed preconditioning matrix, which in turn confirms the applicability of the convergence result to the practical situation. Since the total-variation (TV) based penalty function can well preserve the edges and details of the reconstructed object, we particularly concentrate on the example with TV regularization. Based on the numerical experiments performed in this work, we observe that the alternating projection algorithm with the EM-preconditioner significantly outperforms the conventional EM-TV in all aspects including the convergence speed, the noise in the reconstructed images and the image quality. It also outperforms the recently developed algorithm - the nested EM-TV - in the convergence speed while having a comparable image quality.

We conclude that the developed alternating projection algorithm with dynamic or semi-dynamic EM-preconditioner might allow very significant reduction in the radiation dose to the patients imaged using ECT by providing the same contrast-to-noise ratio for hot and cold lesions as conventional EM-TV, but with the total administered radiotracer activity 2 to 6 times lower than presently used in ECT examinations reconstructed using the conventional EM-TV.

Acknowledgement: This work is supported in part by US Air Force Office of Scientific Research under grant FA9550-09-1-0511, by the US National Science Foundation under grants DMS-0712827, DMS-1115523, and by the Natural Science Foundation of China under grants 11071286 and 91130009, by Guangdong Provincial Government of China through the "Computational Science Innovative Research Team" program.

This work is also supported in part by Award 5-28527 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH) and the NIH Roadmap for Medical Research, and by the Center of Emerging and Innovative Sciences (CEIS), a NYSTAR-designated Center for Advanced Technology. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Research Resources or the National Institutes of Health. This work is supported in part by Carol M. Baldwin Breast Cancer Research Award.

References

[1] J. Bardsley. An efficient computational method for total variation-penalized poisson likelihood estimation. *Inverse Problems and Imaging*, 2:167–185, 2008.

[2] J. Bardsley and J. Goldes. Regularization parameter selection and an efficient algorithm for total variation-regularized positron emission tomography. *Numerical Algorithms*, 57:255–271, 2011.

[3] J. Bardsley and A. Luttman. Total variation-penalized Poisson likelihood estimation for ill-posed problems. *Advances in Computational Mathematics*, 31:35–59, 2009.

[4] H. Barrett. Objective assessment of image quality: effects of quantum noise and object variability. *Journal of the Optical Society of America A*, 7:1266–1278, 1990.

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

[6] D. P. Bertsekas. Projected newton methods for optimization problems with simple constraints. *SIAM Journal on Control and Optimization*, 20:221–246, 1982.

[7] S. Bonettini and V. Ruggiero. An alternating extragradient method for total variation-based image restoration from poisson data. *Inverse Problems*, 27(9):095001, 2011.

[8] J. M. Borwein and A. S. Lewis. *Convex Analysis and Nonlinear Optimization, Theory and Examples.* CMS Books in Mathematics. Springer-Verlag, New York, 2006.

[9] J. Browne and A. de Pierro. A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography. *IEEE Transactions on Medical Imaging*, 15:687–699, 1996.

[10] C. Byrne. Block-iterative interior point optimization methods for image reconstruction from limited data. *Inverse Problems*, 16:1405–1419, 2000.

[11] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20:89–97, 2004.

[12] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40:120–145, 2011.

[13] C. Chaux, J. Pesquet, and N. Pustelnik. Nested iterative algorithms for convex constrained image recovery problems. *Optimization and Control*, arXiv:0806.3920, 2009.

[14] G. Chinn and S. Huang. A general class of preconditioners for statistical iterative reconstruction of emission computed tomography. *IEEE Transactions on Medical Imaging*, 16:1–10, 1997.

[15] N. Clinthorne, T. Pan, P. Chiao, W. Rogers, and J. Stamos. Preconditioning methods for improved convergence rates in iterative reconstructions. *IEEE Transactions on Medical Imaging*, 12:78–83, 1993.

[16] A. de Pierro. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *IEEE Transactions on Medical Imaging*, 14:132–137, 1995.

[17] J. Fessler. Penalized weighted least squares image reconstruction for positron emission tomography. *IEEE Transactions on Medical Imaging*, 13:290–300, 1994.

[18] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[19] S. Geman and D. MacClure. Bayesian image analysis: an application to single photon emission tomography. *Proceedings of the Statistical Computing Section*, pages 12–18. American Statistical Association, Alexandria, 1985.

[20] P. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Transactions on Medical Imaging*, 9:84–93, 1990.

[21] T. Hebert and S. Gopal. The GEM MAP algorithm with 3-D SPECT system response. *IEEE Transactions on Medical Imaging*, 11:81–90, 1992.

[22] T. Hebert and R. Leahy. A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs *priors. IEEE Transactions on Medical Imaging*, 8:194–202, 1989.

[23] T. Hebert and R. Leahy. Statistic-based MAP image reconstruction from Poisson data using Gibbs *priors. IEEE Transactions on Signal Processing*, 40:2290–2303, 1992.

[24] D. Higdon, J. Bowsher, V. Johnson, T. Turkington, D. Gilland, and R. Jaszczak. Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data. *IEEE Transactions on Medical Imaging*, 16:516–526, 1997.

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

[26] V. Johnson, W. Wong, X. Hu, and C. Chen. Image restoration using Gibbs *priors*: boundary modeling, treatment of blurring, and selection of hyperparameters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:413–425, 1991.

[27] V. Johnson, W. Wong, X. Hu, and C. Chen. Bayesian restoration of PET images using Gibbs *priors. Progress in clinical and biological research*, 363:15–28, 1991.

[28] E. Jonsson, C. Huang, and T. Chan. Total variation regularization in positron emission tomography. UCLA CAM Report 98-48, 1998.

[29] C. Kao, X. Pan, C. Chen, and W. Wong. Image restoration and reconstruction with a Bayesian approach. *Medical Physics*, 25:600–613, 1998.

[30] D. Lalush and B. Tsui. Simulation evaluation of Gibbs *prior* distributions for use in maximum *a posteriori* SPECT reconstructions. *IEEE Transactions on Medical Imaging*, 11:267–275, 1992.

[31] D. Lalush and B. Tsui. A generalized Gibbs *prior* for maximum *a posteriori* reconstruction in SPECT. *Physics in Medicine and Biology*, 38:729–741, 1993.

[32] D. Lalush and B. Tsui. Block-iterative techniques for fast 4D reconstruction using a priori motion models in gated cardiac SPECT. *Physics in Medicine and Biology*, 43:875–886, 1998.

[33] K. Lange. Convergence of EM image reconstruction algorithms with Gibbs smoothing. *IEEE Transactions on Medical Imaging*, 9:439–446, 1990.

[34] K. Lange, M. Bahn, and R. Little. A theoretical study of some maximum likelihood algorithms for emission and transmission tomography. *IEEE Transactions on Medical Imaging*, 6:106–114, 1987.

[35] K. Lange and R. Carson. EM reconstruction algorithms for emission and transmission tomography. *Journal of Computer Assisted Tomography*, 8:306–316, 1984.

[36] R. Leahy and J. Qi. Statistical approaches in quantitative positron emission tomography. *Statistics and Computing*, 10:147–165, 2000.

[37] S. Lee. Performance comparison of convex-nonquadratic priors for Bayesian tomographic reconstruction. *Journal of Electronic Imaging*, 9:242–250, 2000.

[38] S. Lee, M. Lee, V. Nguyen, S. Kim, and J. Lee. Three-dimensional edge-preserving regularization for compton camera reconstruction. *IEEE Nuclear Science Symposium Conference Record*, pages 4223-4228, 2008.

[39] 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, to appear, 2012.

[40] L. Loo, K. Doi, and C. Metz. A comparison of physical image quality indices and observer performance in the radiographic detection of nylon beads. *Physics in Medicine and Biology*, 29:837–856, 1984.

[41] J. Ma, Q. Feng, Y. Feng, J. Huang, and W. Chen. Generalized Gibbs *priors* based positron emission tomography reconstruction. *Computers in Biology and Medicine*, 40:565–571, 2010.

[42] C. A. Micchelli, L. Shen, and Y. Xu. Proximity algorithms for image models: Denoising. *Inverse Problems*, 27:045009(30pp), 2011.

[43] C. A. Micchelli, L. Shen, Y. Xu, and X. Zeng. Proximity algorithms for image models II: L1/TV denosing. *Advances in Computational Mathematics*, published online, (2011).

[44] J.-J. Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965.

[45] E. Mumcuoglu, R. Leahy, S. Cherry, and Z. Zhou. Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images. *IEEE Transactions on Medical Imaging*, 13:687–701, 1994.

[46] K. Myers, H. Barrett, M. Borgstrom, D. Patton, and G. Seeley. Effect of noise correlation on detectability of disk signals in medical imaging. *Journal of the Optical Society of America A*, 2:1752–1759, 1985.

[47] J. Nuyts, D. Bequé, P. Dupont, and L. Mortelmans. A concave *prior* penalizing relative differences for maximum-*a-posteriori* reconstruction in emission tomography. *IEEE Transactions on Nuclear Science*, 49:56–60, 2002.

[48] V. Panin, G. Zeng, and G. Gullberg. Total variation regulated EM algorithm. *IEEE Transactions on Nuclear Science*, 46:2202–2210, 1999.

[49] M. Phelps. *PET: Molecular Imaging and Its Biological Applications.* Springer-Verlag, Germany, 2004.

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

[51] J. Qi and R. Leahy. A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data. *IEEE Transactions on Medical Imaging*, 18:293–305, 1999.

[52] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.

[53] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

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

[55] L. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1:113–122, 1982.

[56] E. Sidky, J. Jorgensen, and X. Pan. Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm. *Physics in Medicine and Biology*, 57:3065, 2012.

[57] E. Tanaka and H. Kudo. Subset-dependent relaxation in block-iterative algorithms for image reconstruction in emission tomography. *Physics in Medicine and Biology*, 48:1405–1422, 2003.

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

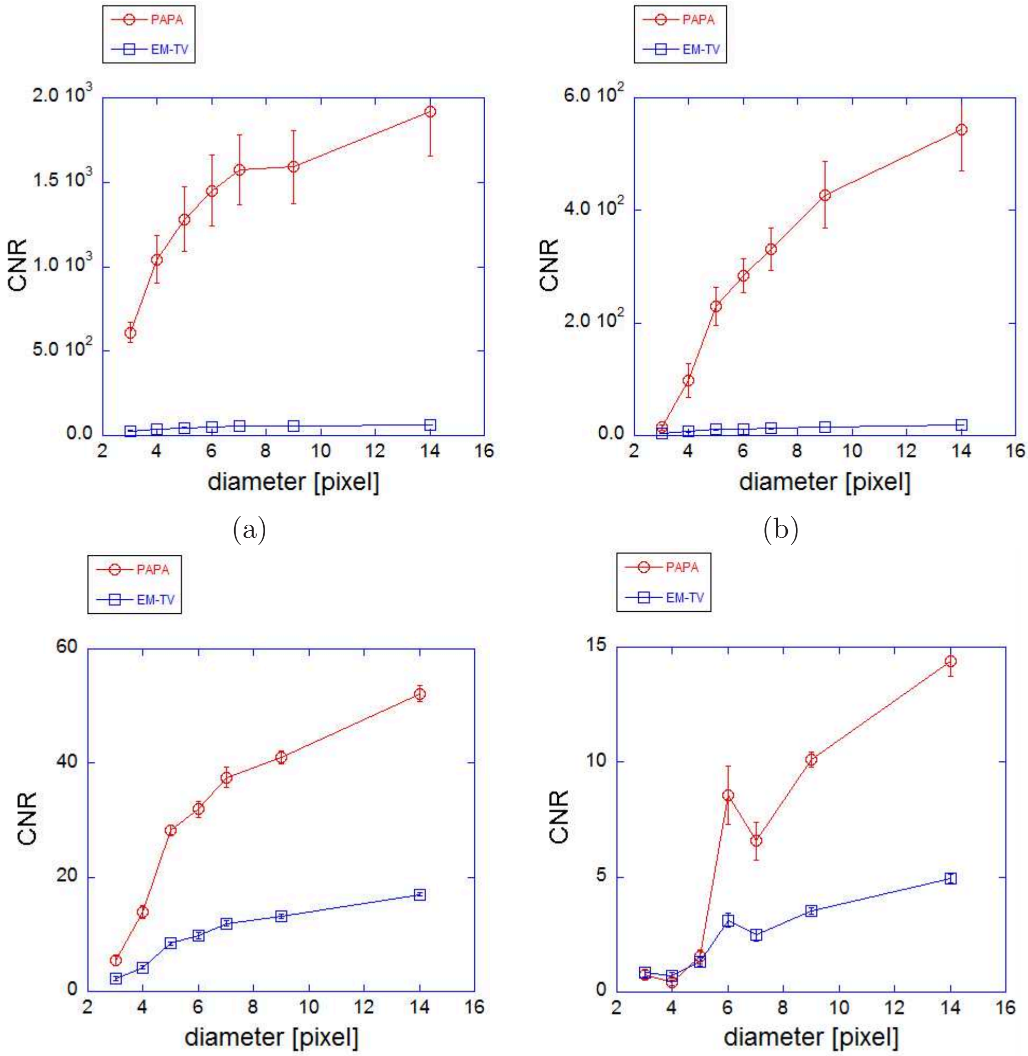
[59] L. Vogelsang. *Development of SPECT and CT tomographic image reconstruction.* PhD thesis, Syracuse University, 2011.

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

[61] Z. Zhou, R. Leahy, and E. Mumcuoglu. Maximum likelihood hyperparameter estimation for Gibbs *priors* with applications to PET. *Information Processing in Medical Imaging*, pages 39–42. Kluwer Academic Press, Dordrecht, 1995.

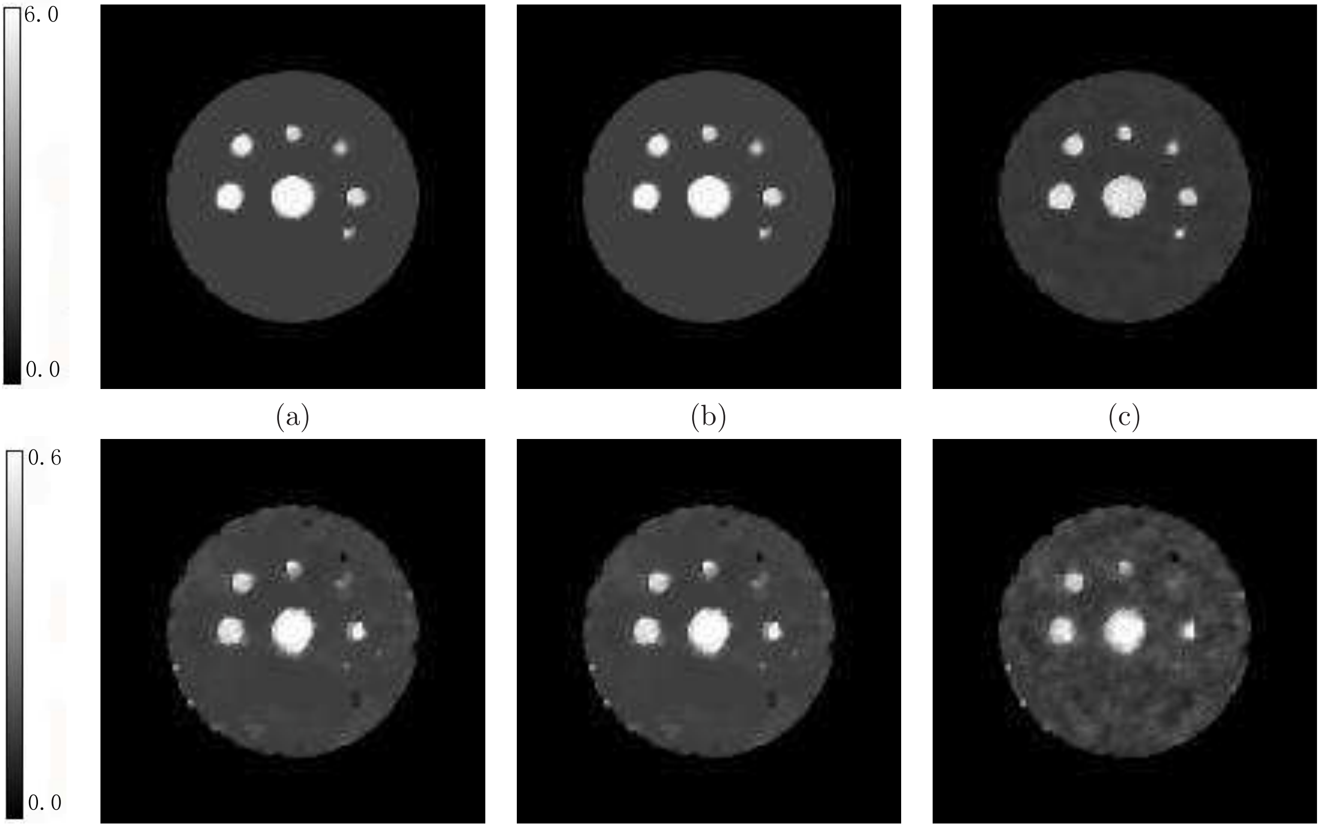
[62] Z. Zhou, R. Leahy, and J. Qi. Approximate maximum likelihood hyperparameter estimation for Gibbs *priors. IEEE Transactions on Image Processing*, 6:844–861, 1997.

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



(a) (b) (c) (d)

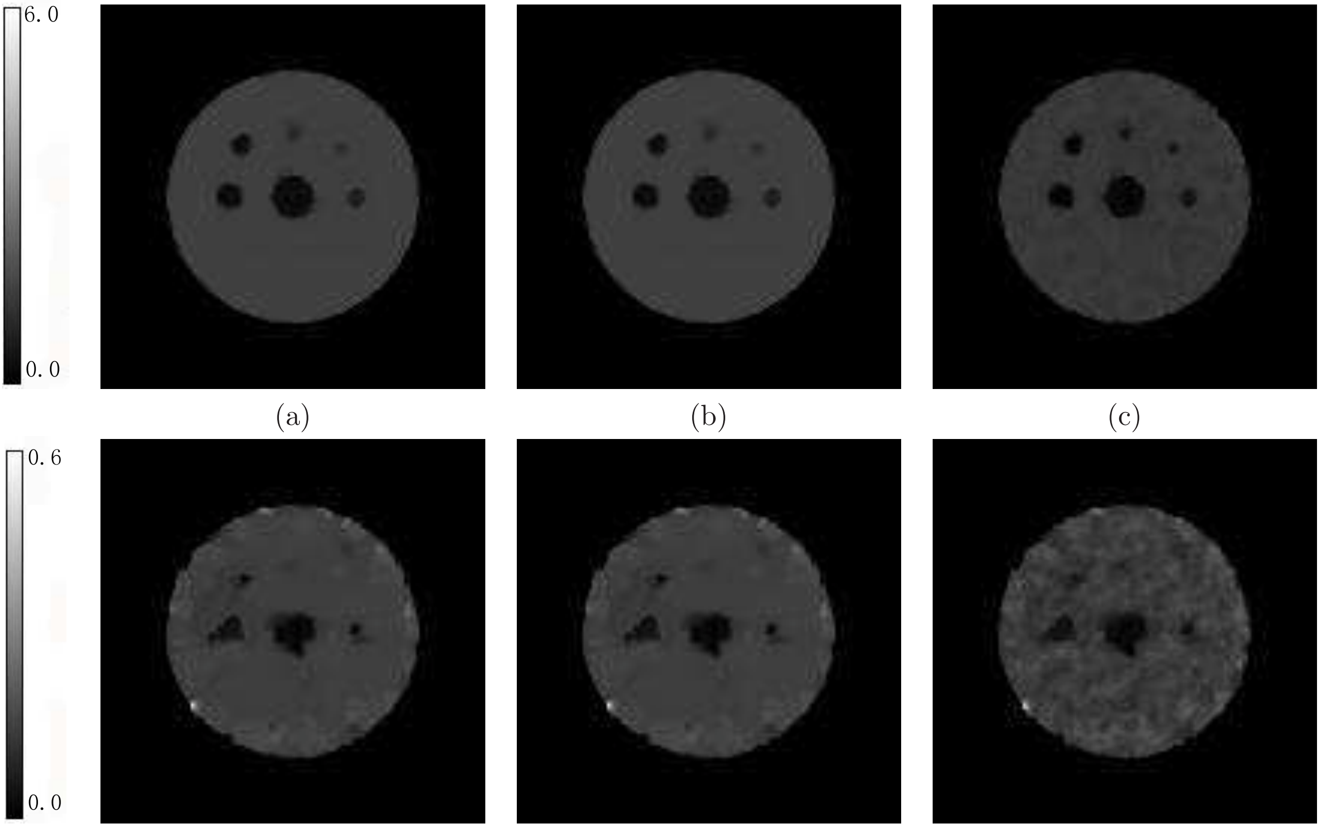
Figure 4: Reconstructed image quality assessment for hot and cold spheres: Plots of mean (with respect to 5 different noise realizations) contrast-to-noise ratio (CNR) vs. diameter of sphere. The vertical error bars denote the calculated standard deviation of the mean. For the EM-TV the error bars are smaller than the symbols used: (a) hot spheres in low-noise phantom; (b) cold spheres in low-noise phantom; (c) hot spheres in high-noise phantom; (d) cold spheres in high-noise phantom. Open squares - the conventional EM-TV. Open circles - PAPA.



(a) (b) (c)

(d) (e) (f)

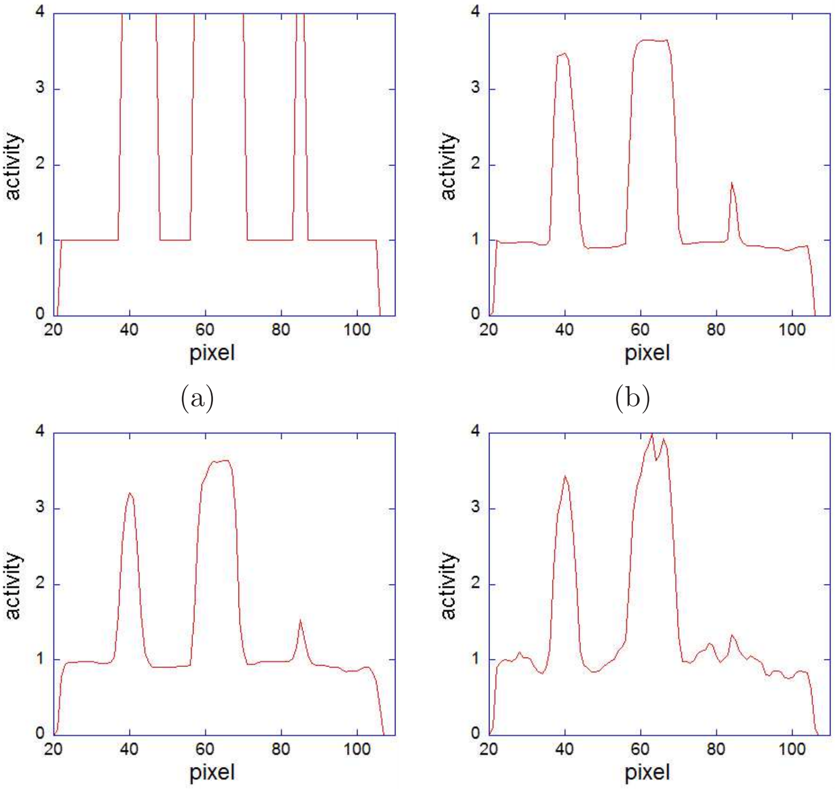
Figure 5: Transaxial cross-sections through the centers of the hot spheres (slice 17) in images reconstructed for low-noise phantom: (a) PAPA; (b) the nested EM-TV; and (c) the conventional EM-TV. Transaxial cross-sections through the centers of the hot spheres (slice 17) in images reconstructed for high-noise phantom: (d) PAPA; (e) the nested EM-TV; and (f) the conventional EM-TV. The calibration bars indicates the linear mapping between reconstructed activity and the gray scale used.



(a) (b) (c)

(d) (e) (f)

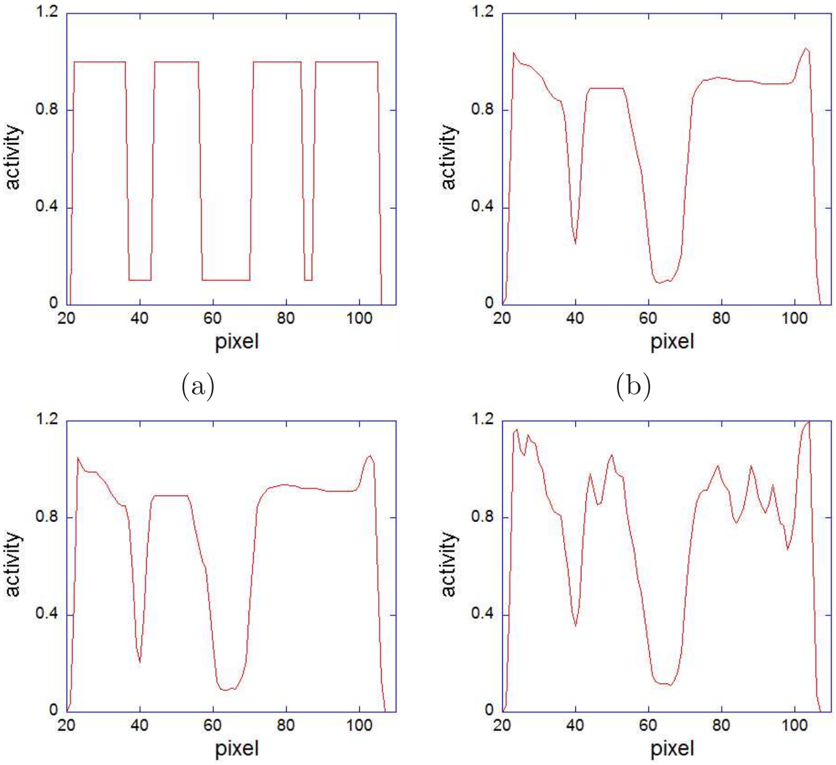
Figure 6: Transaxial cross-sections through the centers of the cold spheres (slice 49) in images reconstructed for low-noise phantom: (a) PAPA; (b) the nested EM-TV; and (c) the conventional EM-TV. Transaxial cross-sections through the centers of the cold spheres (slice 49) in images reconstructed for high-noise phantom: (d) PAPA; (e) the nested EM-TV; and (f) the conventional EM-TV. The calibration bars indicates the linear mapping between reconstructed activity and the gray scale used.



(a) (b)

(c) (d)

Figure 7: One-pixel wide line profiles through the centers of hot spheres with 7, 14 and 3-pixel radii in transaxial cross-sections (slice 17) of: (a) the high-noise phantom; (b) image reconstructed using PAPA from the high-noise phantom data; (c) image reconstructed using the nested EM-TV from the high-noise phantom data; and (d) image reconstructed using conventional EM-TV from the high-noise phantom data.



(a) (b)

(c) (d)

Figure 8: One-pixel wide line profiles through the centers of cold spheres with 7, 14 and 3-pixel radii in transaxial cross-sections (slice 49) of: (a) the high-noise phantom; (b) image reconstructed using PAPA from the high-noise phantom data; (c) image reconstructed using the nested EM-TV from the high-noise phantom data; and (d) image reconstructed using conventional EM-TV from the high-noise phantom data.