

An ordered-subsets proximal preconditioned gradient algorithm for edge-preserving PET image reconstruction

Abolfazl Mehranian

Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva, Switzerland

Arman Rahmim

Department of Radiology, Johns Hopkins University, Baltimore, Maryland 21287

Mohammad Reza Ay

Department of Medical Physics and Biomedical Engineering, Tehran University of Medical Sciences, Tehran, P.O. Box 14155-6447, Iran; Medical Imaging Systems Group, Research Center for Molecular and Cellular Imaging, Tehran University of Medical Sciences, Tehran, P.O. Box 14185-615, Iran; and Research center for Nuclear Medicine, Tehran University of Medical Sciences, Tehran, P.O. Box 14155-6447, Iran

Fotis Kotasidis

Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva, Switzerland

Habib Zaidi^{a)}

Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva, Switzerland; Geneva Neuroscience Center, Geneva University, CH-1205 Geneva, Switzerland; and Department of Nuclear Medicine and Molecular Imaging, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, the Netherlands

(Received 20 December 2012; revised 26 February 2013; accepted for publication 27 March 2013; published 23 April 2013)

Purpose: In iterative positron emission tomography (PET) image reconstruction, the statistical variability of the PET data precorrected for random coincidences or acquired in sufficiently high count rates can be properly approximated by a Gaussian distribution, which can lead to a penalized weighted least-squares (PWLS) cost function. In this study, the authors propose a proximal preconditioned gradient algorithm accelerated with ordered subsets (PPG-OS) for the optimization of the PWLS cost function and develop a framework to incorporate boundary side information into edge-preserving total variation (TV) and Huber regularizations.

Methods: The PPG-OS algorithm is proposed to address two issues encountered in the optimization of PWLS function with edge-preserving regularizers. First, the second derivative of this function (Hessian matrix) is shift-variant and ill-conditioned due to the weighting matrix (which includes emission data, attenuation, and normalization correction factors) and the regularizer. As a result, the paraboloidal surrogate functions (used in the optimization transfer techniques) end up with high curvatures and gradient-based algorithms take smaller step-sizes toward the solution, leading to a slow convergence. In addition, preconditioners used to improve the condition number of the problem, and thus to speed up the convergence, would poorly act on the resulting ill-conditioned Hessian matrix. Second, the PWLS function with a nondifferentiable penalty such as TV is not amenable to optimization using gradient-based algorithms. To deal with these issues and also to enhance edge-preservation of the TV and Huber regularizers by incorporating adaptively or anatomically derived boundary side information, the authors followed a proximal splitting method. Thereby, the optimization of the PWLS function is split into a gradient descent step (upgraded by preconditioning, step size optimization, and ordered subsets) and a proximal mapping associated with boundary weighted TV and Huber regularizers. The proximal mapping is then iteratively solved by dual formulation of the regularizers.

Results: The convergence performance of the proposed algorithm was studied with three different diagonal preconditioners and compared with the state-of-the-art separable paraboloidal surrogates accelerated with ordered-subsets (SPS-OS) algorithm. In simulation studies using a realistic numerical phantom, it was shown that the proposed algorithm depicts a considerably improved convergence rate over the SPS-OS algorithm. Furthermore, the results of bias-variance and signal-to-noise evaluations showed that the proposed algorithm with anatomical edge information depicts an improved performance over conventional regularization. Finally, the proposed PPG-OS algorithm is used for image reconstruction of a clinical study with adaptively derived boundary edge information, demonstrating the potential of the algorithm for fast and edge-preserving PET image reconstruction.

Conclusions: The proposed PPG-OS algorithm shows an improved convergence rate with the ability of incorporating additional boundary information in regularized PET image reconstruction. © 2013 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4801898>]

Key words: total variation regularization, preconditioning, proximal gradient, image reconstruction

I. INTRODUCTION

Positron emission tomography (PET) is one of the leading molecular imaging modalities for the detection, staging, and monitoring of metabolically active lesions in oncology,¹ the diagnosis and risk stratification of cardiovascular diseases in cardiology,² functional brain imaging in a wide variety of clinical neurological applications³ and many other medical disciplines. Over the years, PET imaging has experienced considerable advances in data acquisition and image reconstruction.^{4,5} The development of model-based statistical image reconstruction techniques has substantially increased the quantitative accuracy of PET images. These reconstruction techniques can explicitly account for the physical and statistical processes involved in the image formation procedure and allow for the incorporation of prior knowledge and side information about the PET activity distribution.^{6,7} In the Bayesian estimation framework (under zero-one loss function), statistical reconstruction techniques are formulated as the maximum *a posteriori* (MAP) estimation of the underlying PET image given the acquired emission data. Specifically, MAP reconstruction attempts to estimate an image that has the maximal statistical consistency with measured data (expressed as maximum Poisson or Gaussian likelihood) and at the same time is in conformance with *a priori* knowledge (expressed as *a priori* distribution or penalty function).⁸

The statistical variability of prompt PET measurements is best described by the Poisson distribution; however, when the measurements are corrected for accidental coincidences or acquired at sufficiently high count rates, the Poisson distribution can be approximated to the Gaussian distribution.⁹ Making use of the Gaussian model or likelihood leads to a weighted least-squares (WLS) data-consistency term, which when augmented by *a priori* or penalty function results in a penalized weighted least-squares (PWLS) cost function.⁹ In the PWLS reconstruction, the mean and variance of the measured data are assumed to be known following the correction for attenuation, detector normalization, and accidental coincidences. The minimization of the resulting cost function maximizes the posterior probability of the underlying activity distribution given the measured data. The penalty function penalizes deviations from an *a priori* knowledge and in fact imposes our expectations about the image being estimated. For this purpose, quadratic Tikhonov-based penalty functions have been widely used to suppress noise and impose smoothness during reconstruction.¹⁰ These priors effectively suppress noise on the assumption that sharp transitions between adjacent voxels are probably due to noise.⁸ However, as a compromise, the sharp edges at the boundaries of organs or lesions are blurred. To avoid the smoothing of boundaries, a variety of edge-preserving nonquadratic penalty

functions have been studied in the literature,^{11–17} among which Huber¹⁸ and total variation (TV) (Ref. 19) have garnered significant attention. These penalty functions or regularizers emphasize on the smoothing or regularization of neighboring voxels with small intensity differences, while retaining edges between the voxels with large intensity differences. In most of the above approaches, boundaries are detected based on the short-range and local interactions of neighboring voxels. In an attempt to encourage the formation and preservation of continuous boundaries, various penalty functions have been proposed to exploit the nonlocal features captured from the image itself^{20–22} or from other side information.^{23,24} Yu and Fessler²⁰ proposed a boundary weighted quadratic prior which adaptively derives boundary weights (curves) between organs and thereby penalizes local differences inside each region more than those across boundaries. With the advent of dual modality PET/CT and more recently PET/MR scanners, the penalty functions exploiting anatomical side information, known as anatomical priors, have gained attention in PET image reconstruction.^{25–28} Anatomical priors are commonly used to encourage spatial smoothness within organs in PET images while discouraging it across organ boundaries. For this purpose, some groups incorporated anatomical labels and boundaries into the quadratic priors.^{24,25,29,30} The mutual information and joint entropy prior between the anatomical and functional images have also been used further to regulate the global distribution of activity within organs during the reconstruction task.^{27,31}

To optimize the PWLS cost function with anatomically weighted quadratic penalty function, Comtat *et al.*³² used successive overrelaxation (SOR) algorithm, while Dewaraja *et al.*²⁹ used separable paraboloidal surrogates accelerated with ordered subsets (SPS-OS) for a similar penalty function. In general, several efficient algorithms have been proposed for the optimization of PWLS cost function, mostly in the context of transmission tomography, such as SPS-OS,³³ block coordinate descent,³⁴ and preconditioned conjugate gradient (PCG) methods.³⁵ However, there are some issues and considerations in developing optimization algorithms for the PWLS cost function. First issue is the ill-conditioning and shift-variance of its second-order derivative (Hessian) due to its weighting matrix (see Sec. II) and the penalty function, particularly with edge-preserving penalties.³⁶ As a result, the paraboloidal surrogate functions end up with high curvatures³³ and gradient-based algorithms with small step-sizes, leading to slow convergence. In addition, preconditioners used to improve the condition number of the Hessian, and thus to speed up the convergence, would act poorly on the resulting ill-conditioned Hessian matrix.³⁶ Some special shift-invariant weighting matrices and preconditioners have been proposed to deal with this issue, however, through making a number of assumptions

and introducing computational complexities.^{35,37} Another issue arises when employing noncontinuously differentiable penalty functions such as TV. As a result, the PWLS cost function would not be amenable to optimization using conventional gradient-based algorithms. A conventional approach to address the nondifferentiability and the so-called staircasing effect of the TV penalty is to smooth it with a Huber norm.³⁸ Staircasing refers to the formation of artificially flat regions in the reconstructed images. Another approach is variable splitting and dual formulation of TV prior as proposed by Chambolle.³⁹ Recently, Ramani and Fessler⁴⁰ employed variable splitting and augmented Lagrangian methods to reduce the PWLS problem into independent subproblems toward addressing the above issues.

In this study, we describe an approach to improve the convergence rate of existing gradient-based algorithms for the PWLS cost function and also for the first time to enhance the edge-preserving effect of the Huber and TV functions by incorporating additional anatomical boundary information (as weights) into their dual formulations. To address the ill-conditioning of the Hessian matrix due to the penalty function, and also to provide a framework for the dual formulation of weighted TV and Huber priors, we followed a proximal point technique to surrogate the PWLS cost function with a proxy function. The proxy or surrogate was split into a gradient descent and a proximal mapping associated with the priors or regularizers. The convergence rate of the gradient descent step was improved by preconditioning, step size optimization, and ordered subsets; thereby, a proximal preconditioned gradient algorithm accelerated with ordered subsets (PPG-OS) was derived. The performance of the proposed algorithm was studied for three different preconditioners and compared with the SPS-OS algorithm in a simulation study. Finally, the proposed image reconstruction algorithm with boundary weighted regularization was assessed using simulation and clinical studies.

II. THEORY

II.A. The PWLS cost function

Let $x \in \mathbb{R}^N$ be the true activity distribution being observed by a PET imaging system, described by the geometric system matrix $G \in \mathbb{R}^{M \times N}$. Under Poisson photon counting statistics (no dead time or data correction), the negative Poisson log-likelihood for observing the measured prompt coincidences $y^p \in \mathbb{Z}^M$ given x is described by

$$\Phi(x) = \sum_{j=1}^M (n_j a_j [Gx]_j + \bar{r}_j) - y_j^p \log(n_j a_j [Gx]_j + \bar{r}_j), \quad (1)$$

where $[Gx]_j = \sum_{i=1}^N G_{ji} x_i$ is the expected number of counts in j th line-of-response (LOR), and n_j , a_j , \bar{r}_j account for detector efficiency, photon attenuation, and expected random and scatter coincidences in the j th LOR, respectively. The quadratic approximation of $\Phi(x)$ in Eq. (1) around an estimate of the trues at the j th LOR, i.e., \hat{y}_j , yields a WLS cost

function

$$\begin{aligned} \phi(x) &= \frac{1}{2} \sum_{j=1}^M w_j ([Gx]_j - \hat{y}_j)^2, \\ w_j &= \frac{n_j^2 a_j^2}{\max(1, y_j^p)}, \quad \hat{y}_j = \frac{y_j^p - \bar{r}_j}{n_j a_j}, \end{aligned} \quad (2)$$

where \hat{y} is a precorrected emission sinogram. The details of the above derivation are given in Appendix A. To penalize estimates of x deviating from *a priori* knowledge, the cost function in Eq. (2) can be augmented by a penalty or regularizer, $R(x)$, which results in the following penalized WLS cost function:⁹

$$\psi(x) = \frac{1}{2} \|Gx - \hat{y}\|_W^2 + \beta R(x), \quad (3)$$

where $W = \text{diag}\{w_1, \dots, w_M\}$ and $\beta > 0$ is a regularization parameter that controls the balance between the fidelity of Gx to the data and the regularity of x to *a priori* knowledge. In this study, we describe an algorithm for the optimization of PWLS cost function which: (i) allows for efficient implementation of an edge-preserving regularization capable of incorporating anatomical edge information and (ii) possesses an improved convergence rate. We consider a 3D regularizer of the general form

$$R(x) = \sum_{i=1}^{3N} \omega_i \varphi([Dx]_i), \quad (4)$$

where $D = [D_v^T, D_h^T, D_a^T]^T \in \mathbb{R}^{3N \times N}$ is a derivative matrix, composed of first-order finite difference matrices in horizontal, vertical, and axial directions, $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ is a nonconvex potential function, differentiable on $\mathbb{R} \setminus \{0\}$ and the ω_i s are weighting factors that control the strength of regularization between voxels. The conventional quadratic regularization uses $\varphi_Q(t) = t^2/2$, which tends to blur edges by assigning high penalties on large voxel differences. The anisotropic TV regularizer makes use of an absolute value potential function $\varphi_{TV}(t) = |t|$ to preserve sharp edges while suppressing noise.¹⁹ There are, however, two issues with the TV regularizer. First, as this regularizer (with unit weights) is locally edge preserving, it sometimes results in staircasing in the reconstructed images. Second, it is not continuously differentiable because of the nondifferentiability of the absolute value function at points $[Dx]_i = 0$. As a result, the PWLS cost function is not amenable to optimization using conventional gradient-based approaches. To address these issues, the absolute value function is usually smoothed to $|t| \approx \sqrt{t^2 + \epsilon}$ or an anisotropic Huber potential function is employed, which is given by¹⁸

$$\varphi_H(t) = \begin{cases} |t| - \frac{\delta}{2}, & |t| \geq \delta \\ \frac{1}{2\delta} |t|^2, & |t| < \delta \end{cases}, \quad (5)$$

where $\epsilon > 0$ and $\delta > 0$ control the smoothing and thus the edge-preserving effects. Note that when ϵ or $\delta = 0$, the smoothed TV penalty reduces to a nonsmooth TV penalty. TV and Huber regularizations with unit weights detect and preserve edges based on information derived from short-range interactions between neighboring voxels. In this study, we describe an approach to improve the global edge-preservation

property of these regularizers using the weights, ω_i , based on the information captured from the images being regularized²⁰ or from side information of CT or MRI.^{23,29}

II.B. The proposed algorithm

As mentioned earlier, there are two issues in the optimization of PWLS cost function making use of a (weighted) TV or Huber regularization, (i) the ill-conditioning of its Hessian matrix, which ends up with the slow convergence of optimization algorithms and (ii) the noncontinuous differentiability of TV regularizer. The presence of weighting factors can introduce additional shift-variance into the Hessian matrix and thus make the optimization of this cost function more challenging.

To optimize the PWLS function in Eq. (3), we follow a proximal point method,^{41,42} wherein, one iteratively adds a proximal penalty term to an original cost function $\psi(x)$. As a result, an easy-to-optimize proxy or surrogate function, $\psi(x, x^k)$, can be obtained whose optimization guarantees that of the original cost function. In this sense, the proximal point technique follows the same idea of optimization transfer techniques.⁴³ Therefore, we iteratively estimate the solution of the problem (3) by

$$\begin{aligned} x^{k+1} &= \operatorname{argmin}_{x \geq 0} \psi(x, x^k), \\ \psi(x, x^k) &\triangleq \psi(x) + \frac{1}{2} \|x - x^k\|_L^2, \end{aligned} \quad (6)$$

where $L = \pi I_N - G^T W G$ is a positive definite matrix, in which I_N is an $N \times N$ identity matrix, $G^T W G$ is the Fisher information term of the Hessian, $\mathcal{H} = G^T W G + \beta \nabla^2 R(x)$, of the PWLS functional and π satisfies $\pi \geq \lambda_{\max}(G^T W G)/2$. $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue. Note that the surrogate function in Eq. (6) satisfies the monotonicity conditions for surrogates,^{43,44} i.e., it (i) coincides with $\psi(x)$ at point $x = x^k$, (ii) has a matched gradient as $\psi(x)$ at that point, and (iii) lies above this function at other points. By some algebra, the surrogate can be written as

$$\begin{aligned} \psi(x, x^k) &= \frac{\pi}{2} \left\| x - \left(x^k - \frac{1}{\pi} \nabla \phi(x^k) \right) \right\|^2 \\ &\quad - \frac{1}{2\pi} \|\nabla \phi(x^k)\|^2 + \nabla \phi(x^k) + \beta R(x), \end{aligned} \quad (7)$$

where $\nabla \phi(x^k) = G^T W (G x^k - \hat{y})$. Hence, we split the optimization problem in Eq. (6) into two steps: (i) a gradient descent

$$\tilde{x} = x^k - \tau \nabla \Phi(x^k) \quad (8)$$

where $\tau = 1/\pi$ is a step size, and (ii) a proximal mapping associated with the regularizer $R(x)$

$$x^{k+1} = \operatorname{argmin}_{x \geq 0} \frac{1}{2} \|x - \tilde{x}\|^2 + \tau \beta R(x) + C, \quad (9)$$

where C is a constant. This algorithm is known as the proximal gradient algorithm.^{45,46} To improve the convergence rate of this algorithm for PET image reconstruction, we propose to upgrade the step in Eq. (8) to a proximal preconditioned

gradient algorithm accelerated with (i) ordered subsets (PPG-OS) and (ii) step size optimization.

II.B.1. Preconditioning and step size optimization

Generally, an ideal preconditioner, $P \in \mathbb{R}^{N \times N}$, for PG-type algorithms, that aims to optimize the PWLS objective function, would be the inverse of the Hessian matrix such that $P\mathcal{H} = I_N$, where the condition number of the matrix $P\mathcal{H}$, i.e., the ratio of its largest to smallest eigenvalues, is minimal (unity). However, because of the ill-conditioning of the Hessian due to the weighting matrix W and the penalty $R(x)$,³⁵ the exact computation of the inverse of the Hessian is impractical and thus approximations are used. In the applied splitting technique, the Hessian matrix is reduced to $G^T W G$, which is better conditioned than \mathcal{H} , particularly when the regularizer is nonquadratic and edge-preserving. In this work, we employed three different diagonal preconditioners including: $P_1 = \operatorname{diag}\{(G^T W G)_{ii}\}^{-1}$, which uses the inverse of the diagonal elements of the Hessian matrix,⁴⁷ $P_2 = \operatorname{diag}\{G^T W G \mathbf{1}\}^{-1}$, where $\mathbf{1} \in \mathbb{R}^N$ and $P_3 = \operatorname{diag}\{(x^k + \epsilon)/G^T \mathbf{1}\}^{-1}$, where $\epsilon > 0$ ensures the positivity of this EM-type preconditioner and $\mathbf{1} \in \mathbb{R}^M$.⁴⁸ The preconditioner P_2 has been inspired from De Pierro's convexity method to the data fidelity term in Eq. (4) (see Ref. 49). As we extended the step in Eq. (8) to a preconditioned gradient descent, the step size τ should satisfy: $0 < \tau \leq 2/\lambda_{\max}(P G^T W G)$. Proof follows from Theorem 8.3 in Ref. 50. To achieve the maximum amount of decrease of the cost function, the step size can be obtained by the following maximization:

$$\tau^k = \operatorname{argmax}_{\tau \geq 0} \phi(x^k - \tau P \nabla \phi(x^k)), \quad (10)$$

whereby, an optimal step size is obtained by

$$\tau^k = \frac{(P \nabla \phi(x^k))^T \nabla \phi(x^k)}{(P \nabla \phi(x^k))^T G^T W G (P \nabla \phi(x^k))}. \quad (11)$$

II.B.2. Proximal mapping

To solve the *primal* problem (9) for a (nondifferentiable) weighted TV and weighted Huber and penalty functions, we cast it into its *dual* problem by making use of an extended dual formulation of these penalty functions. First, we notice that the weighted absolute value of a scalar $x \in \mathbb{R}$ with the weight $\omega > 0$ can be defined as

$$\omega|x| = \max\{zx : |z| \leq \omega\}. \quad (12)$$

Hence, using the above fact, one can define a weighted TV function in its dual formulation by: $R_{\text{TV}}^\omega(x) = \sum_{i=1}^{3N} \max_{z \in Q_i} \{z[Dx]_i\}$, where $Q_i = \{z \in \mathbb{R} : |z| \leq \omega_i\}$ is a dual feasible set. Since the Huber penalty function generalizes the TV one, we define the weighted version of Huber regularizer in duality context by^{51,52}

$$\begin{aligned} R_{\text{H}}^\omega(x) &= \max_{z \in Q} \left\{ z^T D x - \frac{\delta}{2} z^T z \right\}, \\ Q &= \{z \in \mathbb{R}^{3N} : |z_i| \leq \omega_i, \quad \forall i = 1, \dots, 3N\}, \end{aligned} \quad (13)$$

where the quadratic term $(\delta/2)z^T z$ accounts for the smoothness of the Huber penalty such that $\delta = 0$ reduces it to a weighted TV penalty. Using this formulation, the nondifferentiability of the TV function is transferred to the set Q , which can be easily addressed during the proximal mapping. By plugging Eq. (13) into Eq. (9), one arrives at the following primal-dual problem, which is convex in the primal variable x and concave in the dual variable z :

$$\min_{x \geq 0} \max_{z \in Q} \left\{ \frac{1}{2} \|x - \tilde{x}\|^2 + \tau \beta \left(z^T D x - \frac{\delta}{2} z^T z \right) \right\}. \quad (14)$$

To solve this problem, we exchange the order of the maximization and minimization using the min-max property, reduce $\tau \beta$ to β , and minimize the problem with respect to x . Substituting the resulting solution $x = \tilde{x} - \beta D^T z$ into the max problem gives

$$\max_{z \in Q} \left\{ -\frac{\beta^2}{2} z^T D D^T z + \beta z^T D \tilde{x} - \frac{\beta \delta}{2} z^T z \right\}. \quad (15)$$

We address the non-negativity on this solution later within the body of the algorithm (Algorithm 1, step 6a). The above problem is then cast into the following minimization by multiplying its objective with $-1/\beta^2$ and completing the square:

$$\min_{z \in Q} \left\{ \frac{1}{2} \|D^T z - \tilde{x}/\beta\|^2 + \frac{\delta}{2\beta} \|z\|^2 + C \right\}. \quad (16)$$

To find the solution of the above constrained problem, we follow the proximal point method by adding the proximal penalty $\frac{1}{2} \|z - z^n\|_{\Sigma}^2$ to its objective, where $\Sigma = \alpha I_{3N} - D D^T$, I_{3N} is a $3N \times 3N$ identity matrix and the parameter α should satisfy $\alpha \geq \lambda_{\max}(D D^T)/2$. By some algebra, dropping constant terms and letting $\sigma = \beta/\alpha$, the minimizer of the problem then reads:

$$z^{n+1} = \operatorname{argmin}_{z \in Q} \left\{ \frac{1}{2} z^T z - b^T z \right\}, \quad (17)$$

$$b = \frac{1}{\delta + \sigma} (\sigma z^n + D(\tilde{x} - \beta D^T z^n)).$$

As the above problem is separable, one can solve it componentwise subjected to the constraint $|z_i| \leq \omega_i$. One can show that for $|b_i| \leq \omega_i$, the solution is $z_i^{n+1} = b_i$, and for $|b_i| > \omega_i$, the solution reads $z_i^{n+1} = \omega_i \operatorname{sign}(b_i)$. As a result, the solution of Eq. (19) is vector b projected to the subset Q as follows:

$$z_i^{n+1} = \operatorname{proj}_Q(b_i) = \begin{cases} \omega_i \operatorname{sign}(b_i), & |b_i| > \omega_i \\ b_i, & |b_i| \leq \omega_i \end{cases}. \quad (18)$$

To this end, Algorithm 1 summarizes the proposed PPG algorithm, whose convergence rate was improved by ordered subsets. Note that subsetization is employed within steps 1 through 4 and hence step size optimization is carried out for each subset.

ALGORITHM I. PPG-(OS) algorithm.

Initialize: $x^0 = x^1 = \mathbf{0}$, $z^0 = D x^0$, $\delta, \alpha, \beta, \sigma = \frac{\beta}{\alpha}$, $T, k = n = 0$

While (stopping criterion is not met) **do**

1. $\nabla \Phi = G^T W(G x^k - \hat{y})$
2. $p^k = P \nabla \Phi$
3. $\tau^k = (p^k)^T \nabla \Phi / (p^k)^T G^T W G(p^k)$
4. $x^{k+1} = x^k - \tau^k p^k$.
5. $\tilde{x} \leftarrow x^{k+1}$.
6. **While** $n \leq T$ **do**
 - a. $x^n = [\tilde{x} - \beta D^T z^n]_+$.
 - b. $b = \frac{1}{\delta + \sigma} (\sigma z^n + D x^n)$.
 - c. $z_i^{n+1} = \operatorname{proj}_Q(b_i), i = 1, \dots, N$.
 - d. $n = n + 1$
7. $x^{k+1} \leftarrow x^n$
8. $k = k + 1$.

Output: $\hat{x} \leftarrow x^{k+1}$

In this algorithm, the non-negativity constraint declared in problem (14) is imposed on the primal variable x^n on step 6a, where $[t]_+ = \max(0, t)$ is a non-negativity operator. In this study, we approximated the largest eigenvalue of the matrix $D D^T$ using power iteration method,⁵³ which was found to be less than 8 in 2D regularization and 12 in the 3D case. Therefore, in Algorithm 1, one can initialize the parameter $\alpha \geq 4$ in 2D and 6 in 3D cases. In the thresholding scheme in Eq. (18), the boundary weights $\{\omega_1, \dots, \omega_{3N}\} \in [a, 1]$, $0 < a < 1$, are chosen such that lower weights are assigned to the smoothing of boundary regions and thereby the global edge-preserving effect of the resulting algorithm is improved. The weights can be derived from anatomical side information or from the image being regularized. In this study, we derived weighting factors by the following equation:⁵⁴

$$\omega_i = S_a \left\{ \frac{1}{1 + \operatorname{Edge}(x)} \right\}, \quad i = 1, \dots, N, \quad (19)$$

where $S_a(t)$ is a scaling function that maps the range of t to $[a, 1]$ and $\operatorname{Edge}(\cdot)$ is an edge detector. We used 2D Canny edge detector⁵⁵ using MATLAB's *edge* function with automatic threshold selection. In Eq. (4), three weighting factors were defined for the gradient of each voxel in horizontal, vertical, and axial directions. The weights given by Eq. (19) were therefore concatenated to obtain such directional weights at each voxel. In all experiments, we set $a = 0.01$, since noise can reside on edges and therefore nonzero weights are prudent to suppress noise at edges.

In this study, we compared the convergence rate of the proposed PPG-OS algorithm with the SPS-OS algorithm in the optimization of the PWLS cost function. The details of the implementation of SPS-OS algorithm are given in Appendix B.

III. METHODS

The performance of the proposed algorithm was evaluated using simulated and clinical studies. In this section, we describe the evaluation procedure of the algorithm and provide the methods and definitions relevant to the studies.

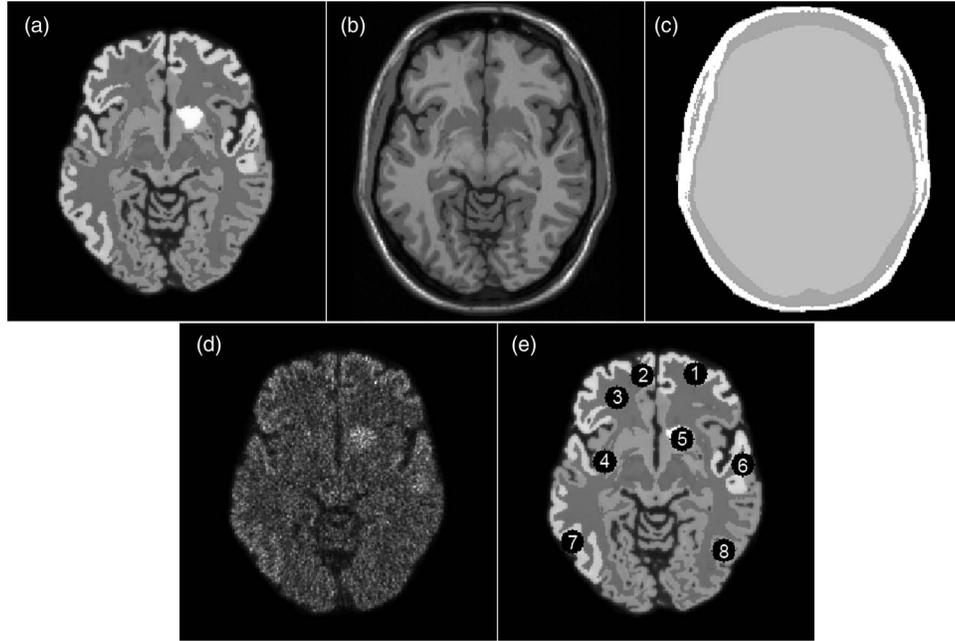


FIG. 1. (a) The true activity distribution derived from (b) a transaxial T1-weighted brain MRI together with (c) its corresponding MR-derived attenuation map. (d) The image reconstructed from a given noise realization using the standard OSEM (15 iterations, 6 subsets) algorithm. (e) The regions of interest used for quantitative bias-variance evaluation of the proposed algorithm with anatomically weighted regularization.

III.A. Simulation study

For the purpose of realistic simulations, we simulated a typical activity distribution with several metabolically active regions and a lesion using a 256×256 T1-weighted brain MR image. The activity concentration in the gyrus and the tumor were assigned an average tumor-to-background ratio of 1.6 and 1.9, respectively. The simulated activity distribution and the MR images together with a corresponding MR-derived attenuation map are shown in Figs. 1(a)–1(c). The attenuation map is composed of linear attenuation coefficients of bone (0.120 cm^{-1}), soft tissue (0.096 cm^{-1}), and air (0 cm^{-1}). The MR image was further used to derive anatomical boundaries for the weighted regularization method elaborated in Sec. II.B.2. The projection data of the simulated activity distribution were computed analytically for a PET scanner with parallel strip-integral geometry (4.06 mm strip width, 258 radial samples, and 404 angular samples uniformly spaced over 180°). The system matrix describing this scanner was generated using the Image Reconstruction Toolbox (IRT) developed by Fessler⁵⁶ and the images reconstructed with image resolution of 256×256 and pixel size of $1.94 \times 1.94 \text{ mm}^2$. Following the attenuated forward projection of the simulated activity distribution, 10 Poisson noise realizations were simulated with a total count of 6×10^6 and 10% random coincidence rate. Figure 1(d) shows the image reconstructed from a given noise realization using the standard OSEM (15 iterations and 6 subsets) algorithm. As expected, the conventional OSEM algorithm enforces maximal consistency between the reconstructed image and the noisy emission data; thereby, the final image tends to be noisy.

In the simulation study, we objectively compared the proposed PPG-OS algorithm (using the preconditioners P_1 , P_2 ,

and P_3^k) with the SPS-OS algorithm for six ordered subsets. All reconstructions were initialized with a zero initial image. For the evaluation of the algorithms, we consider Huber regularization with the smoothing parameter $\delta = 0.5$ for two reasons: (i) to avoid the staircasing artifacts resulting from nonsmooth TV regularization and (ii) to make the resulting PWLS cost function differentiable and thus amenable to optimization using the SPS-OS algorithm. In all reconstructions using the proposed algorithm, the proximal mapping of the regularizer (step 6 in Algorithm 1) was performed by $T = 5$ iterations, which was found to be enough for the regularization step. As presented in Algorithm 1, a global convergence was declared when the relative change of the image estimate between two successive iterations falls below a tolerance of $\eta = 5 \times 10^{-4}$.

To assess the convergence rate of the algorithms, we heuristically chose three regularization parameters that resulted in the highest signal-to-noise ratio (SNR) between the actual image, x^* , and the reconstructed image \hat{x} after convergence. The SNR is defined as

$$\text{SNR} = -20 \log \left(\frac{\|\hat{x} - x^*\|}{\|x^*\|} \right). \quad (20)$$

In the next step, the impact of boundary-weighted regularization on the convergence rate of the PPG-OS algorithm as well as image quality was, respectively, evaluated in terms of SNR and bias-variance trade-off over the simulated noise realizations. We followed a region-of-interest (ROI) based approach to assess the bias vs variance performance of the proposed algorithm. As shown in Fig. 1(e), 8 ROIs ($\sim 15 \text{ mm}$ in diameter) were drawn over different regions of the designed brain phantom. For each ROI, bias was calculated by

$$\text{Bias} = \frac{|\bar{x} - \bar{x}^*|}{\bar{x}^*} \times 100, \quad (21)$$

where $\bar{x} = \frac{1}{m} \sum_{j=1}^m \hat{x}^{(j)}$, $\hat{x}^{(j)} = \frac{1}{n} \sum_i x_i^{(j)}$, and $\bar{x}^* = \frac{1}{n} \sum_i x_i^*$, m is the number of noise realizations, n is the number of voxels in the ROI, $x^{(j)}$ is the image reconstructed from j th noise realization, and x^* is the true object distribution. The variance was also obtained by normalized standard deviation (NSD) for each ROI

$$\text{NSD} = \frac{1}{n} \sum_{i=1}^n \frac{\sqrt{\frac{1}{m-1} \sum_{j=1}^m (x_i^{(j)} - \bar{x}_i)^2}}{\bar{x}_i}, \quad (22)$$

where $\bar{x}_i = \frac{1}{m} \sum_{j=1}^m x_i^{(j)}$ represents the ensemble mean value of voxel i . For each ROI, bias was plotted against NSD values as a function of iteration number and the performance of weighted regularization was compared with the conventional regularization.

III.B. Clinical PET study

The proposed PPG-OS-P2 algorithm was employed for image reconstruction of a clinical brain PET/CT scan. The patient was administrated with 205 MBq of ^{18}F -FDG and a dynamic study (60 min) was acquired on the Biograph 64 True Point PET/CT scanner (Siemens Healthcare, Erlangen). The PET scanner operates in fully 3D mode and records prompt and random coincidences in 32-bit list-mode packets, which include the sinogram bin address of each event. Based on bin addresses, the data were histogrammed for a 15 min (0–15 min) static study ($\sim 78 \times 10^6$ prompts) using an in-house list-mode histogrammer and thereby the prompts, randoms, and net true sinograms were generated with a size of $336 \times 336 \times 313$. Standard corrections were applied on the net-true sinograms (attenuation, normalization, and scatter) to obtain the precorrected sinogram \hat{y} . The data were then reconstructed in 3D mode using the PPG-OS-P2 algorithm with conventional Huber and weighted Huber regularizations. Siddon's algorithm⁵⁷ was used to derive a line integral-based geometric system matrix for the Biograph True Point scanner. A hybrid of system matrix precomputation and inplane and axial symmetry translations was utilized for a fast and memory efficient image reconstruction. At the end, the preconditioner P_2 was precomputed and an image volume of $336 \times 336 \times 81$ size was reconstructed using the proposed algorithm (8 iterations and 14 subsets). For weighted regularization, we followed an approach in which the boundary weights were iteratively derived from the image estimates during the reconstruction using Eq. (19).

IV. RESULTS

IV.A. Simulation studies

IV.A.1. Convergence rate

Table I summarizes the SNR performance of the studied algorithms after their convergence as well as the number of required iterations for the ten noise realizations (ten-trial) and a single noise realization (single-trial). As presented, the algorithms depict nearly similar SNR performance for the chosen regularization parameters; however, the proposed

TABLE I. The quantitative performance of the studied algorithms for ten noise realizations (ten-trial) and a single noise realization (single-trial) of simulated PET emission data.

Algorithm	β	Ten-trial		Single-trial	
		SNR (average)	Iterations (range)	SNR	Iterations
PPG-OS-P1	0.02	17.45	23–25	17.49	23
	0.04	16.76	17–18	16.85	18
	0.06	16.08	15–16	16.17	15
PPG-OS-P2	0.02	17.54	22–24	17.57	23
	0.04	16.87	17–18	16.97	17
	0.06	16.19	15	16.27	15
PPG-OS-P3	0.02	17.33	23–28	17.42	28
	0.04	16.45	18–19	16.52	18
	0.06	15.88	15–17	15.98	16
SPS-OS	0.2	17.77	57–59	17.92	59
	0.4	16.85	50–52	16.77	51
	0.6	16.75	46–48	16.36	47

PPG-OS algorithm achieves its SNR performance in a considerably fewer number of iterations in comparison with the SPS-OS algorithm. Furthermore, the ten-trial average performance shows that the proposed algorithm with the preconditioner P_2 results in a higher SNR and a lower range of iterations in comparison with the other two preconditioners.

Figures 2(a) and 2(b) keep track of the convergence rate of the algorithms by the log of the relative error $\xi = \|x^{k+1} - x^k\|/\|x^k\|$, which was used as a stopping criterion for the algorithms meeting the tolerance of 5×10^{-4} . The convergence curves are shown for the single-trial results in Table I and for two different regularization parameters. It turns out that the proposed algorithm achieves a considerably improved convergence rate over the SPS-OS algorithm and depicts the fastest convergence rate with the preconditioner P_2 . The PPG-OS-P1 and PPG-OS-P2 algorithms present generally with similar convergence rates, while PPG-OS-P3 lags behind at early iterations and approaches them at the last iterations. The same trend was almost observed for the other noise realizations. It should be emphasized that the convergence behavior of the PPG-OS-P3 algorithm highly depends on the parameter ϵ used to ensure the positivity of the preconditioner. In this study, we set $\epsilon = 1 \times 10^{-4}$ and observed that larger values of ϵ improve the initial convergence of the resulting algorithm, however, they could not ensure a convergence rate as fast as smaller values. Figures 2(c) and 2(d) show the performance of the algorithms in terms of the relative error between two successive estimates as a function of computation (CPU) time. The results show that after a CPU time of 120 s, the PPG-OS algorithm reduces the relative error more than the SPS-OS algorithm and hence provides a faster convergence. In the implementation and comparison of both algorithms, the same forward and backward projector as well as the same number of subsets were used. It was found that the average computation time per iteration of our algorithm is 2.92 s while it is 1.45 s in the case of SPS-OS. This increased computation time is mostly due to step size optimization and iterative proximal mapping.

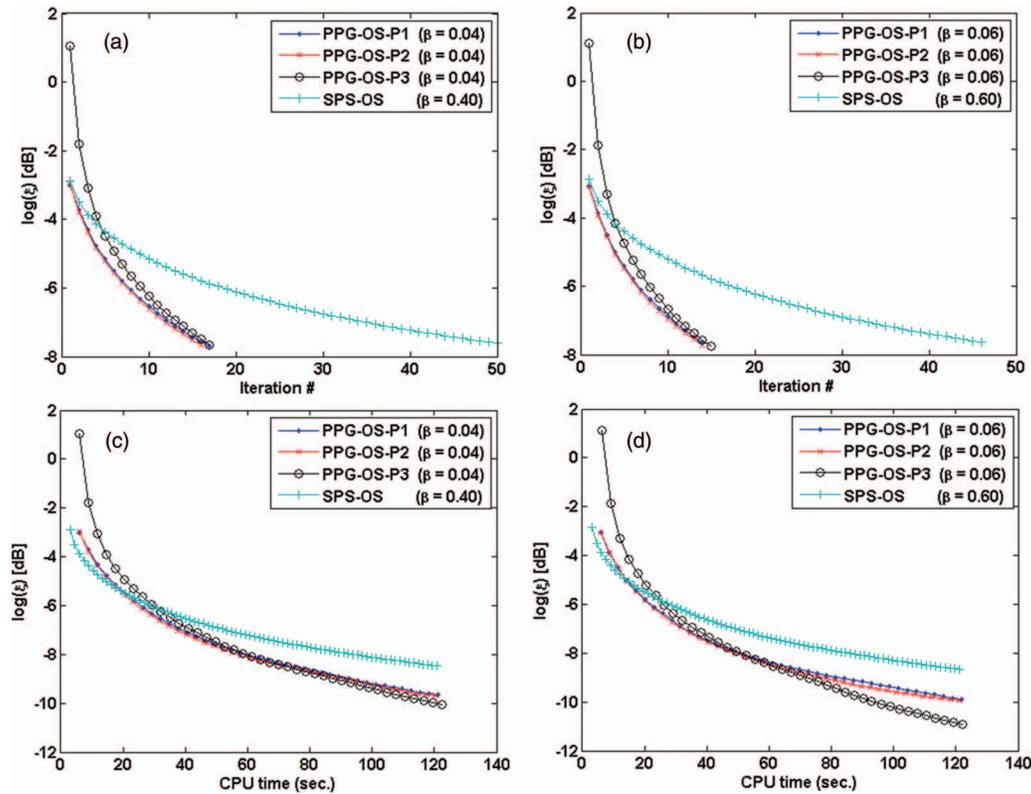


FIG. 2. (a) and (b) The convergence of the studied algorithms for different regularization parameters as a function of iteration number and (c) and (d) computation (CPU) time. The vertical axis is in logarithmic scale.

Figure 3 shows the evolution of the image solution with iteration number for the studied algorithms. Consistent with Table I and Fig. 2, the results show the improved convergence rate of the proposed algorithm, particularly with the preconditioners P_1 and P_2 , over the SPS-OS algorithm. It can be observed that after two iterations, PPG-OS-P1 and PPG-OS-P2 algorithms estimate the hot and cold spots faster than their counterpart. The last column in this figure shows the image estimates after convergence for the chosen regularization parameters.

For image reconstruction of the clinical study presented in Sec. IV.B, we utilized the proposed PPG-OS algorithm with the preconditioner P_2 because (i) it showed a higher SNR and better convergence rate in comparison with the other two preconditioners according to simulation results, and (ii) contrary to preconditioner P_1 , preconditioner P_2 does not need direct access to the Hessian matrix, and hence the precomputation of the system matrix. For clinical datasets, the system matrix is a very large-sized matrix that raises memory shortage issues. As a result, P_2 can be precomputed using on-the-fly forward and backprojections. Generally, the computation of an optimal step size [Eq. (11)] for large-sized clinical data increases remarkably the overall reconstruction time of the proposed subsetized algorithm, because of additional forward projection of each descent direction, $P\nabla\phi(x^k)$, in each subset. To reduce computation time, we designed an experiment to find a

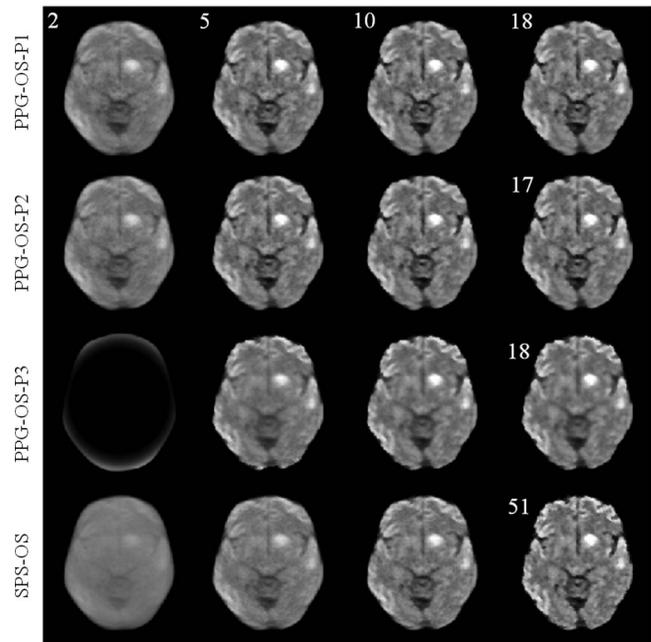


FIG. 3. The evolution of images reconstructed by the studied algorithms as a function of iteration (for $\beta = 0.04$ and 0.4 for PPG-OS and SPS-OS, respectively). The last column shows the image estimates after convergence. As can be seen, the proposed algorithm with the preconditioners P_1 and P_2 show an improved initial and overall convergence. The display window is the same for all images.

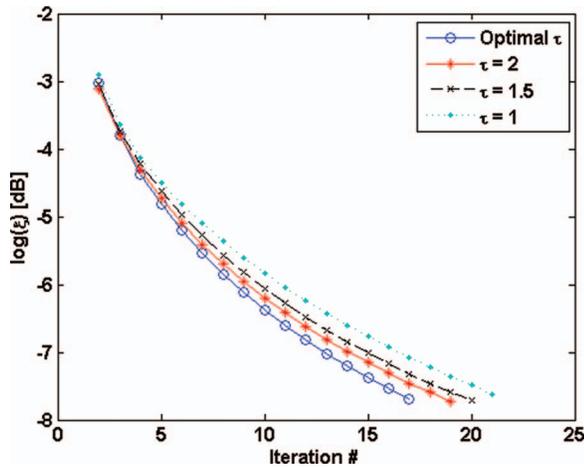


FIG. 4. The convergence rate of the PPG-OS-P2 algorithm for optimal and fixed step lengths as a function of iteration number ($\beta = 0.04$). The results show that step size $\tau = 2$ is near to optimal for this algorithm.

fixed, near-to-optimal step size for the PPG-OS-P2 algorithm according to the condition $0 < \tau^k \leq 2/\lambda_{\max}(PG^T WG)$. Using the power iteration method, it was found that the largest eigenvalue of the $P_2G^T WG$ matrix is close to unity for all noise realizations and for a strip- or line-integral geometry, i.e., $\lambda_{\max}(P_2G^T WG) \cong 1$. As a result, one can declare $0 < \tau \leq 2$. Figure 4 compares the convergence rate of the PPG-OS-P2 algorithm for optimal and fixed step sizes and the same noise realization and regularization parameter ($\beta = 0.06$). The results show that as the step size increases up to 2, the convergence rate of the resulting algorithm approaches the algorithm with optimal step sizes. The same trend was observed with other noise realizations and regularization parameters. As a result, with the step size $\tau = 2$, one can considerably save computation time, yet exploit a favorable convergence rate.

IV.A.2. Regularization with boundary side information

To incorporate anatomical side information into Huber regularization in the form of weighting factors, we derived the anatomical boundaries from the MR image shown in Fig. 1(b) using Canny edge detector with automatic threshold selection according to Eq. (19). The images reconstructed using the PPG-OS-P2 algorithm with and without boundary weighting are shown in Fig. 5. The first row shows the actual activity distribution and the images reconstructed using conventional Huber regularizer with regularization parameters $\beta = 0.04$ and 0.06, respectively. The second row shows, respectively, the derived weighting factors and the image reconstructed using weighted Huber regularization with the above mentioned β parameters. The visual comparison of the images reveals that the incorporation of boundary information in the reconstruction process improves the accuracy of the estimated images. As can be seen, the weights tend to suppress the smoothing effect of the regularization across edges and therefore to increase image contrast.

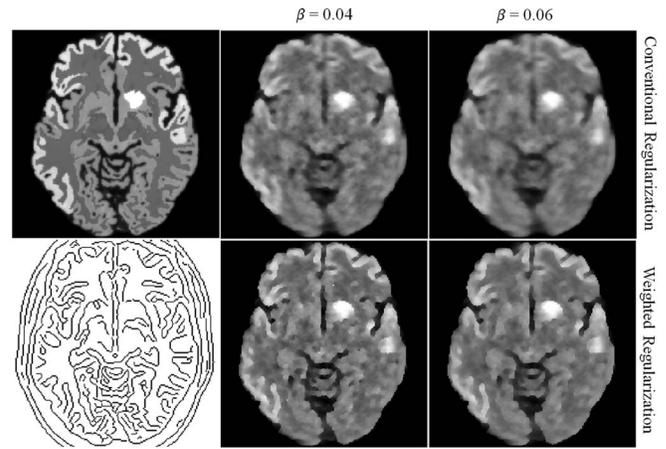


FIG. 5. Huber regularization with anatomical boundary information. First row (from left to right) shows the true activity distribution and the images reconstructed using the PPG-OS-P2 with conventional regularization and regularization parameters of $\beta = 0.04$ and 0.04. Second row (from left to right) shows the boundary weights used in the regularization and the images reconstructed using the PPG-OS-P2 algorithm with weighted regularization and the same β s. The display window is the same for all images.

Figure 6 compares line profiles through the true activity distribution and the images reconstructed with and without boundary weighted regularization. The profiles show that the weighted regularization (R_W) increases the contrast particularly in regions where the conventional unweighted regularization (R) reduces contrast. Note that the metabolically active regions with weak boundary weights or without any corresponding anatomical boundary (e.g., the hot tumor) depict qualitatively nearly the same line profiles in the reconstructed images. Note that the edge enhancement in R_W profiles depends on the strength of edges; thereby, strong edges are more enhanced than the weak ones.

To further quantitatively compare the regularization techniques, we evaluated their bias-variance performance for the ROIs shown in Fig. 1(e) over ten noise realizations of the emission data reconstructed using the PPG-OS-P2 algorithm. Figure 7 shows the bias-variance performance of the regularization techniques as a function of iteration number for regularization parameters of 0.04 and 0.06. Note that each point corresponds to one iteration and six subsets. The results show that the incorporation of the anatomical boundary information in the regularization substantially reduces the bias, thus allowing for more accurate estimation of the true activity distribution; however, it increases the noise. The regions with weak boundaries or without any corresponding anatomical boundaries (e.g., ROIs 3 and 5) show an improved bias-variance performance, while the regions with strong boundaries (e.g., ROIs 2 and 6) depict a substantial improvement in bias. The plots also show that in this simulation study, the regularization parameter $\beta = 0.06$ generally decreases the variance while increasing the bias. Therefore, it can be concluded that weighted regularization with a higher regularization parameter can give rise to a better bias-variance trade off.

We also evaluated the impact of the weighting scheme on the SNR and convergence performance of the proposed

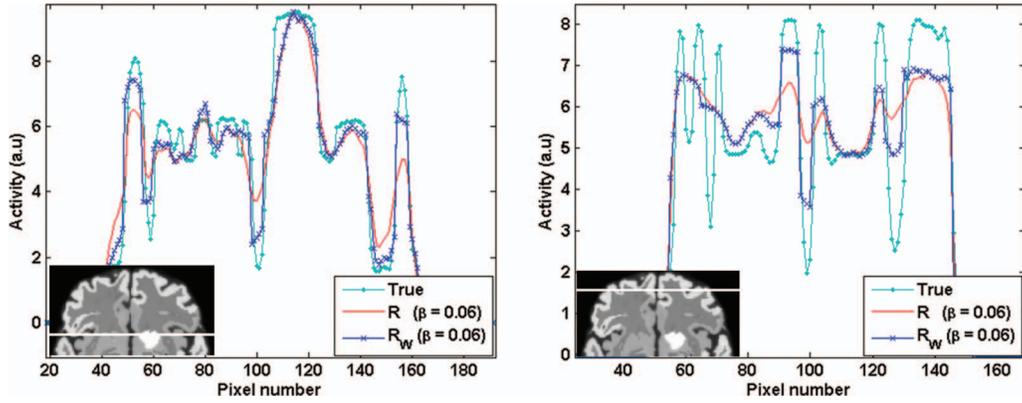


FIG. 6. Activity line profiles through the true activity distribution and the images reconstructed using the PPG-OS-P2 algorithm with conventional regularization (R) and anatomically weighted regularization (R_w), along the lines shown in subfigures.

PPG-OS algorithm. Table II presents the results for the preconditioners P_1 , P_2 , and P_3 and the ten noise realizations (ten-trial) and a single noise realization (single-trial). In comparison with Table I, the results show that the incorporation of the anatomical boundary information in the regularization step reduces the convergence rate of the resulting algorithms, however, prudently improves their SNR performance for $\beta = 0.04$ and 0.06 . In the case of $\beta = 0.02$, both SNR and convergence rate decrease, which should be attributed to the fact that the noise tends to reside on edges where regularization is suppressed [according to the lower bound a in Eq. (19)] and

this low β is not sufficient to remove the noise and hence to establish the convergence path of the algorithm.

IV.B. Clinical PET study

Figure 8 shows the transverse, coronal, and sagittal planes of the clinical study reconstructed using the PPG-OS-P2 algorithm with conventional (unweighted) and weighted Huber regularizations ($\beta = 0.04$, $\delta = 0.08$) together with the weighting factors at the last iteration. The algorithm has successfully reconstructed this precorrected PET study at a typical

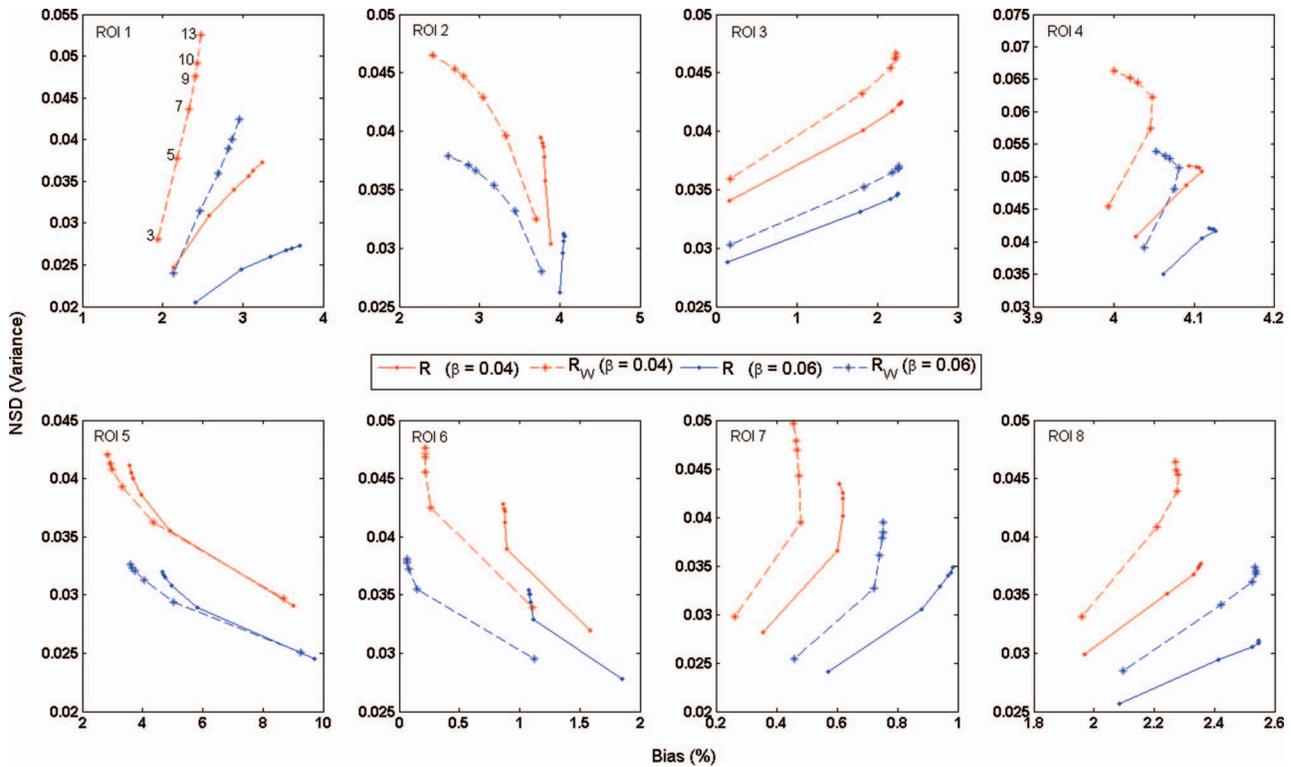


FIG. 7. The bias versus variance trade-off of anatomically weighted Huber regularization (R_w) in comparison with conventional Huber regularization (R) as a function of iteration number and regularization parameters. The trade-offs were obtained for eight ROIs drawn on different regions of the brain phantom [see Fig. 1(e)] and calculated from ten noise realizations of simulated PET emission data reconstructed using the PPG-OS-P2 algorithm.

TABLE II. The quantitative performance of the proposed algorithm with anatomically weighted Huber regularization for ten noise realizations (ten-trial) and a single noise realization (single-trial) of simulated PET emission data.

Algorithm	β	Ten-trial		Single-trial	
		SNR (average)	Iterations (range)	SNR	Iterations
PPG-OS-P1	0.02	16.65	39–43	16.71	39
	0.04	17.24	26–27	17.26	27
	0.06	16.81	22–23	16.87	23
PPG-OS-P2	0.02	16.65	39–44	16.62	39
	0.04	17.38	25–26	17.41	26
	0.06	16.96	22–23	17.03	22
PPG-OS-P3	0.02	16.78	53–68	16.89	53
	0.04	17.10	35–62	17.21	35
	0.06	16.65	26–37	16.74	35

noise level encountered in clinical setting. It can be observed that the incorporation of boundary weighting factors during reconstruction improves the contrast and hence the quality of the reconstructed images. As mentioned in Sec. III.B, for this clinical study, the weighting factors were iteratively derived for each axial slice during reconstruction by means of MATLAB's 2D Canny edge detection operator. This operator has properly captured the inplane topological features of the brain and thus resulted in contiguous boundaries mostly in the transverse plane. However, as can be seen in the coronal and sagittal planes displayed in Fig. 8, the derived weighting factors have also considerably improved the image contrast in these planes.

V. DISCUSSION

In this study, we developed an algorithm for the optimization of PWLS cost function in PET image reconstruction with an improved convergence rate and described the capability of

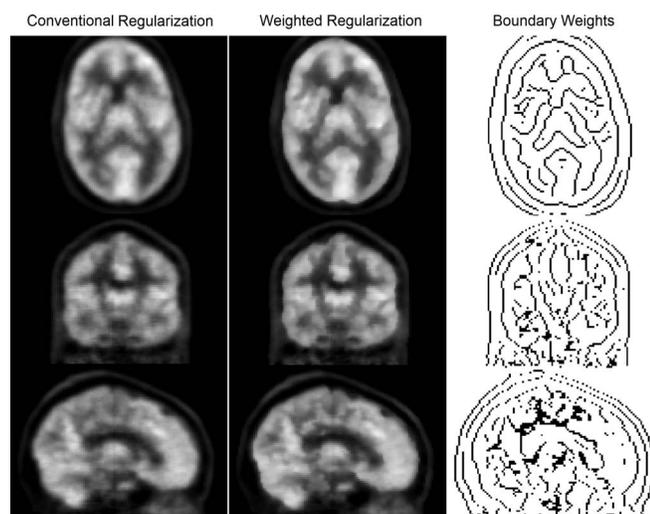


FIG. 8. PET images (transverse, coronal, and sagittal slices) of the clinical brain study reconstructed using the proposed PPG-OS-P2 algorithm with Huber and weighted Huber regularizations. The boundary weighting factors derived from the last iteration are also shown.

the algorithm to incorporate anatomical edge information into the regularization step exploiting total variation and Huber regularizations.

V.A. Convergence rate and parameter selection

In principle, an iterative algorithm is globally convergent if for any arbitrary initial estimate the algorithm is guaranteed to converge to a solution where the gradient of the cost function is zero.⁵⁰ However, in practice, an iterative algorithm is said to be convergent if it reaches a solution where the image estimates do not change (or practically change within a tolerance) with succeeding iteration.⁵⁸ The convergence rate is often referred to as the number of iterations it takes the algorithm to reach a convergent solution. In this study, it was shown that the proposed algorithm depicts a fast convergence rate in the optimization of PWLS cost function. The average SNR results in Table I show that for the same regularization parameter, the proposed algorithm, and the SPS-OS algorithm converge roughly to the same solution. However, the range of iterations show that the proposed algorithm converges in a fewer number of iterations. Generally, the convergence rate of gradient-based algorithms depends highly on the condition number of the Hessian matrix of the cost function.⁵⁰ In this regard, preconditioners are used to improve the conditioning of the problem. The Hessian matrix of the PWLS cost function is ill-conditioned due to the weighting matrix and the penalty function. As elaborated in Sec. II.B, we split the optimization of PWLS into two steps. This procedure reduced the Hessian matrix of PWLS to its Fisher information matrix and hence removing the additional shift variance induced by the regularizer. As a result, the preconditioners can more effectively approximate the inverse of the resulting Hessian matrix, and therefore, the order of convergence of the resulting algorithm is improved. According to Eq. (B1), in Appendix B, the SPS-OS algorithm has also a preconditioned gradient form. In this algorithm, the Hessian matrix, however, includes the second-order derivative of the regularizer and therefore the preconditioner should approximate the inverse of this matrix with a comparatively higher condition number. Generally, the preconditioners increase the convergence rate of the algorithm and in principle should not have an effect on the final solution.³⁵ The results shown in Tables I and II show that the proposed algorithm with the studied preconditioners achieves nearly the same average SNR performance over 10 Poisson noise realizations. In accordance with Fig. 3 (right column), this implies that the proposed algorithm with the studied preconditioners converges to nearly the same solution. The results, however, show that the preconditioner P_2 , inspired by De Pierro's convexity trick, depicts the best convergence rate. It should be noted that the SPS-OS algorithm exploits a preconditioner with the same structure [see Eq. (B1)].

The convergence properties of an optimization algorithm depend on the choice of involved parameters. The results presented in Table I show that by increasing the regularization parameter β , the convergence rate of the algorithms is improved. This is mainly due to a higher regularization and faster

stabilization of the solution in terms of noise and fluctuations between two successive iterations. In this work, we studied the edge-preserving Huber prior, which reduces the staircasing of the TV prior. In the simulation study, we experimentally set the smoothing parameter $\delta = 0.5$ to avoid the staircasing while in the clinical dataset we used $\delta = 0.08$. Generally, this parameter should be chosen according to the level of noise and the intrinsic smoothing effect of basis functions used in the discretization of activity distribution. In noisy datasets, Huber regularization with lower values of δ tends to preserve false details formed by noise, particularly with rectangular basis functions (voxels) that do not possess an intrinsic smoothing effect. Hence, the parameter δ should be increased accordingly to effectively reduce noise and staircasing effect. The convergence rate of algorithms such as SPS-OS depends on δ , since the condition number of the problem is inversely related to this parameter (see Theorem 3.1 in Ref. 59). This implies that the convergence rate of the proposed splitting based algorithm should, in principle, be independent of δ . The PPG-OS algorithm also includes two other parameters, α and T , which are, respectively, the largest eigenvalue of the DD^T matrix (see Sec. II.B.2), and the number of subiterations in the proximal mapping (see Algorithm 1). The parameter α was set to 5 (larger than 4) in the 2D regularization in the simulation study and 7 (larger than 6) in the 3D regularization for clinical studies. It was found that setting the values of this parameter lower than the abovementioned limits results in artifacts and in some cases the divergence of the algorithm. The number of subiterations T was set to 5 in both simulation and clinical studies. Generally, this parameter implicitly controls the amount of regularization and can potentially impact the convergence rate; however, we found that a higher number of subiterations has the same impact as the abovementioned value. Finally, the number of ordered subsets was set to six and kept constant in our comparisons. Increasing the number of subsets can in principle improve the convergence rate of the proposed algorithm and generally the algorithms exploiting the subsetization.

V.B. Incorporation of anatomical edge information

The major motivation for the incorporation of boundary information, derived from CT and MR images^{28,29} or from the emission/transmission image being regularized²⁰ into the regularization step is to promote the formation and preservation of continuous boundaries and therefore to improve image quantification and tumor delineation. In this study, we showed that the proposed algorithm can be extended to incorporate such information into the dual formulation of the edge-preserving TV and Huber regularizer according to our observations in Eq. (12). The Canny edge detection operator was used to derive such boundary information from a MR image, in our simulation study, and from the emission image, in our clinical study. The comparison of the results shown in Tables I and II demonstrates that the average SNR with weighted regularization is improved, however, at the expense of increased convergence rate. This is due to the fact that noise also resides in boundary regions and reduced values

of a in Eq. (19) tend to preserve noise in those regions. The bias-variance evaluations in the simulation study showed that a boundary-weighted Huber regularization noticeably reduces bias; however, it can increase variance particularly in regions with strong boundaries. In general, an improved bias-variance trade-off can be achieved by adjusting the parameter a as well as the regularization parameter β . Canny's edge detection method is based on the assumption that in regions where the amplitude of image gradient is high, an edge probably exists. It exploits hysteresis thresholding to detect edges and thus is somewhat robust to noise. However, at higher noise levels, it might pick up false edges. To enhance the robustness of this edge detector, particularly for the task of simultaneous edge-detection and image reconstruction, the procedure can be improved by edge detection methods based on local mutual information.⁶⁰ In this study, we introduced the PPG-OS algorithm and demonstrated its improved convergence rate as well as its ability to incorporate the anatomical boundary information. Future work is therefore to evaluate the algorithm with more adaptive edge detection techniques in clinical PET/MR datasets.

VI. CONCLUSION

In this study, we proposed a splitting-based preconditioned conjugate gradient algorithm accelerated with ordered subsets for the PWLS reconstruction of PET data. The splitting allowed for decoupling of the penalty function from the PWLS cost function, thereby (i) the ill-conditioning of its Hessian and thus convergence rate of the resulting algorithm was improved and (ii) a proximal subproblem was obtained for implementing adaptively or anatomically weighted TV and Huber regularizations in their dual formulations. The proposed algorithm was studied for three different diagonal preconditioners and compared with the SPS-OS algorithm. In simulation studies, it was demonstrated that the proposed algorithm achieves a considerable improved convergence rate over its counterpart. The performance of the proposed algorithm with boundary-weighted regularization was also evaluated using both simulation and clinical studies. It was concluded that the proposed PPG-OS algorithm is efficient in terms of convergence rate and capable of incorporating boundary side information in the regularized PET image reconstruction.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation under Grant No. SNSF 31003A-135576, Geneva Cancer League, and the Indo-Swiss Joint Research Programme ISJRP 138866.

APPENDIX A: DERIVATION OF WLS COST FUNCTION

Let us rewrite the negative Poisson log-likelihood in Eq. (1) as follows:

$$\begin{aligned}\Phi(x) &= \sum_{j=1}^M h_j([Gx]_j), \\ h_j(y_j) &= (n_j a_j y_j + \bar{r}_j) - y_j^p \log(n_j a_j y_j + \bar{r}_j).\end{aligned}\quad (\text{A1})$$

A quadratic approximation of the above objective function is obtained by applying a second-order Taylor's expansion to the marginal function $h(y_j)$ around an estimate of net-true coincidences at the j th line-of-response, \hat{y}_j , which yields

$$h_j(y_j) \approx h_j(\hat{y}_j) + h'_j(\hat{y}_j)(y_j - \hat{y}_j) + \frac{1}{2}h''_j(\hat{y}_j)(y_j - \hat{y}_j)^2, \quad (\text{A2})$$

where the first and second derivatives of h_j are given by

$$h'_j(y_j) = n_j a_j \left[1 - \frac{y_j^p}{n_j a_j y_j + \bar{r}_j + \bar{s}_j} \right],$$

$$h''_j(y_j) = y_j^p \left(\frac{n_j a_j}{n_j a_j y_j + \bar{r}_j + \bar{s}_j} \right)^2. \quad (\text{A3})$$

Using the estimate $\hat{y}_j = (y_j^p - \bar{r}_j - \bar{s}_j)/n_j a_j$ into Eq. (A2), we then arrive at

$$h_j(y_j) \approx y_j^p - y_j^p \log(y_j^p) + \frac{1}{2} \frac{n_j^2 a_j^2}{y_j^p} (y_j - \hat{y}_j)^2. \quad (\text{A4})$$

By dropping the terms independent of y_j in Eq. (A4), the log-likelihood reduces to the following weighted least squares cost function:

$$\phi(x) = \frac{1}{2} \sum_{j=1}^M w_j ([Gx]_j - \hat{y}_j)^2, \quad w_j = \frac{n_j^2 a_j^2}{\max(1, y_j^p)}. \quad (\text{A5})$$

In the above equation, we add one count to the empty sinogram bins to avoid division by zero. Note when the prompt emission sinogram is not available, the prompt coincidences can be estimated from an available precorrected sinogram \hat{y} , which yields $y_j^p \approx n_j a_j \hat{y}_j + \bar{r}_j + \bar{s}_j$.

APPENDIX B: SPS-OS ALGORITHM

The one-subset version of the SPS-OS algorithm applied for the PWLS cost function with a continuously differentiable penalty function has the following preconditioned gradient descent form:²⁹

$$x^{k+1} = x^k - P(\nabla\phi(x^k) + \beta\nabla R(x^k)),$$

$$P = \text{diag}\{G^T W G \mathbf{1} + \beta \nabla^2 R(x^k)\}^{-1}, \quad (\text{B1})$$

where $\nabla R(x^k) = \sum_{ij} d_{ji} \omega_i \phi'([Dx]_i)$, $\nabla^2 R(x^k) = \sum_{ij} |d_{ji}| \omega_i \sum_k |d_{ki}|$, where d_{ji} s are the elements of matrix D and ω_i are weighting factors to control the strength of regularization at voxel i .

^{a)}Author to whom correspondence should be addressed. Electronic mail: habib.zaidi@hcuge.ch; Telephone: +41 22 372 7258; Fax: +41 22 372 7169.

¹J. Czernin, M. Allen-Auerbach, and H. R. Schelbert, "Improvements in cancer staging with PET/CT: Literature-based evidence as of September 2006," *J. Nucl. Med.* **48**, 78S–88S (2007).

²M. F. Di Carli, S. Dorbala, J. Meserve, G. El Fakhri, A. Sitek, and S. C. Moore, "Clinical myocardial perfusion PET/CT," *J. Nucl. Med.* **48**, 783–793 (2007).

³J. C. Price, "Molecular brain imaging in the multimodality era," *J. Cereb. Blood Flow Metab.* **32**, 1377–1392 (2012).

⁴H. Zaidi and B. H. Hasegawa, "Determination of the attenuation map in emission tomography," *J. Nucl. Med.* **44**, 291–315 (2003).

⁵A. J. Reader and H. Zaidi, "Advances in PET image reconstruction," *PET Clinics* **2**, 173–190 (2007).

⁶R. M. Lewitt and S. Matej, "Overview of methods for image reconstruction from projections in emission computed tomography," *Proc. IEEE* **91**, 1588–1611 (2003).

⁷J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography," *Phys. Med. Biol.* **51**, R541–R578 (2006).

⁸D. S. Lalush and M. Wernick, "Iterative image reconstruction," in *Emission Tomography: The Fundamentals of PET and SPECT*, edited by M. Wernick and J. Aarsvold (Academic, San Diego, 2004), pp. 443–472.

⁹J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. Med. Imaging* **13**, 290–300 (1994).

¹⁰W. C. Karl, "Regularization in image restoration and reconstruction," in *Handbook of Image and Video Processing*, edited by A. Bovik (Academic, San Diego, 2000).

¹¹P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imaging* **9**, 84–93 (1990).

¹²V. E. Johnson, W. H. Wong, X. Hu, and C. T. Chen, "Image restoration using Gibbs priors: Boundary modeling, treatment of blurring, and selection of hyperparameter," *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 413–425 (1991).

¹³C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.* **2**, 296–310 (1993).

¹⁴S. J. Lee, A. Rangarajan, and G. Gindi, "Bayesian image reconstruction in SPECT using higher order mechanical models as priors," *IEEE Trans. Med. Imaging* **14**, 669–680 (1995).

¹⁵P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.* **12**, 629–639 (1990).

¹⁶F. Catte, P.-L. Lions, J.-M. Morel, and T. Coll, "Image selective smoothing and edge detection by nonlinear diffusion," *SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal.* **29**, 182–193 (1992).

¹⁷S. T. Acton, "Locally monotonic diffusion," *IEEE Trans. Signal Process.* **48**, 1379–1389 (2000).

¹⁸P. J. Huber, *Robust Statistics* (Wiley, New York, 1981).

¹⁹L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D* **60**, 259–268 (1992).

²⁰D. F. Yu and J. A. Fessler, "Edge-preserving tomographic reconstruction with nonlocal regularization," *IEEE Trans. Med. Imaging* **21**, 159–173 (2002).

²¹H. H. Feng, W. C. Karl, and D. A. Castanon, "A curve evolution approach to object-based tomographic reconstruction," *IEEE Trans. Image Process.* **12**, 44–57 (2003).

²²A. Sawatzky, "(Nonlocal) total variation in medical imaging," Ph.D. thesis, Westfälischen Wilhelms-Universität Münster, Germany, 2011, p. 215.

²³J. E. Bowsher, V. E. Johnson, T. G. Turkington, R. J. Jaszczak, C. E. Floyd, and R. E. Coleman, "Bayesian reconstruction and use of anatomical a priori information for emission tomography," *IEEE Trans. Med. Imaging* **15**, 673–686 (1996).

²⁴G. Gindi, M. Lee, A. Rangarajan, and I. G. Zubal, "Bayesian reconstruction of functional images using anatomical information as priors," *IEEE Trans. Med. Imaging* **12**, 670–680 (1993).

²⁵J. A. Fessler, N. H. Clinthorne, and W. L. Rogers, "Regularized emission image reconstruction using imperfect side information," in *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference Record, Santa Fe, NM, 1991* (IEEE, New York, 1991), Vol. 3, pp. 1991–1995.

²⁶K. Baete, J. Nuyts, W. Van Paesschen, P. Suetens, and P. Dupont, "Anatomical-based FDG-PET reconstruction for the detection of hypometabolic regions in epilepsy," *IEEE Trans. Med. Imaging* **23**, 510–519 (2004).

²⁷J. Tang and A. Rahmim, "Bayesian PET image reconstruction incorporating anato-functional joint entropy," *Phys. Med. Biol.* **54**, 7063–7075 (2009).

²⁸K. Vunckx, A. Atre, K. Baete, A. Reilhac, C. M. Deroose, K. Van Laere, and J. Nuyts, "Evaluation of three MRI-based anatomical priors for quantitative PET brain imaging," *IEEE Trans. Med. Imaging* **31**, 599–612 (2012).

²⁹Y. K. Dewaraja, K. F. Koral, and J. A. Fessler, "Regularized reconstruction in quantitative SPECT using CT side information from hybrid imaging," *Phys. Med. Biol.* **55**, 2523–2539 (2010).

- ³⁰F. Bataille, C. Comtat, S. Jan, F. C. Sureau, and R. Trebossen, "Brain PET partial-volume compensation using blurred anatomical labels," *IEEE Trans. Nucl. Sci.* **54**, 1606–1615 (2007).
- ³¹S. Somayajula, E. Asma, and R. M. Leahy, "PET image reconstruction using anatomical information through mutual information based priors," in *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference Record, Puerto Rico, 2005* (IEEE, New York, 2005), pp. 2722–2726.
- ³²C. Comtat, P. E. Kinahan, J. A. Fessler, T. Beyer, D. W. Townsend, M. De-frise, and C. Michel, "Clinically feasible reconstruction of 3D whole-body PET/CT data using blurred anatomical labels," *Phys. Med. Biol.* **47**, 1–20 (2002).
- ³³H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.* **44**, 2835–2851 (1999).
- ³⁴T. M. Benson, B. De Man, F. Lin, and J. B. Thibault, "Block-based iterative coordinate descent," in *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference Record, Knoxville, 2010* (IEEE, New York, 2010), pp. 2856–2859.
- ³⁵J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Image Process.* **8**, 688–699 (1999).
- ³⁶N. H. Clinthorne *et al.*, "Preconditioning methods for improved convergence rates in iterative reconstructions," *IEEE Trans. Med. Imaging* **12**, 78–83 (1993).
- ³⁷A. H. Delaney and Y. Bresler, "A fast and accurate Fourier algorithm for iterative parallel-beam tomography," *IEEE Trans. Image Process.* **5**, 740–753 (1996).
- ³⁸J. Savage and K. Chen, "On multigrids for solving a class of improved total variation based staircasing reduction models," in *Proceedings of the International Conference on PDE-based Image Processing and Related Inverse Problems, Norway, 2005* (Springer, New York, 2005), pp. 60–94.
- ³⁹A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imaging Vision* **20**, 89–97 (2004).
- ⁴⁰S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical x-ray CT reconstruction," *IEEE Trans. Med. Imaging* **31**, 677–688 (2012).
- ⁴¹J. Eckstein and D. P. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming* **55**, 293–318 (1992).
- ⁴²S. Setzer, "Operator splittings, Bregman methods and frame shrinkage in image processing," *Int. J. Comput. Vis.* **92**, 265–280 (2011).
- ⁴³K. Lange, D. R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," *J. Comput. Graph. Stat.* **9**, 1–20 (2000).
- ⁴⁴D. Hunter and K. Lange, "A tutorial on MM algorithms," *Am. Stat.* **58**, 30–37 (2004).
- ⁴⁵A. Beck and M. Teboulle, "Gradient-based algorithms with applications in signal recovery problems," in *Convex Optimization in Signal Processing and Communications*, edited by D. Palomar and Y. Eldar (Cambridge University Press, New York, 2010), pp. 33–88.
- ⁴⁶W. Zuo and Z. Lin, "A generalized accelerated proximal gradient approach for total-variation-based image restoration," *IEEE Trans. Image Process.* **20**, 2748–2759 (2011).
- ⁴⁷C. A. Johnson, J. Seidel, and A. Sofer, "Interior-point methodology for 3-D PET reconstruction," *IEEE Trans. Med. Imaging* **19**, 271–285 (2000).
- ⁴⁸L. Kaufman, "Maximum likelihood, least squares, and penalized least squares for PET," *IEEE Trans. Med. Imaging* **12**, 200–214 (1993).
- ⁴⁹D. Kim and J. A. Fessler, "Accelerated ordered-subsets algorithm based on separable quadratic surrogates for regularized image reconstruction in x-ray CT," in *Proceedings of the IEEE International Symposium on Biomedical Imaging, Valencia, Spain, 2011* (IEEE, New York, 2011), pp. 1134–1137.
- ⁵⁰E. P. Chong and S. H. Zak, *An Introduction to Optimization*, 2nd ed. (Wiley, New York, 2001).
- ⁵¹A. Chambolle and T. Pock, "A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging," *J. Math. Imaging Vis.* **40**(1), 120–145 (2011).
- ⁵²S. Becker, J. Bobin, and E. J. Candes, "NESTA: A fast and accurate first-order method for sparse recovery," *SIAM J. Imaging Sci.* **4**, 1–39 (2011).
- ⁵³J. Kiusalaas, *Numerical Methods in Engineering with MATLAB*, 2nd ed. (Cambridge University Press, New York, 2010).
- ⁵⁴J. Cheng-Liao and J. Qi, "PET image reconstruction with anatomical edge guided level set prior," *Phys. Med. Biol.* **56**, 6899–6918 (2011).
- ⁵⁵J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-8**(6), 679–698 (1986).
- ⁵⁶J. Fessler, "Image reconstruction toolbox," University of Michigan, 2013 (available URL <http://www.eecs.umich.edu/~fessler/code/index.html>).
- ⁵⁷F. Jacobs, E. Sundermann, B. D. Sutter, M. Christiaens, and I. Lemahieu, "A fast algorithm to calculate the exact radiological path through a pixel or voxel space," *J. Comput. Inf. Technol.* **6**, 89–94 (1998).
- ⁵⁸D. S. Lalush and B. M. W. Tsui, "A fast and stable maximum *a posteriori* conjugate gradient reconstruction algorithm," *Med. Phys.* **22**, 1273–1284 (1995).
- ⁵⁹T. Jensen, J. Jørgensen, P. Hansen, and S. Jensen, "Implementation of an optimal first-order method for strongly convex total variation regularization," *BIT Numer. Math.* **52**, 329–356 (2012).
- ⁶⁰W. Guo and F. Huang, "A local mutual information guided denoising technique and its application to self-calibrated partially parallel imaging," in *Proceedings of the Medical Image Computing Assisted Intervention, Part II, New York, 2008*, Lecture Notes in Computer Science Vol. 5242 (Springer-Verlag, Berlin Heidelberg, 2008), pp. 939–947.