A Fast Convergent Ordered-Subsets Algorithm With Subiteration-Dependent Preconditioners for PET Image Reconstruction

Jianfeng Guo, C. Ross Schmidtlein, *Member, IEEE*, Andrzej Krol, *Senior Member, IEEE*, Si Li, Yizun Lin, Sangtae Ahn, *Member, IEEE*, Charles Stearns, *Fellow, IEEE*, and Yuesheng Xu

***Abstract*— We investigated the imaging performance of a fast convergent ordered-subsets algorithm with subiteration-dependent preconditioners (SDPs) for positron emission tomography (PET) image reconstruction. In particular, we considered the use of SDP with the block sequential regularized expectation maximization (BSREM) approach with the relative difference prior (RDP) regularizer due to its prior clinical adaptation by vendors. Because the RDP regularization promotes smoothness in the reconstructed image, the directions of the gradients in smooth areas more accurately point toward the objective function's minimizer than those in variable areas. Motivated by this observation, two SDPs have been designed to increase iteration step-sizes in the smooth areas and reduce iteration step-sizes in the variable areas relative to a conventional expectation maximization preconditioner. The momentum technique used for convergence acceleration can be viewed as a special case of SDP. We have proved the global convergence of SDP-BSREM algorithms by assuming certain characteristics of the preconditioner. By means of numerical experiments using both simulated and clinical PET data, we have shown that the SDP-BSREM algorithms substantially improve the convergence rate, as compared to conventional BSREM and a vendor's implementation as Q.Clear. Specifically, SDP-BSREM algorithms converge 35%-50% faster in reaching the same objective function value than conventional BSREM and commercial Q.Clear algorithms. Moreover, we showed in phantoms with hot, cold and background regions that the SDP-BSREM algorithms approached the values of a highly converged reference image faster than conventional BSREM and commercial Q.Clear algorithms.**

***Index Terms*— Image reconstruction, ordered-subsets, positron emission tomography, preconditioner, relative difference prior.**

Manuscript received 18 April 2022; accepted 6 June 2022. Date of publication 9 June 2022; date of current version 27 October 2022. The work of Jianfeng Guo was supported in part by the Special Project on High-Performance Computing under the National Key Research and Development Program under Grant 2016YFB0200602 and in part by the National Natural Science Foundation of China under Grant 11771464. The work of C. Ross Schmidtlein was supported in part by the Memorial Sloan Kettering Cancer Center's Support Grant from the National Cancer Institute under Grant P30 CA008748 and in part by the National Cancer Institute of the National Institutes of Health under Award R21CA263876. The work of Andrzej Krol was supported in part by the National Cancer Institute of the National Institutes of Health under Award R21CA263876. The work of Si Li was supported in part by the Natural Science Foundation of Guangdong Province under Grant 2022A1515012379, in part by the Opening Project of Guangdong Province Key Laboratory of Computational Science at Sun Yat-sen University under Grant 2021007, and in part by the National Natural Science Foundation of China under Grant 11771464. The work of Yizun Lin was supported in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515110541, in part by the Fundamental Research Funds for the Central Universities of China under Grant 21620352, and in part by the Opening Project of Guangdong Province Key Laboratory of Computational Science at Sun Yat-sen University under Grant 2021006.The work of Yuesheng Xu was supported in part by the U.S. National Science Foundation under Grant DMS-1912958 and in part by the National Cancer Institute of the National Institutes of Health under Award R21CA263876. *(Corresponding author: Yuesheng Xu.)*

Jianfeng Guo is with the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510275, China (e-mail: guojf3@mail2.sysu.edu.cn).

C. Ross Schmidtlein is with the Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY 10065 USA (e-mail: schmidtr@mskcc.org).

Andrzej Krol is with the Departments of Radiology and Pharmacology, SUNY Upstate Medical University, Syracuse, NY 13210 USA (e-mail: krola@upstate.edu).

Si Li is with the School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China (e-mail: sili@gdut.edu.cn).

Yizun Lin is with the Department of Mathematics, College of Information Science and Technology, Jinan University, Guangzhou 510632, China (e-mail: linyizun@jnu.edu.cn).

Sangtae Ahn is with GE Research, Niskayuna, NY 12309 USA (e-mail: ahns@ge.com).

Charles Stearns is with GE Healthcare, Waukesha, WI 53188 USA (e-mail: stearnsc@alum.mit.edu).

Yuesheng Xu is with the Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529 USA (e-mail: y1xu@odu.edu).

Digital Object Identifier 10.1109/TMI.2022.3181813

I. Introduction

POSITRON emission tomography (PET) data are inherently count limited due to health consideration, basic physical processes, and patient tolerance. Moreover, these data must be reconstructed into images within a few minutes of acquisition. This creates a challenging situation in which vendors strive to produce high quality images in a clinically viable time frame. In this study, we introduce a method for accelerating the reconstruction of high quality PET images.

Over last 20 years, the non-penalized maximum-likelihood (ML) statistical approaches have become a preferred model for the reconstruction of PET [1], [2]. However, when iterated to full convergence, ML methods produce extremely noisy images, and are sensitive to small statistical perturbations

in the data. Hence, these methods are seldom run to full convergence and iterations are stopped before fitting noise becomes unacceptable at the expense of excessive blur in the reconstructed images. It has been demonstrated that applications of penalized likelihood (PL) models that include a data fidelity term (Kullback-Leibler divergence) and a regularization term leads to improved quantification and better noise suppression, as compared to non-penalized reconstructions [3]. To reduce the computational expense, ordered-subsets expectation maximization (OSEM) algorithms proposed by Hudson and Larkin are widely used in unregularized PET image reconstruction [4]. However, OSEM is unsuitable for regularized image reconstruction leading to the development of relaxation [5], [6] and the block sequential regularized expectation maximization (BSREM) algorithm [7].

Initially, quadratic penalties were explored [8], [9], but they had resulted in over-smoothed edges and loss of details in the reconstructed images. Later, a number of other penalties were developed to address these problems but they often had undesirable properties including nonsmooth [10], [11], noncovex [12], or requiring additional hyper-parameters [13]. An example of such an approach is the total variation penalty that is able to preserve sharp boundaries between low-variability regions [10], [14]. Thus, ability to deal with non-smooth priors became an urgent issue, However, only a few reconstruction algorithms have been able to combine the Poisson noise model and non-smooth priors [15]–[17].

In an alternative approach, Nuyts *et al.* [18] introduced the relative difference prior (RDP) that preserved high spatial frequencies in reconstructed images while still being smooth and convex. This RDP was adopted by General Electric (GE) Healthcare as the penalty term in their PL PET reconstruction model. The penalty term is controlled by a single user-defined parameter called beta. The GE Healthcare introduced a modified BSREM algorithm [8] to solve the PL model in their commercial clinical software, called Q.Clear, that is currently available on GE PET/CT scanners [3]. Other interesting methods, suitable for optimization with smooth penalties, include the optimization transfer descent algorithm (OTDA) [19], [20] and the preconditioned limited-memory Broyden-Fletcher-Goldfarb-Shanno with boundary constraints algorithm (L-BFGS-B-PC) [9]. While these algorithms converge very rapidly, they represent a substantial departure from the BSREM algorithm complicating their implementation.

The choice of preconditioners in the algorithm is well known to strongly affect the convergence rate [9], [16], [21], [22]. The widely used preconditioners have been designed based on the EM matrix [16], [21]–[23] or the Hessian matrix [9]. As part of the BSREM convergence proof, Ahn and Fessler [8] presented the subiteration-independent preconditioner, which can be viewed as a uniform operator of the image for all subiteration. However, a subiteration-independent preconditioner is overly restrictive and may result in a slower convergence rate. We believe that a well-designed subiteration-dependent preconditioner (SDP) will accelerate the algorithm convergence.

In the present study, we propose a subiteration-dependent preconditioned BSREM (SDP-BSREM) for the RDP regularized PET image reconstruction. We prove that it is convergent under certain assumptions imposed on the preconditioner. According to the smoothness-promoting property of the RDP regularization in the reconstructed image, the directions of the gradients in smooth areas more accurately point toward the objective function's minimizer than those in variable areas. Inspired by this observation, we propose two SDPs satisfying the assumptions needed for the convergence proof. We note that the momentum technique is a special case of SDP. We have used the numerical gradient of the image to measure its smoothness. These two SDPs achieve larger step-sizes in the smooth areas of the image and smaller step-sizes in the variable areas of the image. The proposed algorithms have been compared with BSREM for simulations and with the Q.Clear method with data acquired from a GE PET/CT. In simulations, two numerical phantoms were used. In the clinical comparisons, data from a whole-body PET patient and an American College of Radiology (ACR) quality assurance phantom (Esser Flangeless PET phantom) were used both with and without time-of-flight data.

This paper is organized in five sections. In section II, we first describe the RDP regularized PET image reconstruction model and the modified BSREM algorithm and then develop our new SDP-BSREM algorithm. In section III, proofs for convergence of SDP-BSREM are provided with and without an interior assumption and four SDPs satisfying the convergence conditions are presented as well. Comprehensive comparisons of the results obtained for simulated and clinical data obtained by means of our proposed SDP-BSREM methods versus BSREM and Q.Clear are provided in section IV. The conclusions are presented in section V. Two appendices with additional details are also provided.

II. Methodology

In this section, we develop the SDP-BSREM algorithm for solving the RDP regularized PET image reconstruction model.

A. RDP Regularized PET Image Reconstruction Model

We denote by $\mathbb{R}_+$ the set of all nonnegative real numbers, by $\mathbb{R}_{++}$ the set of all positive real numbers, by $\mathbb{N}$ the set of positive integers, and by $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. For $p, q \in \mathbb{N}$, we let $\boldsymbol{A} \in \mathbb{R}_+^{p\times q}$ denote the PET system matrix whose entries are the probability of detection of the positron annihilation gamma photon pairs emitted from a particular voxel containing PET radiotracer, and let $\boldsymbol{\gamma} \in \mathbb{R}_{++}^p$ denote the mean value of the background events produced by random and scatter coincidences. The relation of the radiotracer distribution $\boldsymbol{f} \in \mathbb{R}_+^q$ within a patient with the projection data $\boldsymbol{g} \in \mathbb{R}_+^p$ acquired by a PET scanner is described by the Poisson model

$$\boldsymbol{g} = \text{Poisson}(\boldsymbol{A}\boldsymbol{f} + \boldsymbol{\gamma}), \tag{1}$$

where $\text{Poisson}(\boldsymbol{x})$ denotes a Poisson-distributed random vector with mean $\boldsymbol{x}$.

Model (1) may be solved by minimizing the fidelity term

$$F(\boldsymbol{f}) := \langle \boldsymbol{A}\boldsymbol{f}, \boldsymbol{1}_p \rangle - \langle \ln(\boldsymbol{A}\boldsymbol{f} + \boldsymbol{\gamma}), \boldsymbol{g} \rangle, \tag{2}$$

where $\mathbf{1}_p \in \mathbb{R}^p$ denotes the vector with all components 1, $\ln x := [\ln x_1, \ln x_2, \ldots, \ln x_n]^\top$ is the logarithmic function at a vector $x \in \mathbb{R}^n_{++}$ and x_i is the i-th component of x, and $\langle x, y\rangle := \sum_{i=1}^n x_i y_i$ denotes the inner product of $x, y \in \mathbb{R}^n$. It is well-known that model (2) is ill-posed [24] and its solutions may result in over-fitting in reconstructed images. Regularization is often used to avoid the over-fitting problem. A commonly used regularized PET image reconstruction model has the following form:

$$\arg\min_{f\in\mathbb{R}^q_+} \Phi(f), \tag{3}$$

where

$$\Phi(f) := F(f) + \beta R(f), \tag{4}$$

with $\beta \in \mathbb{R}_+$ being the regularization parameter and $R(f)$ representing the regularization term. In this study, we will consider the RDP [18] regularization term that is given by

$$R(f) := \sum_{j=1}^{q} \sum_{k\in N_j} \frac{(f_j - f_k)^2}{(f_j + f_k) + \gamma_R \left|f_j - f_k\right| + \epsilon}, \tag{5}$$

where $\gamma_R \in \mathbb{R}_+$ controls the degree of edge preservation, and N_j is the neighborhood of pixel j.

In model (5), a small constant $\epsilon > 0$ is added to the denominator to avoid singularities when both f_j and f_k are equal to zero. By its definition, RDP is a function of both differences of neighboring pixels and their sums. The inclusion of the sums term makes RDP differ from conventional regularization terms and causes the regularizer to be activity-dependent. We note that the function $\Phi(f)$ is twice differentiable since both $F(f)$ and $R(f)$ are twice differentiable [16], [18]. The inclusion of a small constant ϵ in the denominator of RDP provides the objective function Φ with two useful properties (the proofs are provided in appendix I): (i) it is strictly convex under an assumption that $A^\top g$ is a nonzero vector; and (ii) it has a Lipschitz continuous gradient on $\mathbb{R}^q_+$.

B. Modified BSREM Algorithm

A modified BSREM [8] was adopted by GE Healthcare as the optimizer in the Q.Clear method [3] for solving the model (3). Here we describe and review the modified BSREM.

Minimization problem (3) is often solved by the gradient descent method. However, computing the whole gradient $\nabla\Phi$ is computationally expensive. To alleviate this issue, the ordered-subsets (OS) algorithm was developed to accelerate its convergence [4]–[7]. For $n \in \mathbb{N}$, let $\mathbb{N}_n := \{1, 2, \ldots, n\}$. For $M \in \mathbb{N}$, let $\mathcal{I} := \{I_i : i \in \mathbb{N}_M\}$ be a collection of disjoint subsets of $\mathbb{N}_p$ such that $\bigcup_{i=1}^M I_i = \mathbb{N}_p$. The partition $\mathcal{I}$ is chosen as in [8]. According to the partition $\mathcal{I}$, we partition the system matrix A into M row sub-matrices A_i, and g and γ into M sub-vectors g_i and γ_i, respectively, for $i \in \mathbb{N}_M$. We use $|\Omega|$ to denote the cardinality of set Ω. For $i \in \mathbb{N}_M$, we define

$$\Phi_i(f) := \langle A_i f, \mathbf{1}_{|I_i|}\rangle - \langle \ln\left(A_i f + \gamma_i\right), g_i\rangle + \frac{\beta}{M} R(f). \tag{6}$$

It follows that $\Phi(f) = \sum_{i=1}^M \Phi_i(f)$. An OS algorithm computes only one $\nabla\Phi_i$ at each subiteration step.

A subiteration-independent preconditioned OS algorithm was proposed in [8]. The preconditioner is designed by using an upper bound of the solution set of minimization problem (3). It was proved in [8] that for any projection data g, there exists a constant $U > 0$ such that the solution set $\mathcal{S}^*$ of minimization problem (3) is contained in the bounded set

$$\mathcal{B} := \left\{f : f \in \mathbb{R}^q_+,\ 0 \leqslant f_j \leqslant U,\ j \in \mathbb{N}_q\right\}. \tag{7}$$

That is, $\mathcal{S}^* \subset \mathcal{B}$. For $f \in \mathcal{B}$, a subiteration-independent diagonal preconditioner $S(f)$ is defined as

$$S(f)_{jj} := \begin{cases} f_j/p_j & \text{if } 0 \leq f_j < U/2, \\ (U - f_j)/(p_j) & \text{if } U/2 \leq f_j \leq U, \end{cases} \tag{8}$$

where p_j are defined by

$$p_j := \begin{cases} (A^\top \mathbf{1}_p)_j/M & \text{if } (A^\top \mathbf{1}_p)_j > 0, \\ 1/M & \text{if } (A^\top \mathbf{1}_p)_j = 0, \end{cases} \quad \text{for } j \in \mathbb{N}_q. \tag{9}$$

Note that the preconditioner S is uniform for all iterations.

For a small $t \in (0, U)$ and $f \in \mathbb{R}^q$, an operator $P_t : \mathbb{R}^q \to \mathcal{B}$ is defined by

$$P_t(f)_j := \begin{cases} t & \text{if } f_j \leqslant 0, \\ U - t & \text{if } f_j \geqslant U, \\ f_j & \text{otherwise.} \end{cases} \tag{10}$$

Using operator P_t, the modified BSREM algorithm [8] may be described as for $k \in \mathbb{N}_0$, $i \in \mathbb{N}_M$,

$$\begin{cases} \tilde{f}^{k,i} = f^{k,i-1} - \lambda_k S(f^{k,i-1})\nabla\Phi_i(f^{k,i-1}), \\ f^{k,i} = P_t(\tilde{f}^{k,i}), \end{cases} \tag{11}$$

with $f^{k,0} := f^k, f^{k+1} := f^{k,M}$, where $\lambda_k > 0$ is the relaxation parameter. For simplicity of notation, we will refer to the modified form of BSREM simply as BSREM.

C. BSREM With Subiteration-Dependent Preconditioners

In this subsection we propose subiteration-dependent preconditioners (SDPs). To motivate them, we review the momentum approach. The momentum is an acceleration technique widely used in optimization [25]–[27]. The Nesterov momentum [25] has been combined with OS by Kim *et al.* [28] for CT image reconstruction. Recently, Lin *et al.* [22] successfully applied a different form of momentum to PET image reconstruction. However, no explicit convergence proof has been provided for the OS combined momentum methods. Instead, Kim *et al.* proved that the expectation of the successive steps converged, while Lin *et al.* proved convergence for the non-OS method. The momentum technique used in [22] can be described as follows: for $k \in \mathbb{N}_0$, $i \in \mathbb{N}_M$,

$$\begin{cases} \tilde{f}^{k,i} = \max[f^{k,i-1} - \lambda_k S(f^{k,i-1})\nabla\Phi_i(f^{k,i-1}), 0], \\ f^{k,i} = (1 - \alpha_{k,i})f^{k,i-1} + \alpha_{k,i}\tilde{f}^{k,i}, \end{cases} \tag{12}$$

where $\alpha_{k,i} > 1$ is the momentum sequence. Under an assumption that $\tilde{f}^{k,i}$ are non-negative, one can obtain

$$f^{k,i} = f^{k,i-1} - \lambda_k \alpha_{k,i} S(f^{k,i-1})\nabla\Phi_i(f^{k,i-1}). \tag{13}$$

TABLE I
SDP-BSREM ALGORITHM

Preparation: $\boldsymbol{f}^0, M, T, P_t$ is defined in (10)
for $k = 0, 1, 2, \ldots, T$
  $\boldsymbol{f}^{k,0} = \boldsymbol{f}^k$
  for $i = 1, 2, \ldots, M$
    $\tilde{\boldsymbol{f}}^{k,i} = \boldsymbol{f}^{k,i-1} - \lambda_k \boldsymbol{S}^{k,i}(\boldsymbol{f}^{k,i-1})\nabla\Phi_i(\boldsymbol{f}^{k,i-1})$
    $\boldsymbol{f}^{k,i} = P_t(\tilde{\boldsymbol{f}}^{k,i})$
  end
  $\boldsymbol{f}^{k+1} = \boldsymbol{f}^{k,M}$
end

By letting $S^{k,i}(\boldsymbol{f}) := \alpha_{k,i}S(\boldsymbol{f})$, we may reinterpret $S^{k,i}$ as an SDP and (13) as a BSREM algorithm with $S^{k,i}$. Inspired by this, we introduce SDPs by setting $S^{k,i}(\boldsymbol{f}) := \text{diag}(\boldsymbol{\alpha}^{k,i})S(\boldsymbol{f})$, where $\boldsymbol{\alpha}^{k,i}$ is a positive vector sequence and $\text{diag}(\boldsymbol{y})$ denotes the diagonal matrix with the diagonal entries being the components of the vector $\boldsymbol{y}$. Using the same notation as for BSREM above, we have arrived at the SDP-BSREM algorithm for solving model (3) given in Table I.

Bearing in mind the momentum concept defined in [22], it is clear that the iteration sequence provided in (13) is a special case of our proposed SDP-BSREM algorithm with $\boldsymbol{\alpha}^{k,i} := \alpha_{k,i}\mathbf{1}_q$. For this reason, we expect that our proposed SDP-BSREM algorithm setting will allow us to choose an SDP to yield the convergence acceleration. We will evaluate its performance by means of numerical experiments to be presented in section IV.

III. CONVERGENCE OF SDP-BSREM ALGORITHM

In this section we present convergence properties of the SDP-BSREM algorithm. We also describe four specific SDPs that satisfy the convergence condition. To this end, we assume that the objective function Φ satisfies the hypothesis:

(i). Φ has a unique minimizer on $\mathcal{B}$;
(ii). Φ is convex and twice differentiable on $\mathcal{B}$;
(iii). $\nabla\Phi_i$ are Lipschitz continuous on $\mathcal{B}$ for all $i \in \mathbb{N}_M$.

The use of SDPs results in scaled subset gradients with their sum inconsistent with the scaled full gradient. It complicates the convergence proof and requires additional assumptions on the preconditioner to make the inconsistencies asymptotically approach zero. For a general SDP, our proposed SDP-BSREM algorithm may not converge [8]. Nevertheless, by imposing certain assumptions on $S^{k,i}$, convergence to the desired optimum point can be ensured. For a SDP $S^{k,i}(\boldsymbol{f}) = \text{diag}(\boldsymbol{\alpha}^{k,i})S(\boldsymbol{f})$ with $S(\boldsymbol{f})$ defined in (8), the required additional assumptions are as follows:

(iv). The relaxation sequence satisfies $\sum_{k=0}^{\infty}\lambda_k = \infty$ and $\sum_{k=0}^{\infty}\lambda_k^2 < \infty$;
(v). There exists a positive vector $\boldsymbol{\alpha}$ such that $\lim_{k\to\infty}\boldsymbol{\alpha}^{k,i} = \boldsymbol{\alpha}$ for all $i \in \mathbb{N}_M$;
(vi). The vector series $\sum_{k=0}^{\infty}\lambda_k(\boldsymbol{\alpha} - \boldsymbol{\alpha}^{k,i})$ converge for all $i \in \mathbb{N}_M$.

Note that condition (iv) was imposed in [8], [29] for convergence proofs for relaxed OS algorithms. Conditions (v) and (vi) were imposed to overcome difficulties caused by the use of different preconditioners in different subiterations.

We now state a lemma regarding the Lipschitz continuity properties of SDPs. The proof is included in Appendix II.

Lemma 1: If conditions (iii) and (v) are satisfied, then $S^{k,i}(\boldsymbol{f})\nabla\Phi_i(\boldsymbol{f})$ are uniformly bounded and Lipschitz continuous on $\mathcal{B}$ with Lipschitz constants bounded above by a uniform constant, for all $k \in \mathbb{N}_0, i \in \mathbb{N}_M$.

The inclusion of the operator P_t in the SDP-BSREM algorithm complicates the convergence proof. Here, we present our approach in dealing with this difficulty. Let $\text{int}\,\mathcal{B}$ denote the interior of $\mathcal{B}$. We prove the convergence of SDP-BSREM in two steps. We first prove it with the interior assumption $\tilde{\boldsymbol{f}}^{k,i} \in \text{int}\,\mathcal{B}$ for all $k \in \mathbb{N}_0, i \in \mathbb{N}_M$, and then prove it by showing that the interior assumption holds true under certain conditions.

We now proceed the first step. If $\tilde{\boldsymbol{f}}^{k,i} \in \text{int}\,\mathcal{B}$ for all $k \in \mathbb{N}_0, i \in \mathbb{N}_M$, then $\boldsymbol{f}^{k,i} = P_t(\tilde{\boldsymbol{f}}^{k,i}) = \tilde{\boldsymbol{f}}^{k,i}$ and the iteration scheme can be formulated as

$$\boldsymbol{f}^{k,i} = \boldsymbol{f}^k - \lambda_k \sum_{l=1}^{i} \boldsymbol{S}^{k,l}(\boldsymbol{f}^{k,l-1})\nabla\Phi_t(\boldsymbol{f}^{k,l-1}). \tag{14}$$

We first establish a technical lemma.

Lemma 2: Suppose conditions (iii) and (v) are satisfied. If $\lim_{k\to\infty}\lambda_k = 0$, and $\tilde{\boldsymbol{f}}^{k,i} \in \text{int}\,\mathcal{B}$, for all $k \in \mathbb{N}_0$, $i \in \mathbb{N}_M$, then $\lim_{k\to\infty}(\boldsymbol{f}^{k,i} - \boldsymbol{f}^k) = 0$, for all $i \in \mathbb{N}_M$.

Proof: By Lemma 1, $S^{k,i}(\boldsymbol{f})\nabla\Phi_i(\boldsymbol{f})$, $k \in \mathbb{N}_0, i \in \mathbb{N}_M$, are uniformly bounded on $\mathcal{B}$. This combined with $\lim_{k\to\infty}\lambda_k = 0$ and (14) yields the desired result. ∎

Let $\delta^{k,i} := \boldsymbol{\alpha} - \boldsymbol{\alpha}^{k,i}$, $\delta_k := \max_{i\in\mathbb{N}_M, j\in\mathbb{N}_q}|\delta_j^{k,i}|$. We state a technical lemma whose proof is included in Appendix II.

Lemma 3: Suppose conditions (iii)-(vi) are satisfied and $\tilde{\boldsymbol{f}}^{k,i} \in \text{int}\,\mathcal{B}$ for all $k \in \mathbb{N}_0, i \in \mathbb{N}_M$. If $f_j^{k,i-1} \in (0, U/2)$ for all $i \in \mathbb{N}_M$, then

$$f_j^k - f_j^{k+1} = \lambda_k f_j^k\left[\frac{\alpha_j}{p_j}\frac{\partial\Phi(\boldsymbol{f}^k)}{\partial f_j} + \mathcal{O}(\delta_k) + \mathcal{O}(\lambda_k)\right]. \tag{15}$$

If $f_j^{k,i-1} \in [U/2, U)$ for all $i \in \mathbb{N}_M$, then

$$f_j^k - f_j^{k+1} = \lambda_k(U - f_j^k)\left[\frac{\alpha_j}{p_j}\frac{\partial\Phi(\boldsymbol{f}^k)}{\partial f_j} + \mathcal{O}(\delta_k) + \mathcal{O}(\lambda_k)\right]. \tag{16}$$

We recall that a cluster point of a sequence $\boldsymbol{f}^k$ is defined as the limit of a convergent subsequence of $\boldsymbol{f}^k$ and state a lemma whose proof is included in Appendix II.

Lemma 4: If conditions (i)-(vi) are satisfied and $\tilde{\boldsymbol{f}}^{k,i} \in \text{int}\,\mathcal{B}$ for all $k \in \mathbb{N}_0, i \in \mathbb{N}_M$, then (a) $\Phi(\boldsymbol{f}^k)$ converges in $\mathbb{R}$, (b) there exists a cluster point $\boldsymbol{f}^*$ of $\boldsymbol{f}^k$ with $S(\boldsymbol{f}^*)\nabla\Phi(\boldsymbol{f}^*) = \mathbf{0}$, and (c) such a cluster point $\boldsymbol{f}^*$ described in (b) is a global minimizer of Φ over $\mathcal{B}$.

We are ready to prove convergence of $\boldsymbol{f}^k$ and $\boldsymbol{f}^{k,i}$ with the interior assumption.

Proposition 5: If conditions (i)-(vi) are satisfied and $\tilde{\boldsymbol{f}}^{k,i} \in \text{int}\,\mathcal{B}$ for all $k \in \mathbb{N}_0, i \in \mathbb{N}_M$, then both $\boldsymbol{f}^k$ and $\boldsymbol{f}^{k,i}$ converge to the global minimizer of Φ on $\mathcal{B}$.

Proof: According to Lemma 4 (c), f^* is a global minimizer of Φ over $\mathcal{B}$. Suppose there exists another cluster point $f^{**} \neq f^*$. By Lemma 4 (a), $\Phi(f^k)$ converges in $\mathbb{R}$, which implies that $\Phi(f^*) = \Phi(f^{**})$. Then f^{**} is also a minimizer of $\Phi(f)$, which is a contradiction since $\Phi(f)$ has a unique minimizer on $\mathcal{B}$. Then we obtain $\lim_{k\to\infty} f^k = \arg\min_{f\in\mathcal{B}} \Phi(f)$. The convergence of $f^{k,i}$ follows from Lemma 2 and the convergence of f^k. ∎

We have shown that condition $\tilde{f}^{k,i} \in \operatorname{int}\mathcal{B}$ for all $k \in \mathbb{N}_0, i \in \mathbb{N}_M$ is sufficient for the convergence of SDP-BSREM algorithm. Next, we prove the convergence of SDP-BSREM without the interior assumption. The proof of convergence is completed by proving $\tilde{f}^{k,i} \in \operatorname{int}\mathcal{B}$ for all $i \in \mathbb{N}_M$ and $k > K$ for some $K > 0$. We now state a lemma to prove it.

Lemma 6: Suppose condition (iii) is satisfied. If $\boldsymbol{\alpha}^{k,i}$ is bounded and $\lim_{k\to\infty} \lambda_k = 0$, then $\tilde{f}^{k,i} \in \operatorname{int}\mathcal{B}$ for all $i \in \mathbb{N}_M$ and $k > K$ for some $K > 0$.

Proof: It suffices to prove that $\tilde{f}_j^{k,i} \in (0, U)$ for all $i \in \mathbb{N}_M, j \in \mathbb{N}_q, k > K$, for some $K > 0$. By condition (iii), $(\partial/\partial f_j)\Phi_i(f)$ is bounded over $\mathcal{B}$ for all $i \in \mathbb{N}_M, j \in \mathbb{N}_q$. Combining this with the boundedness of $\boldsymbol{\alpha}^{k,i}$, there exists $c_1 > 0$ such that $|\alpha_j^{k,i}/p_j(\partial/\partial f_j)\Phi_i(f)| \leqslant c_1$, for all $i \in \mathbb{N}_M, j \in \mathbb{N}_q, k \in \mathbb{N}_0$, and all $f \in \mathcal{B}$. Because $\lim_{k\to\infty} \lambda_k = 0$, there exists $K > 0$ such that $|\lambda_k| < 1/c_1$ for all $k > K$, so that $|\lambda_k \alpha_j^{k,i}/p_j(\partial/\partial f_j)\Phi_i(f^{k,i-1})| < 1$. Hence, for $k > K, i \in \mathbb{N}_M$, if $f_j^{k,i-1} \in (0, U/2)$, the preconditioner $S^{k,i}(f^{k,i-1})_{jj} = \alpha_j^{k,i} f_j^{k,i-1}/p_j$ gives rise to $\tilde{f}_j^{k,i} = f_j^{k,i-1}[1 - \lambda_k \alpha_j^{k,i}/p_j(\partial/\partial f_j)\Phi_i(f^{k,i-1})]$, from which we can show that $\tilde{f}_j^{k,i} \in (0, U)$. Likewise, if $f_j^{k,i-1} \in [U/2, U)$, the preconditioner $S(f^{k,i-1})_{jj} = \alpha_j^{k,i}(U - f_j^{k,i-1})/p_j$ gives that $(U - \tilde{f}_j^{k,i}) = (U - f_j^{k,i-1})[1 + \lambda_k \alpha_j^{k,i}/p_j(\partial/\partial f_j)\Phi_i(f^{k,i-1})]$, from which we can show that $\tilde{f}_j^{k,i} \in (0, U)$. ∎

We now arrive at the following theorem for the convergence of SDP-BSREM algorithm without an interior assumption.

Theorem 7: If conditions (i)-(vi) are satisfied, then $f^{k,i}$ converges to the global minimizer of Φ on $\mathcal{B}$.

Proof: We have that $\lim_{k\to\infty} \lambda_k = 0$ and $\boldsymbol{\alpha}^{k,i}$ is bounded from conditions (iv) and (v) respectively. Thus, by Lemma 6, there exists $K > 0$ such that $\tilde{f}^{k,i} \in \operatorname{int}\mathcal{B}$ for all $k > K, i \in \mathbb{N}_M$. Then the proof follows from Proposition 5. ∎

We next propose four specific SDPs which satisfy the convergence conditions (iv)-(vi). For $S^{k,i}(f) = \operatorname{diag}(\boldsymbol{\alpha}^{k,i})S(f)$, let $\boldsymbol{\alpha}^{k,i} := \alpha_{k,i}\boldsymbol{v}^{k,i}$, where $\alpha_{k,i}$ is a scalar sequence and $\boldsymbol{v}^{k,i}$ is a vector sequence to be determined. Other, potentially better, choices of $\boldsymbol{\alpha}^{k,i}$ are left as future work.

In this case, inspired by momentum techniques [22], [25], we consider the following two choices of $\alpha_{k,i}$. The first one is derived from Nesterov momentum [25]:

$$\alpha_{k,i} := 1 + (t_{k,i} - 1)/t_{k,i+1}, \tag{17}$$

where $t_{k,i+1} := (1 + \sqrt{1 + 4t_{k,i}^2})/2$, $t_{0,1} := 1$ and $t_{k+1,1} := t_{k,M+1}, k \in \mathbb{N}_0, i \in \mathbb{N}_M$. The second one has the following form:

$$\alpha_{k,i} := (\varrho(kM + i - 1) + \delta_2)/(kM + i - 1 + \delta_1), \tag{18}$$

for $k \in \mathbb{N}_0, i \in \mathbb{N}_M$, where ϱ, δ_1 and δ_2 are positive parameters. We notice that this $\alpha_{k,i}$ is an extension of the momentum proposed in [22].

The motivation for the design of $\boldsymbol{v}^{k,i}$ is presented as follows. The use of different step-sizes for different regions in the image can accelerate the convergence of the algorithm. The diagonal nonnegative definite preconditioner plays an important role in rescaling the step-sizes of the algorithm. Hence, a good preconditioner can significantly accelerate the convergence of the algorithm. We propose a type of preconditioner that is related to the regularization term that promotes smoothness in the reconstructed image. Our goal is to find the minimizer of the objective function which consists of the fidelity term and the regularization term. The fidelity term estimates the fitting quality of the reconstructed image to the data and the regularization term defined in (5) promotes smoothness in the reconstructed image. Smaller fidelity term makes the reconstructed image more consistent with the data and smaller regularization term leads to a smoother reconstructed image.

Suppose f^* is a minimizer of the objective function Φ and the iteration scheme of the algorithm converges to f^*. Let I_s and I_v be the smooth and variable areas of f^*, respectively. Suppose $\nabla_{I_s}\Phi$ and $\nabla_{I_v}\Phi$ are the subsets of $\nabla\Phi$ defined in the areas I_s and I_v, respectively. Then for the smooth areas I_s, the descent directions of the fidelity term and the regularization term are consistent whereas for the variable areas I_v, the descent directions of the fidelity term and the regularization term are inconsistent. Thus the direction of $\nabla_{I_s}\Phi$ more accurately points toward the minimizer than the direction of $\nabla_{I_v}\Phi$. Therefore, we conclude that the directions of $\nabla_{I_s}\Phi(f^{k,i})$ and $\nabla_{I_s}\Phi(f^{k,i+1})$ are more consistent than the directions of $\nabla_{I_v}\Phi(f^{k,i})$ and $\nabla_{I_v}\Phi(f^{k,i+1})$. We illustrate this conjecture (the subset gradient is used in practice) via numerical experiments (see Fig. 2). The preconditioner is designed to achieve larger iteration step-sizes in the smooth areas I_s and smaller iteration step-sizes in the variable areas I_v. The numerical gradient (the *gradient* function in Matlab) of f^* is applied to measure the smoothness degree of the image f^*. Then larger and smaller step-sizes will be used in the areas having smaller and larger numerical gradients, respectively.

Suppose $f \in \mathbb{R}_+^{q_1 q_2 \times 1}$ is a 2D image with size $q_1 \times q_2$. Let $\operatorname{mat}(f) \in \mathbb{R}_+^{q_1 \times q_2}$ be the matrix form of f. Using the *gradient* function in Matlab, we compute the gradients of $\operatorname{mat}(f)$ along the x and y directions, namely, $\operatorname{grad}_x(\operatorname{mat}(f))$ and $\operatorname{grad}_y(\operatorname{mat}(f))$. Let $\operatorname{grad}(f) := \sqrt{(\operatorname{grad}_x(\operatorname{mat}(f)))^2 + (\operatorname{grad}_y(\operatorname{mat}(f)))^2}$, where the square and square root operations are element-wise. For PET patients data, the minimizer f^* is unknown and we use $f^{k,i}$ to approximate f^*. Based on the consideration that areas with larger numerical gradients should have smaller step-sizes, we first let $\boldsymbol{\mu}^{k,i} := \max(0.01, \operatorname{grad}(f^{k,i})/\operatorname{mean}(f^{k,i}))$, where $\operatorname{mean}(f) := \sum_{j=1}^q f_j/q$ is used to normalize the $f^{k,i}$. Instead of directly letting $\boldsymbol{v}^{k,i}$ be the $1/\boldsymbol{\mu}^{k,i}$, we define a projection operator to avoid too large or too small step-sizes. For two positive numbers $v_m < v_M$, and $f \in \mathbb{R}_+^q$, a projection operator $P_{v_m}^{v_M} : \mathbb{R}^q \to \mathbb{R}^q$ is defined by $P_{v_m}^{v_M}(f)_j := \min\{v_M, \max\{f_j, v_m\}\}$.

Let $J_{k,i} := kM + i, k \in \mathbb{N}_0, i \in \mathbb{N}_M$. For $0 < \nu_1 < \nu_2$, and $0 \leqslant J_0 \leqslant J_1 := k_1 M + i_1$, the $\boldsymbol{v}^{k,i}$ is determined by

$$\boldsymbol{v}^{k,i} := \begin{cases} \mathbf{1}_q & \text{if } J_{k,i} \leqslant J_0, \\ P_{\nu_1}^{\nu_2}(\text{mean}(\boldsymbol{\mu}^{k,i})/\boldsymbol{\mu}^{k,i}) & \text{if } J_0 < J_{k,i} \leqslant J_1, \\ \boldsymbol{v}^{k_1,i_1} & \text{if } J_{k,i} > J_1. \end{cases} \quad (19)$$

For the first J_0 subiterations, the $\boldsymbol{v}^{k,i}$ is set to the identity vector since the approximation of $\boldsymbol{f}^{k,i}$ to $\boldsymbol{f}^*$ is poor. The approximation becomes better as the iteration continues, hence different step-sizes for different regions of the image are used for $J_0 < J_{k,i} \leqslant J_1$. After J_1-th subiteration $\boldsymbol{v}^{k,i}$ is then fixed due to improved approximation. The preconditioners $\boldsymbol{S}^{k,i}(\boldsymbol{f}) = \boldsymbol{S}(\boldsymbol{f})\text{diag}(\alpha_{k,i}\boldsymbol{v}^{k,i})$ are denoted by P1 and P2 depending on $\alpha_{k,i}$ defined in (17) and (18) respectively. The momentum-like preconditioners $\boldsymbol{S}^{k,i}(\boldsymbol{f}) = \boldsymbol{S}(\boldsymbol{f})\text{diag}(\alpha_{k,i}\mathbf{1}_q)$ are denoted by M1 and M2 depending on $\alpha_{k,i}$ defined in (17) and (18) respectively. Then we have the following theorem. The proof can be found in the appendix II.

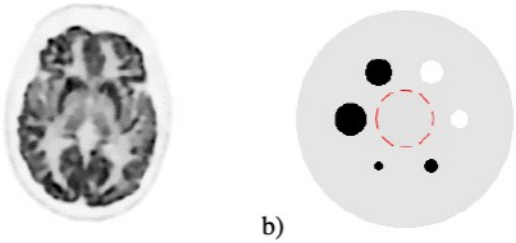
Theorem 8: The SDP-BSREM algorithm with Φ defined in (4), and with relaxation $\lambda_k := \lambda_0/(ak + 1), \lambda_0, a > 0$, and preconditioner P1, P2, M1, or M2 is convergent.

IV. Numerical Results

In this section, we present results of evaluations of the SDP-BSREM algorithms performance obtained by means of numerical experiments using both simulated and clinical PET data, in comparison with BSREM and with the clinical version of BSREM (Q.Clear, GE).

A. Simulation Setup

The algorithms were implemented using a 2D PET simulation model developed in the Matlab environment [22], [30]. The projection matrix, based on a single detector ring of a GE D710 PET/CT, was built using a ray-driven model with 32 parallel rays per detector pair. Cylindrical detector ring, consisting of 576 detectors whose width are 4 mm, was applied. The field of view (FOV) was set to 300 mm and 288 projection angles were used to reconstruct a 256×256 image with pixel size 1.17 mm×1.17mm. The true count projection data were obtained by forward projecting the phantom convolved in image space with an idealized point spread function (PSF). The PSF was a shift-invariant Gaussian function with full width at half maximum (FWHM) equal to 6.59 mm [31]. Uniform water attenuation, with attenuation coefficient 0.096 cm^{-1}, was simulated using the PET image as support. Scatter was simulated by adding highly smoothed and scaled projection of the phantom to the attenuated image sinograms. The scaling factor was equal to the estimated scatter fraction $SF := S/(T + S)$, where T and S are true and scatter counts respectively [32]. Random counts were simulated by adding a uniform distribution to the true and scatter count distributions scaled by a random fraction $RF := R/(T + S + R)$, where R is the random count [32]. The total count was defined as $TC := T + S + R$. In the simulations, it was $TC = 6.8 \times 10^6$ for high count data and $TC = 6.8 \times 10^5$ for low count data. In both cases $SF = RF = 0.25$.



a) b)

Fig. 1. Numerical phantom used in simulations. a) Brain phantom. b) Uniform phantom: uniform background (1 ROI with radius 25 pixels is shown) with 6 uniform spheres of different radii (2 cold spheres and 4 hot spheres).

The individual noise realizations were generated by adding the Poisson noise to the total count distribution. The same system matrix was used to simulate the data and to reconstruct them.

To investigate the convergence acceleration and its impact on reconstructed images fidelity, two figures-of-merit were computed: the objective function ($\Phi(\boldsymbol{f}^k)$) and the normalized root mean square difference (NRMSD). The region of interest (ROI) based NRMSD is defined by $\sqrt{\sum_{j\in\Omega}(f_j^k - f_j^\infty)^2}/\sqrt{\sum_{j\in\Omega}(f_j^\infty)^2}$, where $\boldsymbol{f}^\infty$ is the converged image at 1000 iterations by BSREM algorithm with 24 subsets in simulations, and $\boldsymbol{\Omega}$ is the ROI. In the simulations the global NRMSD is obtained by setting the ROI as the whole image.

Two 256×256 numerical 2D phantoms shown in Fig. 1 were used in simulations. The brain phantom [30] was obtained from a high quality clinical PET image. The uniform phantom consists of 4 uniform hot spheres and 2 uniform cold spheres with distinct radii: 4, 6, 8 (cold), 10 (cold), 12, 14 pixels. The contrast ratio for the cold and hot spheres are 0 : 1 and 1 : 10, respectively. All simulations were performed in a 64-bit Windows 10 OS laptop with Intel Core i7-8550U Processor at 1.80 GHz, 16 GB of DDR4 memory and 512 GB SATA SSD.

The parameter t in P_t was set to 10^{-4}. The constant ϵ was set to 10^{-12}. The regularization parameter β in model (3) was set to 0.1 and 0.8 for high and low count data respectively. In RDP regularization term, the parameter γ_R was set to 2 and 8-point neighborhood was considered. The initialization $\boldsymbol{f}^0$ was set to $\mathbf{1}_q$ to examine the setting of $\boldsymbol{v}^{k,i}$. We used the relaxation sequence defined by $\lambda_0/(ak + 1), a > 0$. In all simulation experiments, for simplicity, we empirically set $\lambda_0 = 1, J_0 = 3, J_2 = 1000$ and $\delta_1 = \delta_2$. Other parameter values, shown in Table II, were chosen based on the performance of objective function value.

B. Simulation Results

1) Comparison of Gradient Consistency: To measure the directional consistency of two vectors, we computed the angle between them. The angle between vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ is defined as $\theta(\boldsymbol{v}_1, \boldsymbol{v}_2) := \arccos(\langle \boldsymbol{v}_1, \boldsymbol{v}_2\rangle / (\|\boldsymbol{v}_1\|_2\|\boldsymbol{v}_2\|_2))$. We define the smooth areas sequence by $I_s^{k,i} := \left\{j \in \mathbb{N}_q : \text{grad}(\boldsymbol{f}^{k,i})_j < 0.01 \cdot \text{mean}(\boldsymbol{f}^{k,i})\right\}$, and the variable areas sequence by $I_v^{k,i} := \left\{j \in \mathbb{N}_q : \text{grad}(\boldsymbol{f}^{k,i})_j > 0.2 \cdot \text{mean}(\boldsymbol{f}^{k,i})\right\}$. In order to estimate the consistency of

TABLE II
ALGORITHMIC PARAMETERS FOR 2D SIMULATION RECONSTRUCTION

Parameters	Brain phantom	Uniform phantom
BSREM(12)	high count: $a = 1/400$ low count: $a = 1/18$	-
SDP-P1(12)	high count: $a = 1/13$, $\nu_1 = 1.6, \nu_2 = 2.4$ low count: $a = 0.5$, $\nu_1 = 1.6, \nu_2 = 2.4$	-
SDP-P2(12)	high count: $a = 1/5, \varrho = 5$, $\delta_1 = 5, \nu_1 = 0.8, \nu_2 = 2.2$ low count: $a = 1.3, \varrho = 7.5$, $\delta_1 = 5, \nu_1 = 1.3, \nu_2 = 2.1$	-
BSREM(24)	high count: $a = 1/35$ low count: $a = 1/5$	high count: $a = 1/35$
SDP-P1(24)	high count: $a = 0.35$, $\nu_1 = 1.6, \nu_2 = 2.4$ low count: $a = 1.3$, $\nu_1 = 1.4, \nu_2 = 2.5$	high count: $a = 0.5$, $\nu_1 = 1.8, \nu_2 = 2.5$
SDP-P2(24)	high count: $a = 0.45, \varrho = 4$, $\delta_1 = 3, \nu_1 = 0.8, \nu_2 = 1.8$ low count: $a = 1.4, \varrho = 2.2$, $\delta_1 = 1, \nu_1 = 1.3, \nu_2 = 2.4$	high count: $a = 0.7$, $\varrho = 3, \delta_1 = 7$, $\nu_1 = 1.4, \nu_2 = 2.3$

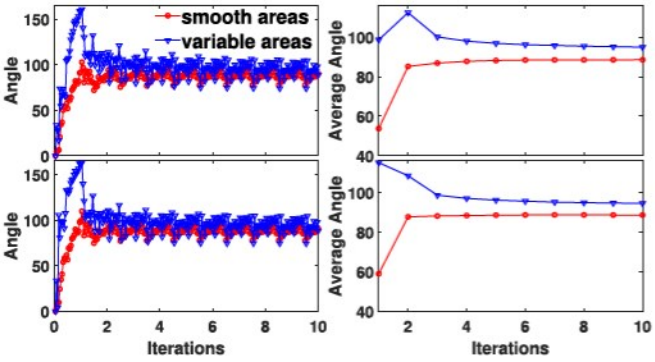


Fig. 2. Angle (left) and average angle (right) between the gradients of the successive subiterations vs. iterations projected in the smooth areas and variable areas in the reconstructed images, respectively, for the brain phantom with high count data. Top row: SDP-P1 (24). Bottom row: SDP-P2(24).

$\nabla_{I_s^{k,i}} \Phi_{i-1}(f^{k,i-1})$ and $\nabla_{I_s^{k,i}} \Phi_i(f^{k,i})$, we computed the angle between them. For smooth areas, we define the angle sequence by $\theta_{k,i} := \theta(\nabla_{I_s^{k,i}} \Phi_{i-1}(f^{k,i-1}), \nabla_{I_s^{k,i}} \Phi_i(f^{k,i}))$, and the average angle in each iteration by $\theta_k := \sum_{i=1}^{M} \theta_{k,i}/M$, where $\Phi_0 := \Phi_M$. Similarly, for variable areas, we computed $\tilde{\theta}_{k,i} := \theta(\nabla_{I_v^{k,i}} \Phi_{i-1}(f^{k,i-1}), \nabla_{I_v^{k,i}} \Phi_i(f^{k,i}))$ and $\tilde{\theta}_k := \frac{1}{M} \sum_{i=1}^{M} \tilde{\theta}_{k,i}$. In Fig. 2, we observed that for SDP-P1 and SDP-P2, the angle and average angle in smooth areas were smaller than those in the areas with more variability. This is consistent with our conjecture that the directions of gradients in the smooth areas more accurately point toward the minimizer than those in the variable areas. Hence, larger step-sizes in the smooth areas and smaller step-sizes in the variable areas are reasonable.

2) Comparison of Preconditioners: In order to reveal the improvement due to the application of a preconditioner, as compared to the use of a momentum, we compared SDP-BSREM algorithm with four different preconditioners: P1, P2, M1, and M2, where M1 and M2 are surrogates of momenta. The parameter a in SDP-M1(12) and SDP-M1(24) was set to $1/50$ and $1/6$, respectively. And the

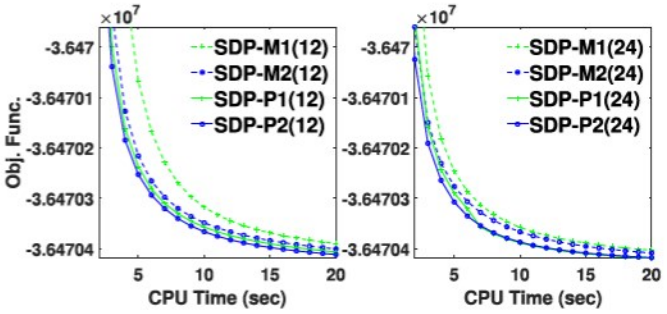


Fig. 3. Comparison of performance of preconditioners investigated in this study. Objective function vs. elapsed CPU time in reconstructions performed with SDP-BSREM algorithm with four preconditioners: M1, M2, P1, and P2, and with 12 subsets (left) and 24 subsets (right), respectively, for the brain phantom with high count data. Preconditioners P1 and P2 were generalized from M1 and M2, respectively.

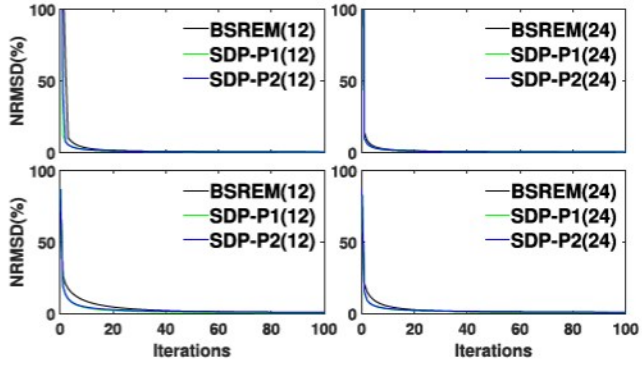


Fig. 4. Global NRMSD vs. Iterations in reconstructions performed with different algorithms: BSREM(12), SDP-P1(12), SDP-P2(12), BSREM(24), SDP-P1(24), SDP-P2(24) for the brain phantom with low (top row) and high (bottom row) count data, respectively.

parameters a, ϱ, δ_1 in SDP-M2(12) and SDP-M2(24) were set to $1/15$, 3, 1 and $1/5$, 2.6, 0.5, respectively. In Fig. 3, one can observe that SDP-P1 and SDP-P2 outperform SDP-M1 and SDP-M2, in reaching the same objective function value, by 25-30% and 25%, respectively.

3) Comparison of SDP-BSREM With BSREM: In this subsection, we analyzed the performance of SDP-BSREM algorithms compared to the BSREM algorithm. First, we showed the global NRMSD for all the algorithms, with 12 and 24 subsets, in Fig. 4, using the brain phantom. It showed that all algorithms converged to the same solution for both low and high count data. Further, this figure showed that SDP-P1 and SDP-P2 outperformed BSREM with respect to global NRMSD. To analyze convergence acceleration, we showed the objective function values of each algorithm in Fig. 5. In this figure, one can observe that both proposed algorithms, SDP-P1 and SDP-P2, outperform the BSREM algorithm, in reaching the same objective function value, by roughly a factor of two for all cases: 12 and 24 subsets for both low and high count data using the brain phantom.

Next, we examined the local convergence performance of SDP-BSREM algorithms by ROI based NRMSD in 8 different ROIs with different contrast ratios in the reconstructions of the uniform phantom. High count data and 24 subset were used in this experiments. In Fig. 6, we observed that the proposed

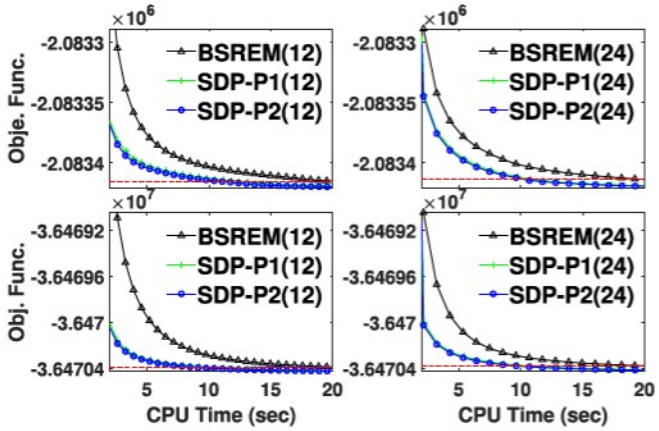


Fig. 5. Comparison of performance of SDP-BSREM vs. BSREM algorithm. Objective function vs. elapsed CPU time in reconstructions performed with BSREM, SDP-P1, and SDP-P2, with 12 (left) and 24 (right) subsets for the brain phantom with low (top row) and high (bottom low) count data, respectively. The dash lines represent the objective function values of BSREM at 20 seconds CPU time.

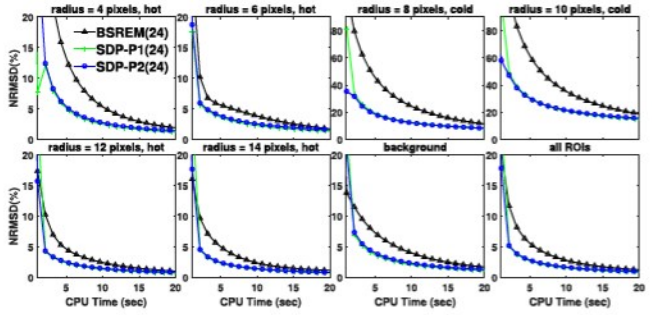


Fig. 6. Comparison of local convergence performance of SDP-BSREM vs. BSREM algorithms for the uniform phantom with high count data. ROI based normalized root mean square difference (NRMSD) vs. iterations is shown. The eight ROIs are the 4 hot spheres, 2 cold spheres, 1 background spheres, and the region consisting of all the former 7 ROIs, named "all ROIs".

SDP-P1 and SDP-P2 algorithms converged fast than BSREM algorithm in all 8 ROIs.

C. Clinical Results

The reconstructions were performed with our SDP-BSREM algorithm with P2 preconditioner and with commercial Q.Clear by means of the GE toolbox [33] with the penalty weight (β) set to the default value of 350. Because the 2D-projectors used in the simulations and 3D-projectors used in clinical data reconstructions were scaled differently, the penalty values used in respective reconstructions differed substantially.

To mimic the GE's clinical implementation of Q.Clear, 25 and 8 iterations were used for non-TOF and TOF data, respectively, with 24 subsets in the experiments. For the same reason we initialized both the non-TOF and TOF reconstructions using OSEM with 2 iterations and 24 subsets. The reconstructions with TOF data were further initialized using 3 iterations with 24 subsets of non-TOF algorithm. This gives a more clinically realistic view of the performance, but at the cost of being able to isolate TOF performance.

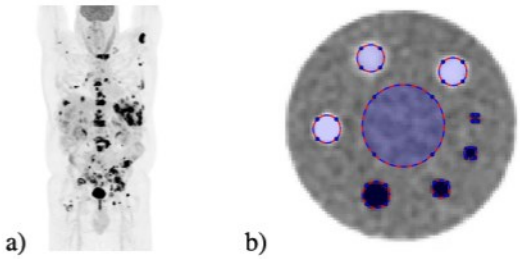


Fig. 7. Clinical PET patient and ACR phantom. a) Clinical PET patient: Coronal maximum intensity projection (MIP) of a clinical whole-body 18F-FDG PET patient image acquired on a GE D710 PET/CT and reconstructed using the Q.Clear clinical method [3]. b) Clinical ACR quality assurance phantom showing the regions of interest for cold/hot cylinders, 0:1 and 2.5:1 activity concentration ratios, respectively, and background [34].

TABLE III
ALGORITHMIC PARAMETERS FOR 3D PATIENT RECONSTRUCTION

Algorithm	Parameters
Q.Clear(nonTOF)	$\lambda_0 = 2, a = 1/5$
Q.Clear(TOF)	$\lambda_0 = 1.2, a = 1/5$
SDP-P2(nonTOF)	$\lambda_0 = 1.6, a = 1/4, \varrho = 2.2,$ $\delta_1 = 0.1, \delta_2 = 1.6, \nu_1 = 0.6, \nu_2 = 1.25$
SDP-P2(TOF)	$\lambda_0 = 1.1, a = 1/3, \varrho = 2.4,$ $\delta_1 = 0.6, \delta_2 = 1.6, \nu_1 = 0.6, \nu_2 = 1.25$

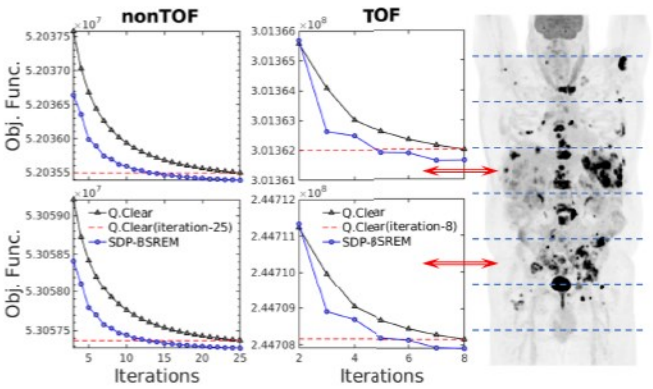


Fig. 8. Comparison of performance of SDP-BSREM vs. Q.Clear (β = 350) algorithms. A whole-body 18F-FDG clinical PET patient scan was used. Eight patient bed positions separated by dashed lines are shown in the coronal maximum intensity projection (MIP) image. The objective function vs. iterations is shown for PET scanner patient bed positions 4 and 6 (red arrows) for nonTOF (left) and TOF data (right). The dashed lines represent the objective function values at the final Q.Clear iterations.

The parameter values are shown in Table III. For simplicity, since good initializations were used, we set $J_0 = 0$ and $J_1 = 1000$. The other algorithmic parameters were found via an iterative golden search procedure using a single bed position (centered on the Derenzo region) from an ACR PET phantom [35] with similar count characteristics as the patient's data. Using this phantom each parameter was sequentially optimized with 5% tolerance and then used in search for the next parameter until parameter values ceased to change ($\sim$3 iterations).

In Fig. 8 we show convergence, via the objective function value as a function of iteration for non-TOF/TOF data, for an ^{18}F-FDG whole-body PET clinical patient (shown in Fig. 7a).

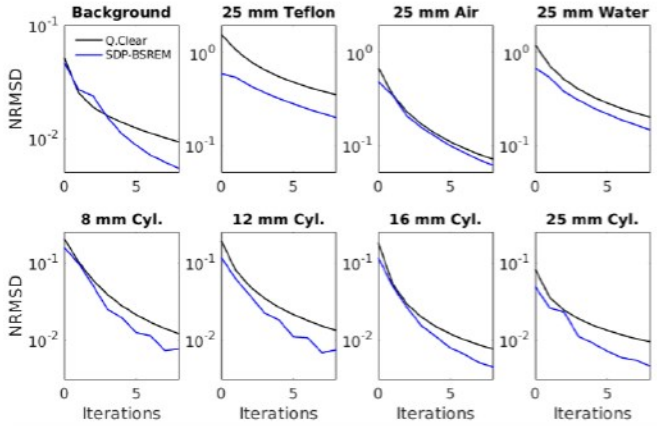


Fig. 9. Local convergence is assessed using the 8 regions of interest from an ACR quality assurance test with TOF data [34]. These regions of interest can be seen in Fig. 7b. Each subplot represents one of the eight regions, which are from left to right and top to bottom: background, cold Teflon/air/water, and hot 8/12/16/25 mm cylinders, respectively.

This data was obtained from 8 bed positions acquired on a GE D710 PET/CT. The nominal administered activity and post-administration acquisition were 444 MBq and 1-hour, respectively, with 3-minute dwell times and 25% overlap, resulting in $[4.1/3.4/4.2/\mathbf{4.5}/4.6/\mathbf{3.8}/3.1/2.6] \times 10^7$ total counts, where the bolded numbers are from the bed positions 4 and 6 shown in Fig. 8.

We observed that our SDP-BSREM method outperformed the Q.Clear algorithm, in reaching the same objective function value, by 40-50% and 35-50% for non-TOF and TOF data, respectively. We note that both the clinical 3D projection and penalty operator have much greater computational complexity than the convergence acceleration scheme described in section III. Hence, the increased computational cost required for the use of the SDP is negligible.

To evaluate the local convergence, TOF data from a quarterly ACR quality assurance test was used. Following ACR guidelines [34] the activity corresponded to a nominal 444 MBq (12 mCi) of ^{18}F-FDG administration used at MSKCC. The upper proton of this phantom contains 8 regions (3 cold, 4 hot, and background) with nominal contrast ratios of 0:1 and 2.5:1 for the cold and hot cylinders, respectively. ROI were defined using the cylinder boundaries from registered CT images. Using the methodology described by Kim *et al.* [36], for each ROI we measured the NRMSD, where the f^∞ in NRMSD is the converged image at 300 iterations by Q.Clear without subsets. These reconstructions used the same parameters as those used in the whole body patient reconstructions (*i.e.*, $J_0 = 0$, $J_1 = 1000$ and Table III). The results are shown in Fig. 9. For each ROI, the SDP-BSREM method converged to f^∞ faster than Q.Clear.

V. Conclusion

In this paper, we have presented the SDP-BSREM algorithms with two SDPs and their global convergence theorems. The two SDPs were designed based on the smoothness-promoting property in the reconstructed images of the regularization term. We tested these algorithms using both simulated and clinical PET data. Using two simulated phantoms, our numerical studies showed that, for solving the RDP regularized PET image reconstruction model, our proposed algorithms converged more quickly than BSREM. Similarly, when using clinical patient and phantom PET data, our proposed algorithm SDP-P2 outperformed Q.Clear. We plan to test the SDP-BSREM algorithm on more varied data sets acquired under a wide range of conditions seen in the clinic.

Appendix I

This appendix includes the proof of strict convexity and Lipschitz continuous gradient of the objective function Φ.

Proposition 9: If $\boldsymbol{A}^\top \boldsymbol{g} \neq \boldsymbol{0}$, then the objective function $\Phi(\boldsymbol{f})$ in (4) is strictly convex on $\mathbb{R}^q_+$.

Proof: From (2), the gradient of fidelity term F is given by $\nabla F = \boldsymbol{A}^\top(\boldsymbol{1}_p - \boldsymbol{g}/(\boldsymbol{A}\boldsymbol{f} + \boldsymbol{\gamma}))$. Then the Hessian matrix of F can be computed as follows:

$$\nabla^2 F(\boldsymbol{f}) = \boldsymbol{A}^\top \boldsymbol{G} \boldsymbol{A} \tag{20}$$

with $\boldsymbol{G}$ a diagonal matrix and $\boldsymbol{G} = \operatorname{diag}(\boldsymbol{g}/(\boldsymbol{A}\boldsymbol{f} + \boldsymbol{\gamma})^2)$. The first-order partial derivative of the RDP term $R(\boldsymbol{f})$ is given by

$$\frac{\partial R}{\partial f_j} = 2\sum_{k \in N_j} \frac{(f_j - f_k)(\gamma_R |f_j - f_k| + f_j + 3f_k + 2\epsilon)}{(f_j + f_k + \gamma_R |f_j - f_k| + \epsilon)^2}. \tag{21}$$

Then we have the second-order partial derivative of R as follows:

$$\frac{\partial^2 R}{\partial f_j \partial f_k} = \begin{cases} \displaystyle\sum_{l \in N_j} \frac{4(2f_l + \epsilon)^2}{(f_j + f_l + \gamma_R |f_j - f_l| + \epsilon)^3} & \text{if } k = j, \\ \displaystyle -\frac{4\left(4f_j f_k + 2\epsilon(f_j + f_k) + \epsilon^2\right)}{(f_j + f_k + \gamma_R |f_j - f_k| + \epsilon)^3} & \text{if } k \in N_j, \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

For any $\boldsymbol{0} \neq \boldsymbol{x} \in \mathbb{R}^q$, ignoring the zero entries of $\nabla^2 R$, we obtain

$$\boldsymbol{x}^\top \nabla^2 R \boldsymbol{x} = \sum_{j=1}^{q} \frac{\partial^2 R}{\partial f_j^2} x_j^2 + \sum_{j=1}^{q} \sum_{k \in N_j} \frac{\partial^2 R}{\partial f_j \partial f_k} x_j x_k. \tag{23}$$

By (22) and (23), we have

$$\boldsymbol{x}^\top \nabla^2 R \boldsymbol{x} = \sum_{j=1}^{q} \sum_{k \in N_j} \frac{2\left((2f_k + \epsilon)x_j - (2f_j + \epsilon)x_k\right)^2}{(f_j + f_k + \gamma_R |f_j - f_k| + \epsilon)^3}. \tag{24}$$

We can see that $\boldsymbol{x}^\top \nabla^2 R \boldsymbol{x} \geqslant 0$ and for any $\boldsymbol{x} \neq \boldsymbol{0}$, $\boldsymbol{x}^\top \nabla^2 R \boldsymbol{x} = 0$ if and only if there exists a nonzero constant c, such that $\boldsymbol{x} = c(2\boldsymbol{f} + \epsilon)$. For any $\boldsymbol{x} = c(2\boldsymbol{f} + \epsilon) \neq \boldsymbol{0}$, we have $\boldsymbol{x}^\top \nabla^2 F \boldsymbol{x} = c^2 \|\boldsymbol{G}^{1/2} \boldsymbol{A}(2\boldsymbol{f} + \epsilon)\|^2$, thus $\boldsymbol{x}^\top \nabla^2 F \boldsymbol{x} > 0$ by using $\boldsymbol{A}^\top \boldsymbol{g} \neq \boldsymbol{0}$. Since $\nabla^2 \Phi(\boldsymbol{f}) = \nabla^2 F(\boldsymbol{f}) + \beta \nabla^2 R(\boldsymbol{f})$, one can obtain that $\boldsymbol{x}^\top \nabla^2 \Phi \boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$. Then Φ is strictly convex on $\mathbb{R}^q_+$. ■

Proposition 10: The objective function $\Phi(\boldsymbol{f})$ in (4) has a Lipschitz continuous gradient on $\mathbb{R}^q_+$.

Proof: From (20), one can obtain that $\|\nabla^2 F\|_2 \leqslant \|A\|_2^2 \|g\|_\infty / \gamma_{\min}$, where $\gamma_{\min} > 0$ is the minimum entry of $\boldsymbol{\gamma}$. For any $\boldsymbol{x} \in \mathbb{R}^q$ with $\|\boldsymbol{x}\|_2 = 1$ and $\boldsymbol{f} \in \mathbb{R}^q_+$,

let $h(f) := x^\top \nabla^2 R(f)x$. From (24), we know that $h(f)$ is continuous on $\mathbb{R}^q_+$ and $\lim_{\|f\|_2\to\infty} h(f) = 0$. Thus there exists $C_1 > 0$ such that $|h(f)| \leqslant C_1$ for any x with $\|x\|_2 = 1$. Then we have $\|\nabla^2 R\|_2 \leqslant C_1$ which, when combined with the boundedness of $\|\nabla^2 F\|_2$, implies that $\|\nabla^2\Phi\|_2 \leqslant \|A\|_2^2\|g\|_\infty/\gamma_{\min} + C_1$. Using Lemma 1.2.2 in [37], one can obtain that $\nabla\Phi(f)$ is Lipschitz continuous with constant $C := (\|A\|_2^2\|g\|_\infty)/(\gamma_{\min}) + C_1$ on $\mathbb{R}^q_+$. ∎

Appendix II

This appendix includes proofs of Lemma 1, Lemma 3, Lemma 4, and Theorem 7. Here is the proof of Lemma 1.

Proof: We present only a proof of the Lipschitz continuity and the uniform boundedness can be shown in a similar manner. For any $f, \tilde{f} \in \mathcal{B}$, we consider the quantities $\Delta := S^{k,i}(f)\nabla\Phi_i(f) - S^{k,i}(\tilde{f})\nabla\Phi_i(\tilde{f})$, $\Delta_1 := \nabla\Phi_i(f) - \nabla\Phi_i(\tilde{f})$, and $\Delta_2 := S^{k,i}(f) - S^{k,i}(\tilde{f})$.

By the triangle inequality and the Cauchy-Schwarz inequality, we have that

$$\|\Delta\|_2 \leqslant \|S^{k,i}(f)\|_2\|\Delta_1\|_2 + \|\nabla\Phi_i(\tilde{f})\|_2\|\Delta_2\|_2. \tag{25}$$

According to condition (v), $\alpha^{k,i}$ is bounded. This together with the definition of $S(f)$ implies that $S^{k,i}(f) := \mathrm{diag}(\alpha^{k,i})S(f)$ is bounded for all $k \in \mathbb{N}_0$, $i \in \mathbb{N}_M$ and $f \in \mathcal{B}$. By condition (iii), there exists $c_2 > 0$ such that for all $k \in \mathbb{N}_0$, $i \in \mathbb{N}_M$

$$\|S^{k,i}(f)\|_2\|\Delta_1\|_2 \leqslant c_2\|f - \tilde{f}\|_2. \tag{26}$$

We next prove that $S^{k,i}(f)$ are Lipschitz continuous on $\mathcal{B}$ with Lipschitz constants bounded above by a constant. For any $f, \tilde{f} \in \mathcal{B}$, if both $f_j, \tilde{f}_j$ are in $[0, U/2)$ or in $[U/2, U]$, then $|S^{k,i}(f)_{jj} - S^{k,i}(\tilde{f})_{jj}| = |\alpha_j^{k,i}/p_j| \cdot |f_j - \tilde{f}_j|$. If f_j and $\tilde{f}_j$ are not in the same interval, without lose of generality, assuming $f_j \in [0, U/2)$, $\tilde{f}_j \in [U/2, U]$, then $|f_j - (U - \tilde{f}_j)| \leqslant |f_j - \tilde{f}_j|$. Therefore, one can get $|S^{k,i}(f)_{jj} - S^{k,i}(\tilde{f})_{jj}| \leqslant |\alpha_j^{k,i}/p_j| \cdot |f_j - \tilde{f}_j|$. Thus, $\|\Delta_2\|_2 \leqslant \|\alpha^{k,i}\|_2/p_0\|f - \tilde{f}\|_2$, where $p_0 := \min\{p_j : j \in \mathbb{N}_q\}$. Since $\alpha^{k,i}$ and $\nabla\Phi_i(\tilde{f})$ are bounded and $p_0 > 0$, there exists a constant $c_3 > 0$ such that for all $k \in \mathbb{N}_0$, $i \in \mathbb{N}_M$,

$$\|\nabla\Phi_i(\tilde{f})\|\|\Delta_2\|_2 \leqslant c_3\|f - \tilde{f}\|_2. \tag{27}$$

Let $L := c_2 + c_3$. It follows from (25), (26) and (27) that $\|\Delta\|_2 \leqslant L\|f - \tilde{f}\|_2$. That is, $S^{k,i}(f)\nabla\Phi_i(f)$ are Lipschitz continuous with Lipschitz constants bounded above by L for all $k \in \mathbb{N}_0$, $i \in \mathbb{N}_M$. ∎

The proof of Lemma 3 is presented as follows.

Proof: Let $b^{k,i} := \mathrm{diag}(\alpha - \delta^{k,i})\nabla\Phi_i(f^{k,i-1})$ for $k \in \mathbb{N}_0, i \in \mathbb{N}_M$. For $f_j^{k,i-1} \in (0, U/2)$, $i \in \mathbb{N}_M$, we have that $S^{k,i}(f^{k,i-1})_{jj} = (\alpha_j - \delta_j^{k,i})f_j^{k,i}/p_j$. By assumption, $f^{k,i} = P_t(f^{k,i}) = \tilde{f}^{k,i}$. The definition of $f^{k,i}$ yields $f_j^{k,i} = f_j^{k,i-1}(1 - \lambda_k/p_j b_j^{k,i})$, which implies $f_j^{k+1} = f_j^k\Pi_{i=1}^M(1 - \lambda_k/p_j b_j^{k,i})$. By the boundedness of $b_j^{k,i}/p_j$, we have $f_j^{k+1} = f_j^k(1 - \lambda_k/p_j\sum_{i=1}^M b_j^{k,i} + \mathcal{O}(\lambda_k^2))$. We next estimate $\sum_{i=1}^M b^{k,i}$. By definition, $b^{k,i} = \mathrm{diag}(\alpha)\nabla\Phi_i(f^{k,i-1}) - \mathrm{diag}(\delta^{k,i})\nabla\Phi_i(f^{k,i-1})$. To estimate the first term of the last equation, we write $\nabla\Phi_i(f^{k,i-1}) = \Delta^{k,i} + \nabla\Phi_i(f^k)$, where $\Delta^{k,i} := \nabla\Phi_i(f^{k,i-1}) - \nabla\Phi_i(f^k)$. It follows that $\sum_{i=1}^M \nabla\Phi_i(f^{k,i-1}) = \sum_{i=1}^M \Delta^{k,i} + \nabla\Phi(f^k)$. By condition (iii), there exists $c_4 > 0$ such that $\|\Delta^{k,i}\|_2 \leqslant c_4\|f^{k,i-1} - f^k\|_2$. Since $\tilde{f}^{k,i} \in \mathrm{int}\,\mathcal{B}$, by (14) and Lemma 1, there exists a constant $c_5 > 0$ such that

$$\|f^{k,i-1} - f^k\|_2 \leqslant c_5 M\lambda_k. \tag{28}$$

Hence $\|\Delta^{k,i}\|_2 = \mathcal{O}(\lambda_k)$ and this gives that $\sum_{i=1}^M \nabla\Phi_i(f^{k,i-1}) = \nabla\Phi(f^k) + \mathcal{O}(\lambda_k)$. Moreover, by condition (iii), we have that $\|\sum_{i=1}^M \mathrm{diag}(\delta^{k,i})\nabla\Phi_i(f^{k,i-1})\|_2 = \mathcal{O}(\delta_k)$. Therefore, $\sum_{i=1}^M b^{k,i} = \mathrm{diag}(\alpha)\nabla\Phi(f^k) + \mathcal{O}(\lambda_k) + \mathcal{O}(\delta_k)$. Thus, we obtain (15).

Equation (16) may be shown in a similar manner. Indeed, for $f_j^{k,i-1} \in [U/2, U)$, $i \in \mathbb{N}_M$, we have that $S^{k,i}(f^{k,i-1})_{jj} = \alpha_j^{k,i}(U - f_j^{k,i})/p_j$. The definition of $f^{k,i}$ yields that $(U - f_j^{k+1}) = (U - f_j^k)\Pi_{i=1}^M(1 + \lambda_k/p_j b_j^{k,i})$. This equation with similar arguments leads to (16). ∎

The proof of Lemma 4 is presented as follows.

Proof: We first prove part (a). Let $H(f)$ denote the Hessian matrix of $\Phi(f)$, '$\odot$' denote the component-wise multiplication of two vectors, and $h^k := f^{k+1} - f^k$. By the Taylor expansion, we have that

$$\Phi(f^{k+1}) = \Phi(f^k) + (\nabla\Phi(f^k))^\top h^k + R_k, \tag{29}$$

where $R_k := (h^k)^\top H(f^k + \theta \odot h^k)h^k$, for some vector $\theta \in \mathbb{R}^q_+$ with $\theta_j \in (0, 1)$ for all $j \in \mathbb{N}_q$. We now estimate R_k. By condition (iii), $\nabla\Phi(f)$ is Lipschitz continuous. Hence $H(f)$ is bounded on $\mathcal{B}$. Then we have $|R_k| = \mathcal{O}(\|h^k\|_2^2)$. We next evaluate the term h^k. For notation simplicity, we let $e^{k,i} := S^{k,i}(f^k)\nabla\Phi_i(f^k) - S^{k,i}(f^{k,i-1})\nabla\Phi_i(f^{k,i-1})$, and $e^k := \sum_{i=1}^M e^{k,i}$. By Lemma 1, we have $\|e^{k,i}\|_2 \leqslant L\|f^{k,i-1} - f^k\|_2$. This combined with (28) implies that $\|e^{k,i}\|_2 = \mathcal{O}(\lambda_k)$, and thus $\|e^k\|_2 = \mathcal{O}(\lambda_k)$. By assumption, $\tilde{f}^{k,i} \in \mathrm{int}\,\mathcal{B}$, from (14), we obtain that

$$h^k = -\lambda_k\sum_{i=1}^M S^{k,i}(f^k)\nabla\Phi_i(f^k) + \lambda_k e^k. \tag{30}$$

Let $d^k := S(f^k)\sum_{i=1}^M \mathrm{diag}(\delta^{k,i})\nabla\Phi_i(f^k)$, and $J_k := (\nabla\Phi(f^k))^\top\mathrm{diag}(\alpha)S(f^k)\nabla\Phi(f^k)$. Then $\|d^k\|_2 = \mathcal{O}(\delta_k)$ by condition (v) and Lemma 1, and $J_k \geqslant 0$ by the positive semi-definiteness of $\mathrm{diag}(\alpha)S(f^k)$. Since $S^{k,i}(f^k) = \mathrm{diag}(\alpha - \delta^{k,i})S(f^k)$, then we have

$$\sum_{i=1}^M S^{k,i}(f^k)\nabla\Phi_i(f^k) = \mathrm{diag}(\alpha)S(f^k)\nabla\Phi(f^k) - d^k. \tag{31}$$

Combining this with (30), we have

$$h^k = -\lambda_k\mathrm{diag}(\alpha)S(f^k)\nabla\Phi(f^k) + \lambda_k(d^k + e^k). \tag{32}$$

By the boundedness of $\mathrm{diag}(\alpha)S(f)\nabla\Phi(f)$ and the norm of e^k and d^k, we have $\|h^k\|_2 = \mathcal{O}(\lambda_k)$ and hence, $|R_k| = \mathcal{O}(\lambda_k^2)$. This combined with (29), (32) and the boundedness of $\nabla\Phi(f)$ yields that

$$\Phi(f^{k+1}) = \Phi(f^k) - \lambda_k J_k + \mathcal{O}(\lambda_k\delta_k) + \mathcal{O}(\lambda_k^2). \tag{33}$$

For $s \in \mathbb{N}$, summing both sides of (33) for k from 0 to s, we obtain that

$$\Phi(\boldsymbol{f}^{s+1}) = \Phi(\boldsymbol{f}^0) + \sum_{k=0}^{s}[-\lambda_k J_k + \mathcal{O}(\lambda_k\delta_k) + \mathcal{O}(\lambda_k^2)]. \quad (34)$$

We now prove the convergence of the right hand side of (34). By condition (vi), $\sum_{k=0}^{\infty}\lambda_k\delta^{k,i}$ is convergent, and hence $\sum_{k=0}^{\infty}\lambda_k\delta_k$ is convergent. Notice the facts we have in hand: (i) $\sum_{k=0}^{\infty}\lambda_k^2 < \infty$ (by condition (iv)); (ii) the convergence of $\sum_{k=0}^{\infty}\lambda_k\delta_k$, it remains to show that $\sum_{k=0}^{\infty}\lambda_k J_k$ is convergent. In view of the facts (i), (ii) and the boundedness of $\Phi(\boldsymbol{f})$ on $\mathcal{B}$, the partial sum $\sum_{k=0}^{s}\lambda_k J_k$ is bounded, which combined with its monotonicity ($\lambda_k J_k \geqslant 0$) implies its convergence.

We next prove part (b). Since for each k, $J_k \geq 0$, there exists a subsequence J^{k_n} such that $\lim_{n\to\infty} J^{k_n} = 0$. In fact, assume to the contrary that there exists $\epsilon_0 > 0$ and $K_0 \in \mathbb{N}_+$ such that $J_k \geqslant \epsilon_0$, for all $k > K_0$. Because $\sum_{k=0}^{\infty}\lambda_k = \infty$, by condition (iv), and $\lambda_k > 0$, we would have $\sum_{k=0}^{s}\lambda_k J_k \geqslant \epsilon_0\sum_{k=0}^{s}\lambda_k \to \infty$, as $s \to \infty$, a contradiction. Moreover, since $\boldsymbol{f}^{k_n}$ is bounded, there exists a convergent subsequence $\boldsymbol{f}^{k'_n}$ having the limit $\boldsymbol{f}^* \in \mathcal{B}$. Thus, $(\nabla\Phi(\boldsymbol{f}^*))^\top \mathrm{diag}(\boldsymbol{\alpha})\boldsymbol{S}(\boldsymbol{f}^*)\nabla\Phi(\boldsymbol{f}^*) = \lim_{n\to\infty} J^{k'_n} = 0$. Letting $r_j := (\partial/\partial f_j)\Phi(\boldsymbol{f}^*)$ and $s_j := \boldsymbol{S}(\boldsymbol{f}^*)_{jj}$, the last equation yields $\sum_{j=1}^{q}\alpha_j s_j r_j^2 = 0$. Since $s_j \geqslant 0$ and $\alpha_j > 0$, we have for all $j \in \mathbb{N}_q$ that $\alpha_j s_j r_j^2 \geqslant 0$, which implies that $s_j r_j = 0$ for all $j \in \mathbb{N}_q$. That is, $\boldsymbol{S}(\boldsymbol{f}^*)\nabla\Phi(\boldsymbol{f}^*) = \boldsymbol{0}$.

Finally, we show part (c). According to [38, Page 203], it suffices to prove that for each $j \in \mathbb{N}_q$, (i) if $0 < f_j^* < U$, then $r_j = 0$; (ii) if $f_j^* = 0$, then $r_j \geqslant 0$; and if $f_j^* = U$, then $r_j \leqslant 0$. Case (i) clearly follows from part (b), from which we have $s_j r_j = 0$ for all $j \in \mathbb{N}_q$. By the definition (8) of $\boldsymbol{S}(\boldsymbol{f})$, $s_j > 0$ for $0 < f_j^* < U$, and thus, $r_j = 0$.

It remains to prove case (ii). To this end, we let $\mathcal{J}_1 := \left\{j' \in \mathbb{N}_q : f_{j'}^* = 0, r_{j'} < 0\right\}$, $\mathcal{J}_2 := \left\{j'' \in \mathbb{N}_q : f_{j''}^* = U, r_{j''} > 0\right\}$, $\mathcal{J} := \mathcal{J}_1 \cup \mathcal{J}_2$, and show $\mathcal{J} = \emptyset$. Assume to the contrary that $\mathcal{J} \neq \emptyset$. Then, either $\mathcal{J}_1 \neq \emptyset$ or $\mathcal{J}_2 \neq \emptyset$. Assume $\mathcal{J}_1 \neq \emptyset$, then for any $j' \in \mathcal{J}_1$, since $\nabla\Phi(\boldsymbol{f})$ is continuous at $\boldsymbol{f}^*$, there exists $\delta \in (0, U/4)$ such that for all $\boldsymbol{f} \in \mathcal{B}_\delta := \left\{\boldsymbol{f} \in \mathcal{B} : \|\boldsymbol{f} - \boldsymbol{f}^*\|_2 < \delta\right\}$, there holds $(\partial/\partial f_{j'})\Phi(\boldsymbol{f}) < 0$. By Lemma 2, there exists $K_1 \in \mathbb{N}$ such that for $k > K_1$, $\|\boldsymbol{f}^{k,i-1} - \boldsymbol{f}^k\|_2 < \delta$. Then for $k > K_1$, if $\boldsymbol{f}^k \in \mathcal{B}_\delta$, we have $\|\boldsymbol{f}^{k,i-1} - \boldsymbol{f}^*\|_2 < 2\delta < U/2$, and hence, $f_{j'}^{k,i-1} \in (0, U/2)$ for all $i \in \mathbb{N}_M$, $j' \in \mathcal{J}_1$. In this case, $S^{k,i}(\boldsymbol{f}^{k,i-1})_{j'j'} = \alpha_{j'}^{k,i} f_{j'}^{k,i}/p_{j'}$ for all $i \in \mathbb{N}_M$, and hence Lemma 3 ensures that (15) holds. Since $(\partial/\partial f_{j'})\Phi(\boldsymbol{f}^k) < 0$ for $\boldsymbol{f}^k \in \mathcal{B}_\delta$ and $\lambda_k, \delta_k \to 0$ as $k \to \infty$, then there exists $K_2 > K_1$ such that if $k > K_2$ and $\boldsymbol{f}^k \in \mathcal{B}_\delta$, we have $f_{j'}^{k+1} > f_{j'}^k$ for any $j' \in \mathcal{J}_1$. Therefore, for $k > K_2$, if $\boldsymbol{f}^k \in \mathcal{B}_\delta$, then we have $f_{j'}^{k+1} > f_{j'}^k$ for any $j' \in \mathcal{J}_1$. Similarly, if $\mathcal{J}_2 \neq \emptyset$, for any $j'' \in \mathcal{J}_2$, there exists $K_3 > K_2$ such that for $k > K_3$, if $\boldsymbol{f}^k \in \mathcal{B}_\delta$, then (16) ensures that $f_{j''}^{k+1} > f_{j''}^k$.

Since $\lim_{n\to\infty}\boldsymbol{f}^{k'_n} = \boldsymbol{f}^*$, there exists $K_4 > K_3$ such that if $k'_n > K_4$, $\boldsymbol{f}^{k'_n} \in \mathcal{B}_\delta$. Suppose $k'_{n_0} > K_4$ for some n_0. Let $t_n := \max\left\{k : K_4 < k < k'_{n+n_0}, \boldsymbol{f}^k \notin \mathcal{B}_\delta\right\}$. If for some n, $\boldsymbol{f}^k \in \mathcal{B}_\delta$ for all $K_4 < k < k'_{n+n_0}$, set $t_n := K_4$, and hence $t_n \geqslant K_4$. Clearly, we have $\boldsymbol{f}^k \in \mathcal{B}_\delta$ if $t_n + 1 \leqslant k \leqslant k'_{n+n_0}$. Moreover, t_n is a monotone increasing sequence. Then, either (a) $\lim_{n\to\infty} t_n = t_0 \geqslant K_4$, or (b) $\lim_{n\to\infty} t_n = +\infty$. If it is the case (a), then $\boldsymbol{f}^k \in \mathcal{B}_\delta$ for all $k > t_0$. Thus, for $m > l > t_0$ that $f_{j'}^m > f_{j'}^l > 0$ for $j' \in \mathcal{J}_1$. This contradicts the fact that $f_{j'}^* = 0$ for $j' \in \mathcal{J}_1$. Hence, it must be the case (b). Since $\boldsymbol{f}^k \in \mathcal{B}_\delta$ for $t_n + 1 \leqslant k \leqslant k'_{n+n_0}$, we have that $f_{j'}^{k'_{n+n_0}} \geqslant f_{j'}^{t_n+1} > 0$ for $j' \in \mathcal{J}_1$ and $f_{j''}^{k'_{n+n_0}} \leqslant f_{j''}^{t_n+1} < U$ for $j'' \in \mathcal{J}_2$. It follows that $\lim_{n\to\infty} f_{j'}^{t_n+1} = 0$ for $j' \in \mathcal{J}_1$ and $\lim_{n\to\infty} f_{j''}^{t_n+1} = U$ for $j'' \in \mathcal{J}_2$. Then, Lemma 2 ensures that $\lim_{n\to\infty} f_{j'}^{t_n} = 0$ for $j' \in \mathcal{J}_1$ and $\lim_{n\to\infty} f_{j''}^{t_n} = U$ for $j'' \in \mathcal{J}_2$. Thus, we can find a convergent subsequence $\boldsymbol{f}^{t_{n_l}}$ of $\boldsymbol{f}^{t_n}$ such that $\lim_{l\to\infty}\boldsymbol{f}^{t_{n_l}} = \boldsymbol{f}^{**}$ with $f_{j'}^{**} = 0$ for $j' \in \mathcal{J}_1$ and $f_{j''}^{**} = U$ for $j'' \in \mathcal{J}_2$. Since $\boldsymbol{f}^{t_n} \notin \mathcal{B}_\delta$, we observe that $\boldsymbol{f}^{**} \notin \mathcal{B}_\delta$, which ensures that $\boldsymbol{f}^{**} \neq \boldsymbol{f}^*$. By part (a), $\Phi(\boldsymbol{f}^k)$ is convergent, which implies that $\Phi(\boldsymbol{f}^{**}) = \Phi(\boldsymbol{f}^*)$. Let $\mathcal{D} := \left\{\boldsymbol{f} \in \mathcal{B} : f_{j'} = 0 \text{ for } j' \in \mathcal{J}_1 \text{ and } f_{j''} = U \text{ for } j'' \in \mathcal{J}_2\right\}$. Thus, $\boldsymbol{f}^*, \boldsymbol{f}^{**} \in \mathcal{D}$. It can be verified for any $\boldsymbol{f} \in \mathcal{D}$ that $\langle \boldsymbol{f} - \boldsymbol{f}^*, \nabla\Phi(\boldsymbol{f}^*)\rangle \geqslant 0$. By [38, Page 203], $\boldsymbol{f}^*$ is a minimizer of Φ over $\mathcal{D}$. Hence, Φ has two different minimizers $\boldsymbol{f}^*$ and $\boldsymbol{f}^{**}$ over the convex set $\mathcal{D}$. This contradicts the assumption that Φ has a unique minimizer on $\mathcal{B}$. Thus, we have that $\mathcal{J} = \emptyset$. ∎

Here is the proof of Theorem 8.

Proof: From Proposition 9 and Proposition 10, we have that Φ satisfies conditions (i)-(iii). One can directly obtain that the relaxation sequence λ_k satisfies condition (iv).

By Theorem 7, to prove the convergence of SDP-BSREM algorithm with λ_k and $\boldsymbol{S}^{k,i}$, it is sufficient to show that λ_k and $\boldsymbol{S}^{k,i}$ satisfy conditions (v) and (vi). To do this, for the subiteration-dependent preconditioner $\boldsymbol{S}^{k,i}(\boldsymbol{f}) = \mathrm{diag}(\alpha_{k,i}\boldsymbol{v}^{k,i})\boldsymbol{S}(\boldsymbol{f})$, one need to show that $\lim_{k\to\infty}\alpha_{k,i} = \alpha > 0$ for all $i \in \mathbb{N}_M$, and $\sum_{k=0}^{\infty}\lambda_k(\alpha - \alpha_{k,i})$ converges for all $i \in \mathbb{N}_M$ since $\boldsymbol{v}^{k,i}$ is a positive vector sequence and $\boldsymbol{v}^{k,i} = \boldsymbol{v}^{k_1,M}$ for $k > k_1$ and $i \in \mathbb{N}_M$.

For $\alpha_{k,i}$ defined in (17), the sequence $t_{k,i}$ is increasing. By induction, we have that $t_{k,i} > (kM + i)/2$. Further, it can be shown that $\lim_{k\to\infty} t_{k,i}/k = M/2$, and thus $\lim_{k\to\infty} t_{k,i}/t_{k,i+1} = 1$ for all $i \in \mathbb{N}_M$. Then we can obtain $\lim_{k\to\infty}\alpha_{k,i} = 1 + \lim_{k\to\infty}(t_{k,i} - 1)/t_{k,i+1} = 2 > 0$. By computing $2 - \alpha_{k,i} = 1/(2t_{k,i+1}(\sqrt{1 + 4t_{k,i}^2 + 4t_{k,i}})) + 3/(2t_{k,i+1})$, we have that $\lim_{k\to\infty} k(2 - \alpha_{k,i}) = 3/M$. Thus the series $\sum_{k=0}^{\infty}\lambda_k(2-\alpha_{k,i})$ converges since $\sum_{k=0}^{\infty} 1/k^2 < \infty$. Therefore, for preconditioners P1 or M1 and relaxation λ_k, conditions (v) and (vi) are satisfied.

For $\alpha_{k,i}$ defined in (18), we have that $\alpha_{k,i}$ is monotone and $\lim_{k\to\infty}\alpha_{k,i} = \varrho > 0$. It can be shown that $\lim_{k\to\infty} k(\varrho - \alpha_{k,i}) = (\varrho\delta_1 - \delta_2)/M$. Hence for precondition P2 or M2 and relaxation λ_k, conditions (v) and (vi) are satisfied. ∎

ACKNOWLEDGMENT

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

[1] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-1, no. 2, pp. 113–122, Oct. 1982.

[2] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomogr.*, vol. 8, no. 2, pp. 306–316, Apr. 1984.

[3] S. Ahn *et al.*, "Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET," *Phys. Med. Biol.*, vol. 60, no. 15, pp. 5733–5751, Aug. 2015.

[4] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–609, Dec. 1994.

[5] J. Browne and A. B. D. Pierro, "A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography," *IEEE Trans. Med. Imag.*, vol. 15, no. 5, pp. 687–699, Oct. 1996.

[6] D. P. Bertsekas, "A new class of incremental gradient methods for least squares problems," *SIAM J. Optim.*, vol. 7, no. 4, pp. 913–926, 1997.

[7] A. R. De Pierro and M. E. B. Yamagishi, "Fast EM-like methods for maximum '*a posteriori*' estimates in emission tomography," *IEEE Trans. Med. Imag.*, vol. 20, no. 4, pp. 280–288, Apr. 2001.

[8] S. Ahn and J. A. Fessler, "Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms," *IEEE Trans. Med. Imag.*, vol. 22, no. 5, pp. 613–626, May 2003.

[9] Y.-J. Tsai *et al.*, "Fast quasi-Newton algorithms for penalized reconstruction in emission tomography and further improvements via preconditioning," *IEEE Trans. Med. Imag.*, vol. 37, no. 4, pp. 1000–1010, Apr. 2018.

[10] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.

[11] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 492–526, Sep. 2010.

[12] S. Geman and D. E. McClure, "Statistical methods for tomographic image reconstruction," *Bull. Int. Statist. Inst.*, vol. 52, no. 4, pp. 5–21, 1987.

[13] E. U. Mumcuoglu, R. M. Leahy, and S. R. Cherry, "Bayesian reconstruction of PET images: Methodology and performance analysis," *Phys. Med. Biol.*, vol. 41, no. 9, pp. 1777–1807, 1996.

[14] Z. Zhang *et al.*, "Investigation of optimization-based reconstruction with an image-total-variation constraint in PET," *Phys. Med. Biol.*, vol. 61, no. 16, pp. 6055–6084, Aug. 2016.

[15] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, 2011.

[16] A. Krol, S. Li, L. Shen, and Y. Xu, "Preconditioned alternating projection algorithms for maximum *a posteriori* ECT reconstruction," *Inverse Problems*, vol. 28, no. 11, Nov. 2012, Art. no. 115005.

[17] S. Li *et al.*, "Effective noise-suppressed and artifact-reduced reconstruction of SPECT data using a preconditioned alternating projection algorithm," *Med. Phys.*, vol. 42, no. 8, pp. 4872–4887, 2015.

[18] J. Nuyts, D. Beque, P. Dupont, and L. Mortelmans, "A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography," *IEEE Trans. Nucl. Sci.*, vol. 49, no. 1, pp. 56–60, Feb. 2002.

[19] G. Wang and J. Qi, "Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization," *IEEE Trans. Med. Imag.*, vol. 31, no. 12, pp. 2194–2204, Dec. 2012.

[20] G. Wang and J. Qi, "Edge-preserving PET image reconstruction using trust optimization transfer," *IEEE Trans. Med. Imag.*, vol. 34, no. 4, pp. 930–939, Apr. 2015.

[21] E. U. Mumcuoglu, R. Leahy, S. R. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 687–701, Dec. 1994.

[22] Y. Lin, C. R. Schmidtlein, Q. Li, S. Li, and Y. Xu, "A krasnoselskii-mann algorithm with an improved EM preconditioner for PET image reconstruction," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2114–2126, Sep. 2019.

[23] M. J. Ehrhardt, P. Markiewicz, and C.-B. Schönlieb, "Faster PET reconstruction with non-smooth priors by randomization and preconditioning," *Phys. Med. Biol.*, vol. 64, no. 22, Nov. 2019, Art. no. 225019.

[24] D. F. Yu and J. A. Fessler, "Edge-preserving tomographic reconstruction with nonlocal regularization," *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 159–173, Feb. 2002.

[25] Y. E. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Sov. Math. Dokl.*, vol. 27, no. 2, pp. 372–376, 1983.

[26] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Mar. 2009.

[27] X. Zeng, L. Shen, and Y. Xu, "A convergent fixed-point proximity algorithm accelerated by FISTA for the l0 sparse recovery problem," in *Imaging, Vision and Learning Based on Optimization and PDEs*, X.-C. Tai, E. Bae, and M. Lysaker, Eds. Cham, Switzerland: Springer, 2018, pp. 27–45.

[28] D. Kim, S. Ramani, and J. A. Fessler, "Combining ordered subsets and momentum for accelerated X-ray CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 34, no. 1, pp. 167–178, Jan. 2015.

[29] Q. Li, E. Asma, S. Ahn, and R. M. Leahy, "A fast fully 4-D incremental gradient reconstruction algorithm for list mode PET data," *EEE Trans. Med. Imag.*, vol. 26, no. 1, pp. 58–67, Jan. 2007.

[30] C. R. Schmidtlein *et al.*, "Relaxed ordered subset preconditioned alternating projection algorithm for PET reconstruction with automated penalty weight selection," *Med. Phys.*, vol. 44, no. 8, pp. 4083–4097, 2017.

[31] W. W. Moses, "Fundamental limits of spatial resolution in PET," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 648, pp. S236–S240, Aug. 2011.

[32] B. Berthon *et al.*, "PETSTEP: Generation of synthetic PET lesions for fast evaluation of segmentation methods," *Phys. Med.*, vol. 31, no. 8, pp. 969–980, 2015.

[33] S. Ross and K. Thielemans, "General electric PET toolbox release 5.0. [License Agreement]," Gen. Electr. Healthcare, Waukesha, WI, USA, Tech. Rep., 2019.

[34] *PET Phantom Instructions for Evaluation of PET Image Quality*, Amer. Assoc. Physicists Med., Alexandria, VI, USA, 2012.

[35] C. R. MacFarlane, "ACR accreditation of nuclear medicine and PET imaging departments," *J. Nucl. Med. Technol.*, vol. 34, no. 1, pp. 18–24, Mar. 2006.

[36] K. Kim *et al.*, "Time of flight PET reconstruction using nonuniform update for regional recovery uniformity," *Med. Phys.*, vol. 46, no. 2, pp. 649–664, Feb. 2019.

[37] Y. E. Nesterov, *Introductory Lectures Convex Optimization: A Basic Course*. New York, NY, USA: Kluwer, 2004.

[38] B. T. Polyak, *Introduction to Optimization*. New York, NY, USA: Optim Software, 1987.