

# Benefits of Using a Spatially-variant Penalty Strength with Anatomical Priors in PET Reconstruction

Yu-Jung Tsai\*, *Student Member, IEEE*, Georg Schramm, Sangtae Ahn, *Member, IEEE*, Alexandre Bousse, Simon Arridge, Johan Nuyts, *Senior Member, IEEE*, Brian F. Hutton, *Senior Member, IEEE*, Charles W. Stearns, *Fellow, IEEE* and Kris Thielemans, *Senior Member, IEEE*

**Abstract**—In this study, we explore the use of a spatially-variant penalty strength in penalized image reconstruction using anatomical priors to reduce the dependence of lesion contrast on surrounding activity and lesion location. This work builds on a previous method to make the local perturbation response (LPR) approximately spatially invariant. While the dependence of lesion contrast on the local properties introduced by the anatomical penalty is intentional, the method aims to reduce the influence from surroundings lying along the lines of response (LORs) but not in the penalty neighborhood structure. The method is evaluated using simulated data, assuming that the anatomical information is absent or well-aligned with the corresponding activity images. Since the Parallel Level Sets (PLS) penalty is convex and has shown promising results in the literature, it is chosen as the representative anatomical penalty and incorporated into the previously proposed preconditioned algorithm (L-BFGS-B-PC) for achieving both good image quality and fast convergence rate. A 2D disc phantom with a feature at the center and a 3D XCAT thorax phantom with lesions inserted in different slices are used to study how surrounding activity and lesion location affect the visual appearance and quantitative consistency. A bias and noise analysis is also performed with the 2D disc phantom. The consistency of the algorithm convergence rate with respect to different data noise and background levels is also investigated using the XCAT phantom. Finally, an example reconstruction for a patient dataset with inserted pseudo lesions is used as a demonstration in a clinical context. We show that applying the spatially-variant penalization with PLS can reduce the dependence of the lesion contrast on the surrounding activity and lesion location. It does not affect the bias and noise trade-

off curves for matched local resolution. Moreover, when using the proposed penalization, significant improvement in algorithm convergence rate and convergence consistency is observed.

## I. INTRODUCTION

IN emission tomography (ET), the use of an image reconstruction algorithm based on penalized maximum-likelihood (PML) is one of the strategies to control noise amplification as iterations increase [1]–[3]. Desired properties such as sharp edges and smoothness in uniform regions can also be encouraged by appropriate choice of the penalty function [4], [5]. However, the trade-off between noise and resolution in reconstructed images can vary with the individual dataset [6], making the tuning of the weight between the likelihood term and the penalty term of the objective function difficult. In addition, the effective penalty strength at each voxel is determined by its activity level, and that of the surrounding voxels [7]. This makes visual comparison and quantitative analysis for regions with different activities or at different locations difficult even for lesions in the same image.

These issues have been studied extensively for quadratic priors, and analytical predictions of image resolution and variance are available [7]–[11]. The dependence of lesion interpretation on location, surrounding object and therefore the individual can be largely eliminated by modifying the weights of the penalty function using analytical models [7], [9], [11]–[15]. Moreover, since the reconstruction properties have become almost object independent, one can tabulate the relationship between the overall penalty weight and the corresponding resolution by performing reconstructions using a set of Monte Carlo (MC) simulated or experimental data [7].

In contrast to the large literature on quadratic penalization, limited studies exist for edge preserving priors. These priors aim to encourage high resolution at edges and smoothness in uniform regions. To be able to predict local resolution (and variance), the analytical models proposed initially for quadratic penalties were further generalized for non-quadratic ones in [16]. The local resolution was characterized using the local perturbation response (LPR). Instead of trying to achieve uniform resolution, the authors of [16] proposed a similar modification of the penalty weights to obtain uniform LPR across the field-of-view (FOV) for lesions having similar local features [16]. Since the local contrast of the “reconstructed”

This work was supported in part by GE Healthcare, in part by CCP PET-MR (EPSRC grant EP/M022587/1) and in part by the NIHR-funded UCLH Biomedical Research Centre. The work of G. Schramm was supported by the NIH project 1P41EB017183-01A1. *Asterisk indicates corresponding author.*

Y.-J. Tsai\* was with the Institute of Nuclear Medicine, University College London, London NW1 2BU, UK and is now with the Yale PET Center, Yale University, New Haven, CT 06520, USA (e-mail: yu-jung.tsai@yale.edu).

G. Schramm and J. Nuyts are with the Department of Imaging and Pathology, KU Leuven, Leuven 3000, Belgium.

S. Ahn is with GE Global Research, Niskayuna, NY 12309 USA.

A. Bousse was with the Institute of Nuclear Medicine, University College London, London NW1 2BU, UK and is now with the LaTIM, INSERM, UMR 1101, Université de Bretagne Occidentale, Brest, France.

S. Arridge is with the Department of Computer Science, University College London, London WC1E 6BT, UK.

B. F. Hutton and K. Thielemans are with the Institute of Nuclear Medicine, University College London, London NW1 2BU, UK

B. F. Hutton is also with the Centre for Medical Radiation Physics, University of Wollongong, Wollongong NSW 2522, Australia.

C. W. Stearns is with MICT Engineering, GE Healthcare, Waukesha, WI 53188 USA.

Copyright (c) 2019 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

lesion, as determined by LPR, is one of the factors that affects visual comparison and quantitative accuracy of the lesion [17], [18], reducing the dependence of LPR on location or surroundings could help to improve consistency and reliability in many applications. For example, it could benefit positron emission tomography (PET)  $^{18}\text{F}$ -FDG scans for treatment response evaluation [19]–[21] or gross tumor volume delineation [22], [23]. Since the modification is essentially a spatial normalization method for the penalty function, it is referred to as a spatially-variant penalty strength in this study.

In [16], the use of the modified penalty function was validated with an edge preserving Huber prior using 2D MC simulations. However, there is increased interest in priors using anatomical information to tune local properties of the penalty.

This paper, an expansion of initial results presented in [24], extends the theory developed in [16] to a wider class of penalty functions such that it can be applied to any anatomical priors defined in a given neighborhood structure. For computational convenience, a different type of spatially-variant penalization scheme is proposed. A further modification for reducing the sensitivity of the scheme to data noise level is made as well. We introduce two metrics of reconstruction algorithm performance: one to measure the improvement of consistency in lesion quantification and one to assess algorithm convergence rate. The former is mainly evaluated with respect to different surroundings and locations, while the latter is studied using simulations representing different noise and background levels. A bias and noise analysis is also performed with multiple realizations for a simple disc phantom. A set of patient data is also used as a demonstration of using an anatomical penalty function with the proposed modification to achieve consistent quantification and faster convergence rate in a clinical context.

## II. METHODS

### A. PML Image Reconstruction

Given the tracer distribution  $\mathbf{x} = [x_1, \dots, x_J]^\top \in \mathbb{R}^J$  and the measured data  $\mathbf{y} = [y_1, \dots, y_I]^\top \in \mathbb{R}^I$ , PML image reconstruction optimizes an objective function  $\Phi$  consisting of the log-likelihood  $L$  and the penalty function  $R$  with a parameter  $\beta$  controlling its strength:

$$\Phi(\mathbf{x}, \mathbf{y}) = -L(\mathbf{x}, \mathbf{y}) + \beta R(\mathbf{x}). \quad (1)$$

We define the PML estimator as:

$$\hat{\mathbf{x}}(\mathbf{y}) = \arg \min_{\mathbf{x} \geq 0} \Phi(\mathbf{x}, \mathbf{y}). \quad (2)$$

In ET, the statistical nature of the measured data  $\mathbf{y}$  can be described using the Poisson distribution. Therefore, the log-likelihood function  $L$ , omitting terms independent of  $\mathbf{x}$ , can be expressed as:

$$L(\mathbf{x}, \mathbf{y}) = \sum_i y_i \log \bar{y}_i(\mathbf{x}) - \bar{y}_i(\mathbf{x}), \quad \bar{\mathbf{y}}(\mathbf{x}) \triangleq \mathbf{A}\mathbf{x} + \mathbf{n} \quad (3)$$

where  $\bar{y}_i$  is the mean measurement in bin  $i$ ,  $\mathbf{A} \in \mathbb{R}^{I \times J}$  is the system matrix and  $\mathbf{n}$  is the expected background, due to events such as scatter and random coincidences. Each element  $[\mathbf{A}]_{i,j} = A_{i,j}$  denotes the probability that an emission from voxel  $j$  is detected by bin  $i$ .

In this work, we focus on penalty functions  $R$  of the form:

$$R(\mathbf{x}) = \sum_j \phi_j(\mathbf{x}), \quad (4)$$

where each  $\phi_j(\mathbf{x})$  only depends on  $x_j$  and voxels in a given neighborhood  $\mathcal{N}_j$  of voxel  $j$ . Such penalty functions include standard priors of the form  $\phi_j(\mathbf{x}) = \sum_{k \in \mathcal{N}_j} \omega_{jk} \psi(x_j - x_k)$  with  $\omega_{jk}$  indicating the weight between voxel  $j$  and its neighbouring voxel  $k$  as well as local anatomical priors such as the Bowsher prior [25] or the Parallel Level Sets (PLS) prior [26]. One requirement on  $R$  is that  $\Phi$  is strictly convex and differentiable so that (2) is uniquely defined.

Since PLS has shown promising results and its convexity is well-established [26], it is chosen to be the representative anatomical prior in this study. The PLS penalty is defined as:

$$\phi_j(\mathbf{x}|\mathbf{z}) = \sqrt{\alpha^2 + \|\nabla \mathbf{x}\|_j^2 - \langle \nabla \mathbf{x}\|_j, [\xi]_j \rangle^2}, \quad (5)$$

$$[\xi]_j := \frac{[\nabla \mathbf{z}]_j}{\sqrt{\|\nabla \mathbf{z}\|_j^2 + \eta^2}}, \quad \alpha \text{ and } \eta > 0$$

where  $\nabla$  is the gradient operator,  $\langle \cdot, \cdot \rangle$  is the Euclidean inner product,  $\mathbf{z} = [z_1, \dots, z_J]^\top \in \mathbb{R}^J$  is the anatomical image and  $\|\cdot\|_2$  denotes the  $\ell^2$ -norm. The edge preserving property of the function is modulated by the pair of parameters  $(\alpha, \eta)$  [26]. Note that PLS does not require a segmentation of the anatomical image and reduces to the (smooth) Total Variation (TV) prior in the absence of the anatomical information.

### B. Linearized Local Perturbation Response

We summarize the theoretical description of the image resolution proposed by Ahn & Leahy [16] for non-quadratically penalized image reconstruction in this section. By further exploring the property of the Fisher information matrix, a different type of the spatially-variant penalty strength suitable for 3D PET reconstruction is derived. To reduce its sensitivity to data noise level, an approximation that avoids calculating the inverse of the measured data is also proposed.

Let  $\mathbf{x}_t$  be a “true” activity image consisting of a background  $\mathbf{x}_b$  and a signal of interest  $\mathbf{x}_s$ :

$$\mathbf{x}_t = \mathbf{x}_b + \mathbf{x}_s. \quad (6)$$

One way to quantify resolution properties around the signal  $\mathbf{x}_s$  is the *local perturbation response* (LPR) [16]:

$$\bar{\tilde{\mathbf{x}}}_s \triangleq E[\hat{\mathbf{x}}_t] - E[\hat{\mathbf{x}}_b], \quad (7)$$

where  $E[\hat{\mathbf{x}}_t]$  and  $E[\hat{\mathbf{x}}_b]$  are the mean reconstructed images for the noisy measurements with and without the presence of the signal. For reasonably high signal-to-noise ratio (SNR), the mean perturbation  $\bar{\tilde{\mathbf{x}}}_s$  defined in (7) can be approximated by the *linearized local perturbation response* (LLPR) [16]:

$$\check{\tilde{\mathbf{x}}}_s \triangleq \check{\tilde{\mathbf{x}}}_t - \check{\tilde{\mathbf{x}}}_b \quad (8)$$

where

$$\check{\tilde{\mathbf{x}}}_t \triangleq \hat{\mathbf{x}}(\bar{\mathbf{y}}(\mathbf{x}_t)), \quad \check{\tilde{\mathbf{x}}}_b \triangleq \hat{\mathbf{x}}(\bar{\mathbf{y}}(\mathbf{x}_b)). \quad (9)$$

With suitable Taylor expansions, the relationship between the LLPR  $\check{\mathbf{x}}_s$  and the true perturbation  $\mathbf{x}_s$  can be described as:

$$[\mathbf{F}(\check{\mathbf{x}}_b) + \beta\mathbf{\Omega}(\check{\mathbf{x}}_b; \check{\mathbf{x}}_s)] \check{\mathbf{x}}_s \approx \mathbf{F}(\check{\mathbf{x}}_b)\mathbf{x}_s, \quad (10)$$

where  $\mathbf{F}(\mathbf{x}) \in \mathbb{R}^{J \times J}$  is the Fisher information matrix for estimating  $\mathbf{x}$ :

$$\mathbf{F}(\mathbf{x}) \triangleq \mathbf{A}^\top D \left[ \frac{1}{\bar{\mathbf{y}}(\mathbf{x})} \right] \mathbf{A} \quad (11)$$

with  $D[\cdot]$  an operator that constructs a diagonal matrix from a vector. In [16],  $\mathbf{\Omega}(\check{\mathbf{x}}_b; \check{\mathbf{x}}_s) \in \mathbb{R}^{J \times J}$  is defined as:

$$\mathbf{\Omega}(\check{\mathbf{x}}_b; \check{\mathbf{x}}_s) \triangleq \int_0^1 \nabla^2 R(\check{\mathbf{x}}_b + \tau\check{\mathbf{x}}_s) d\tau \quad (12)$$

such that

$$\mathbf{\Omega}(\check{\mathbf{x}}_b; \check{\mathbf{x}}_s)\check{\mathbf{x}}_s = \nabla R(\check{\mathbf{x}}_b + \check{\mathbf{x}}_s) - \nabla R(\check{\mathbf{x}}_b). \quad (13)$$

The notation  $\nabla^2$  represents the Hessian operator.

Although  $\mathbf{F}(\mathbf{x})$  is a non-diagonal matrix, it is concentrated about its diagonal [7]. The image of each row (or column) vector of  $\mathbf{F}(\mathbf{x})$  is assumed to have a similar shape (*i.e.*  $1/r$  and  $1/r^2$  blurring kernel for 2D and 3D PET acquisition, respectively) and its peak center value is approximately proportional to the sum of the image. The value depends on the activity distribution and the spatial variations in sensitivity for a shift-variant system. To investigate the spatial variance, it is useful to define an alternative matrix that separates  $\mathbf{F}(\mathbf{x})$  into a data dependent  $\kappa(\mathbf{x})$  and an (approximately) data independent  $\mathbf{F}_0(\mathbf{x})$ :

$$\mathbf{F}_0(\mathbf{x}) \triangleq D[\kappa(\mathbf{x})]^{-1} \mathbf{F}(\mathbf{x}) D[\kappa(\mathbf{x})]^{-1}, \quad (14)$$

where  $\kappa(\mathbf{x})$  is chosen such that  $\mathbf{F}_0(\mathbf{x})$  is an approximately shift-invariant matrix, desirable for 3D PET reconstruction. For example, the following  $\kappa(\mathbf{x})$  was used in [11] in the case where detector blurring is not modeled:

$$\kappa_j(\mathbf{x}) \triangleq \sqrt{\sum_i A_{i,j}^2 / \bar{y}_i(\mathbf{x})}, \quad \forall j = 1, \dots, J. \quad (15)$$

This choice<sup>1</sup> leads to an approximately shift-invariant  $\mathbf{F}_0(\mathbf{x})$  with diagonal elements being 1. Another choice is discussed in Section II-D.

As  $\kappa_j(\mathbf{x})$  varies slowly with  $j$ , the following simplification of (14) is valid around voxel  $j$ :

$$\mathbf{F}_0(\mathbf{x})\mathbf{e}_j \approx \kappa_j^{-2}(\mathbf{x})\mathbf{F}(\mathbf{x})\mathbf{e}_j \quad (16)$$

where  $\mathbf{e}_j$  is the  $j$ -th unit vector of length  $J$ . With (16) and assuming that  $\check{\mathbf{x}}_s$  is concentrated on voxel  $l$ , (10) can be further approximated by:

$$\left[ \mathbf{F}_0(\check{\mathbf{x}}_b) + \frac{\beta}{\kappa_l^2(\check{\mathbf{x}}_b)} \mathbf{\Omega}(\check{\mathbf{x}}_b; \check{\mathbf{x}}_s) \right] \check{\mathbf{x}}_s \approx \mathbf{F}_0(\check{\mathbf{x}}_b)\mathbf{x}_s. \quad (17)$$

<sup>1</sup>In the notation of Qi & Leahy [11], this is  $\kappa/\nu$ .

### C. Modified Prior for Data Independent LLPR

The dependence of LLPR on data in PML reconstructions is due to the presence of  $\kappa_l^2(\check{\mathbf{x}}_b)$  in (17). Several approaches have been proposed to modify  $R$  to eliminate the data-dependency [7], [16]. With the more general form defined in (4), the modified priors can be described as:

$$\tilde{R}(\mathbf{x}) = \sum_j \kappa_j^2(\mathbf{x})\phi_j(\mathbf{x}). \quad (18)$$

Given  $\mathcal{O}_l$  a neighborhood of voxel  $l$  containing the non-zero indices of the corresponding LLPR and  $\check{\mathbf{x}}_s$  the LLPR using  $\tilde{R}$ , (13) can be rewritten as:

$$\tilde{\mathbf{\Omega}}(\check{\mathbf{x}}_b; \check{\mathbf{x}}_s)\check{\mathbf{x}}_s = \sum_{j \in \mathcal{O}_l} \kappa_j^2(\check{\mathbf{x}}_b)\mathbf{c}_j(\check{\mathbf{x}}_b; \check{\mathbf{x}}_s) \quad (19)$$

where  $\tilde{\mathbf{\Omega}}$  is defined as in (12) by replacing  $R$  with  $\tilde{R}$  and

$$\mathbf{c}_j(\check{\mathbf{x}}_b; \check{\mathbf{x}}_s) = \nabla\phi_j(\check{\mathbf{x}}_b + \check{\mathbf{x}}_s) - \nabla\phi_j(\check{\mathbf{x}}_b).$$

Since  $\mathbf{c}_j(\check{\mathbf{x}}_b; \check{\mathbf{x}}_s)$  has non-zero entries  $[\mathbf{c}_j(\check{\mathbf{x}}_b; \check{\mathbf{x}}_s)]_m$  only for  $m$  in the neighborhood of  $l$  and  $\kappa_l$  varies slowly with location, (19) is further approximated by:

$$\begin{aligned} \tilde{\mathbf{\Omega}}(\check{\mathbf{x}}_b; \check{\mathbf{x}}_s)\check{\mathbf{x}}_s &\approx \kappa_l^2(\check{\mathbf{x}}_b) \sum_{j \in \mathcal{O}_l} \mathbf{c}_j(\check{\mathbf{x}}_b; \check{\mathbf{x}}_s) \\ &\approx \kappa_l^2(\check{\mathbf{x}}_b)\mathbf{\Omega}(\check{\mathbf{x}}_b; \check{\mathbf{x}}_s)\check{\mathbf{x}}_s. \end{aligned} \quad (20)$$

Substituting (20) into (17) gives:

$$[\mathbf{F}_0(\check{\mathbf{x}}_b) + \beta\mathbf{\Omega}(\check{\mathbf{x}}_b; \check{\mathbf{x}}_s)] \check{\mathbf{x}}_s \approx \mathbf{F}_0(\check{\mathbf{x}}_b)\mathbf{x}_s, \quad (21)$$

This result shows that the modified penalty  $\tilde{R}$  is able to eliminate the LLPR dependence on data. Note that it still depends on the local properties introduced by the penalty.

### D. Alternative choice for $\kappa(\mathbf{x})$

The above derivation holds for any ‘‘smooth’’  $\kappa(\mathbf{x})$  that makes  $\mathbf{F}_0(\mathbf{x})$  approximately shift-invariant and data-independent. Existing formulations such as (15) are inconvenient as they need access to the square of the elements of the matrix  $\mathbf{A}$ . Inspired by the ‘‘precomputed denominator’’ in [27], we choose a different type of  $\kappa(\mathbf{x})$  using the square root of the row-sums of  $\mathbf{F}(\mathbf{x})$ . In matrix notation:

$$\kappa(\mathbf{x}) \triangleq \sqrt{\mathbf{F}(\mathbf{x})\mathbf{1}} = \sqrt{\mathbf{A}^\top D \left[ \frac{1}{\bar{\mathbf{y}}(\mathbf{x})} \right] \mathbf{A} \mathbf{1}}, \quad (22)$$

where  $\mathbf{1}$  is a vector of ones. Note that this type of  $\kappa(\mathbf{x})$  requires only forward and back-projection operations. Substituting (22) into (16), we find that the row-sums of  $\mathbf{F}_0(\mathbf{x})$  will be approximately equal to 1:

$$\mathbf{F}_0(\mathbf{x})\mathbf{1} \approx D[\kappa(\mathbf{x})]^{-2} \mathbf{F}(\mathbf{x})\mathbf{1} \approx \mathbf{1}. \quad (23)$$

In other words, this type of  $\kappa(\mathbf{x})$  also leads to a shift-invariant  $\mathbf{F}_0(\mathbf{x})$  for 3D PET reconstruction as the image of each row (or column) vector of  $\mathbf{F}_0(\mathbf{x})$  would have a similar shape and peak center value.

Since  $\kappa(\mathbf{x})$  in (22) is calculated with the mean or noiseless measurement  $\bar{\mathbf{y}}(\mathbf{x})$  which is unavailable in practice, a plug-in

method that substitutes  $\bar{\mathbf{y}}(\mathbf{x})$  by the noisy measurement  $\mathbf{y}$  to compute  $\kappa(\mathbf{x})$  was introduced in [16]. This leads to:

$$\tilde{\kappa} \triangleq \sqrt{\mathbf{A}^\top D \left\{ \frac{1}{\mathbf{y}} \right\} \mathbf{A} \mathbf{1}}, \quad (24)$$

However, as it requires a division by the measured data, its performance is sensitive to the noise level. The following approximation is therefore proposed in this study:

$$\hat{\kappa} \triangleq \sqrt{\mathbf{A}^\top D \left[ \frac{\mathbf{y}}{\bar{\mathbf{y}}^2(\mathbf{x}^{\text{init}})} \right] \mathbf{A} \mathbf{1}}, \quad (25)$$

where  $\mathbf{x}^{\text{init}}$  is the initial image. The matrix  $\hat{\kappa}$  is used to construct the transformation matrix in Section II-E and in [28]. As  $\mathbf{A}$  and  $\mathbf{A}^\top$  are smoothing operators that reduce the noise in  $\mathbf{x}^{\text{init}}$  and  $\mathbf{y}$ , both  $\tilde{\kappa}$  and  $\hat{\kappa}$  is a reasonable approximation of  $\kappa(\mathbf{x})$  defined in (22) when the initial image is close to the solution and the noise level of data  $\mathbf{y}$  is reasonably low. To demonstrate that the proposed spatially-variant penalty strength (25) is a better approximation of (22) compared to the plug-in method (24) in the presence of noise, we have investigated the difference between them using a simulated noisy dataset in Appendix A.

#### E. Optimization Method: L-BFGS-B-PC

The modified penalty function (18) with the spatially-variant penalty strength (25) was incorporated into a fast convergent reconstruction algorithm, L-BFGS-B-PC, which was previously proposed by our group [28]–[30]. The algorithm performed L-BFGS-B [31], [32] in a transformed coordinate system to circumvent its potential slow convergence and sensitivity to global scale factors. As suggested in [28], the transformation matrix  $\mathbf{P}$  can be chosen to be the square root of a diagonal approximation of the Hessian of  $\Phi$ , which consists of the Hessian of the likelihood function (3) and the Hessian of the penalty function (4). Since (25) is also related to a diagonal approximation of the Hessian of (3), we have used it to construct  $\mathbf{P}$  to reduce the total computation. As it is not necessary that  $\mathbf{P}$  is a precise approximation of the Hessian of  $\Phi$  [28], we further replaced the Hessian of the penalty function by a small constant  $\epsilon$  for simplicity and to avoid division by zero. The transformation can therefore be described as:

$$\mathbf{x}' = \mathbf{P} \mathbf{x}, \quad \mathbf{P} = D \left[ \sqrt{\hat{\kappa}^2 + \epsilon} \right], \quad (26)$$

where  $\epsilon = 10^{-4}$  in this study.

An update at each iteration is then found along a search direction  $\mathbf{d}$  with the transformed image  $\mathbf{x}'$ :

$$\mathbf{x}'_{t+1} = \mathbf{x}'_t + \delta^* \mathbf{d}_t, \quad (27)$$

where  $\delta^*$  is the step length and  $\mathbf{d}_t = -\mathbf{B}_t^{-1} \nabla \Phi'(\mathbf{x}'_t)$  with  $\mathbf{B}_t$ , an approximation of the inverse of the Hessian of  $\Phi$  at  $\mathbf{x}'_t$ , constructed by L-BFGS-B. Here  $\nabla \Phi'(\mathbf{x}'_t)$  is the transformed gradient for  $\Phi$  at  $\mathbf{x}'_t$ . In this section, the subscript  $t$  indicates the iteration number instead of the voxel location for compactness. To ensure convergence and sufficient progress,  $\delta^*$  is generally obtained by performing a backtracking algorithm in

which a series of gradually decreasing  $\delta$  from an initial value  $\delta^{\text{init}} \leq 1$  are tested until the Wolfe conditions are met:

$$\Phi'(\mathbf{x}'_t + \delta \mathbf{d}_t) \leq \Phi'(\mathbf{x}'_t) + \lambda_1 \delta \nabla \Phi'(\mathbf{x}'_t)^\top \mathbf{d}_t \quad (28)$$

$$\|\nabla \Phi'(\mathbf{x}'_t + \delta \mathbf{d}_t)^\top \mathbf{d}_t\|_2 \leq \lambda_2 \|\nabla \Phi'(\mathbf{x}'_t)^\top \mathbf{d}_t\|_2, \quad (29)$$

where  $\Phi'(\mathbf{x}'_t) = \Phi(\mathbf{x}_t)$ ,  $0 < \lambda_1 < \lambda_2 < 1$ . In this study,  $\lambda_1$  and  $\lambda_2$  were set to  $10^{-4}$  and 0.9, respectively. We stopped the algorithm if no step length  $\delta^* > 0$  that satisfies the Wolfe conditions can be found in 20 trials of backtracking. This stopping criterion was checked using the results from [28]. More information about the algorithm can be found in [28].

Since both the objective function and its gradient have to be recomputed for each candidate  $\delta$ , extra forward and back-projection operations are needed for the backtracking.

### III. EVALUATION

In this study, we investigate the performance of the anatomical prior, PLS, with and without the new approximation of  $\kappa(\mathbf{x})$  in (25). The potential of applying the modified PLS to improve the quantitative consistency and algorithm convergence rate are investigated with two digital phantoms, a 2D disc phantom and a more realistic 3D XCAT phantom. As changes in the local resolution will influence the image noise properties, a contrast and variance analysis with respect to various penalty strengths using several noise realizations is included for the disc phantom. The multiple realizations are reconstructed with and without anatomical information to simulate the mismatch between emission and anatomical images as well. Finally, a patient dataset with inserted pseudo lesions is used as an example reconstruction closer to the clinical context.

#### A. Data

1) *2D disc phantom*: To demonstrate that the uniformity of the local perturbation for lesions in different surroundings can be improved by applying the modified PLS, a 2D disc phantom with a sphere (value = 3) inserted right at the center of a large hot (value = 5) or cold (value = 1) uniform region was used (see Fig. 3). The phantom was a  $111 \times 111$  matrix with voxel size of  $2.397 \times 2.397$  mm<sup>2</sup> and the diameter of the sphere was 21.573 mm. An attenuation map ( $\mu$ ) was also used to provide anatomical information, consisting of 0.096 and 0.172 cm<sup>-1</sup> for the feature and the surrounding matter, respectively. Note that, relative to the surrounding region, the sphere had the same absolute activity difference and, hence, the effect of PLS would be the same for the spheres in both cold and hot surroundings.

The projection data were generated by using STIR [33] projectors to simulate data from a GE Discovery STE in 2D acquisition mode [34]. The phantom with either high or low surrounding activity was used to generate 100 noise realizations using the Poisson noise model. The total counts were around 302 K and 424 K for the phantom with cold and hot surroundings, respectively. Physical effects, such as attenuation and system blurring (modeled as smoothing in image space with full width at half maximum (FWHM) = 5.2 mm in tangential and radial directions) were simulated.

A uniform projection with a constant intensity was added to the generated data to simulate the background events. It is equivalent to 90% and 64% of the total prompts for the phantom with cold and hot surroundings. For analysis purposes (see Section III-C for more information), we also generated data using the same phantom but with no sphere at the center of both activity and anatomical images. The total counts were 301 K and 422 K for the cold and hot sphere-free phantom.

2) *3D XCAT thorax phantom*: An XCAT phantom [35] was used to generate data representing typical  $^{18}\text{F}$ -FDG scans in the thorax. The resulting volume was of size  $152 \times 152 \times 47$  with voxel size 3.125 mm in all directions. We rescaled the phantom to give voxel intensities between 0 to 2.

To evaluate the dependence of local perturbation on location, 6 hot lesions with size of  $3 \times 3 \times 3$  voxels were inserted in different slices (see Fig. 1). None of the lesions were in the central slice and 2 of them were in the liver. The uptake of the liver was either high (value = 1.6) or low (value = 0.4) to simulate change of surroundings for lesions in the liver. Each lesion had the same absolute difference to its surrounding in both activity (difference = 0.8) and anatomical images (difference = 20 Hounsfield units (HU)) hence the same influence of the anatomical prior. As lesions in liver would have a similar linear attenuation coefficient to liver, our simulations roughly correspond to using CT with injected contrast to provide anatomical information. Note that the distribution of the contrast agent in the enhanced CT images was assumed to be locally uniform and all lesions in the activity image were exactly aligned with those in the anatomical image. Fig. 1 shows the XCAT phantom with high liver uptake and its corresponding contrast enhanced CT image as an example. The phantom was forward projected using vendor supplied software into sinograms simulating data from the GE Discovery STE in 3D acquisition. The background events were simulated by adding a constant value to the generated sinograms. The amount of background events was equivalent to 56% and 55% of total prompts for phantom with low and high liver uptake, respectively. Here we performed the evaluation of local perturbation on noiseless datasets. Similarly, the phantom with and without lesions was used to generate data for analysis purposes.

The phantom with high liver uptake was also used for investigating the influence of the modified PLS on the performance of L-BFGS-B-PC with respect to different noise and background levels. These factors had been proven to influence the convergence rate of L-BFGS-B-PC when using non-anatomical priors with spatially invariant weights in our previous study [28]. For assessing effects of the noise level, we generated three datasets (G0) with total counts  $S_{\text{tot}}$  of 50 M, 250 M and 1252 M. The Poisson noise model was used. The true to background event ratio (TBR) for each of them was fixed at 0.6. The possible influence of background events was studied using four more datasets, which were divided into two groups. For the first group (G1), each dataset had the same total counts as the data representing medium noise level, *i.e.*,  $S_{\text{tot}} = 250$  M, but with 5 times lower or higher TBR. For data in the second group (G2), we kept  $S_{\text{true}}$  the same as that in the data with  $S_{\text{tot}} = 250$  M counts, but changed  $S_{\text{bg}}$

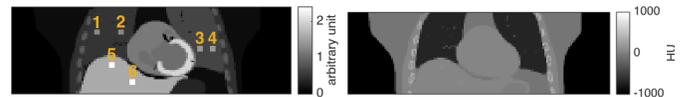


Fig. 1. The central coronal view of the XCAT phantom with high liver uptake (left) and the corresponding contrast enhanced CT image used to provide anatomical information (right). Note that given the small difference in HU value between the lesions and their surroundings compared to the image scale, the enhanced lesions are not apparent without using a specific display window for lesions in different surroundings.

TABLE I  
A SUMMARY OF THE SIMULATED DATA FOR EVALUATING THE INFLUENCE OF THE BACKGROUND.

		$S_{\text{true}}$	$S_{\text{bg}}$	$S_{\text{tot}}$
G0	TBR = 0.6	19 M	31 M	50 M
	TBR = 0.6	94 M	156 M	250 M
	TBR = 0.6	469 M	783 M	1252 M
G1	TBR = 3	187 M	63 M	250 M
	TBR = 0.12	27 M	223 M	250 M
G2	TBR = 3	94 M	31 M	125 M
	TBR = 0.12	94 M	783 M	877 M

by 5 times lower or higher. Note that these two groups had identical TBR for the same background level: TBR = 0.12 for the high background data and TBR = 3 for the low background data. A summary of the simulated data in these two groups can be found in Table I. Note that all simulations took into account the attenuation effect and system blurring modeled in image space using FWHM = 5.2 mm in tangential and radial directions and 5.7 mm in the axial direction.

3) *Patient data*: A patient dataset of the thorax region acquired on the GE Discovery STE PET/CT scanner was used for this retrospective study. The patient was injected with 315 MBq of  $^{18}\text{F}$ -FDG approximately 1 hour before the scan started. The acquisition included a cine-CT scan (140 kVp, 60 mA, 4 s duration, 0.5 s rotation period, 0.45 s time between reconstructed images, 9 bed positions, 8 axial slices per bed position), followed by a PET scan in fully 3D mode. The total number of counts in the PET data was  $S_{\text{tot}} = 219\text{M}$ . Since the true lesion value and location are unknown, three pseudo lesions with size of 9.375 mm in all directions were forward projected and attenuated using vendor-provided software (the GE PET Toolbox, not commercial). A Poisson noise realization was then generated based on the projection data using a built-in function in MATLAB. The final dataset was a sum of the noisy data for the pseudo lesions and the measured data. The locations of the pseudo lesions (two in the liver and one in the right lung of the patient, none of these in the central slice) were determined from a preliminary reconstruction. The difference between each pseudo lesion and its surroundings was 0.8 MBq/cc. The corresponding contrast enhanced CT with lesions inserted at the same locations was simulated from the average cine-CT to provide anatomical information. Based on the results in [36]–[38], an absolute difference of 150 HU to the surroundings was assigned to all pseudo lesions in the simulated enhanced CT image. We ignored the possible non-uniform distribution of the contrast agent in clinical practice and the attenuation change induced by the inserted lesions. The vendor-provided software was also

used for binning the PET data into sinograms and modeling the detection efficiency, attenuation, scatter and randoms.

### B. Reconstruction

Based on our previous study [28], L-BFGS-B-PC initialized by 1 full iteration of ordered-subsets expectation maximization (OS-EM) with 35 subsets was adopted for image reconstruction. The initial image was also used to calculate the spatially-variant penalty strength  $\hat{\kappa}$  and the transformation matrix  $\mathbf{P}$ . We reconstructed each dataset with and without using the proposed  $\hat{\kappa}$  to modify PLS. For simulated data, the reconstructed image had the same matrix size as the corresponding phantom. The reconstructed voxel size was  $3.125 \times 3.125 \text{ mm}^2$  for the disc phantom and  $3.947 \times 3.947 \times 3.27 \text{ mm}^3$  for the XCAT phantom. We chose different voxel sizes for phantom definitions and image reconstructions to avoid artifacts induced by discretization. The reconstructed image for the patient datasets had  $192 \times 192 \times 47$  voxels with voxel size of  $3.125 \times 3.125 \times 3.27 \text{ mm}^3$ . The system blurring was also modeled during reconstruction. To exclude the dependence of the penalty function on the selection of parameters, the parameter set  $(\alpha, \eta)$  in PLS was chosen according to the scale of the anatomical and activity images, respectively. We kept the strength of  $\alpha$  and  $\eta$  to 12.5% and 25% of the intensity difference between the lesion and its surroundings in anatomical and activity images for all reconstructions. Therefore, the set of parameters was  $(0.25, 0.019 \text{ cm}^{-1})$  for the disc phantom,  $(0.1, 5 \text{ HU})$  for the XCAT phantom and  $(0.1, 37.5 \text{ HU})$  for the real patient data. We defined a different global penalty strength  $\hat{\beta} = \beta \hat{\kappa}_0^2$  for reconstructions without applying  $\hat{\kappa}$ , where  $\hat{\kappa}_0$  represents the value at the center of  $\hat{\kappa}$  calculated with a reference dataset. Therefore, the effective penalty strength at the image center is identical in both reconstructions with and without using  $\hat{\kappa}$  when the reference dataset is considered. The influence of the global penalty strength on quantification is discussed further in Section V. We selected a different reference dataset for different evaluations (see section IV for more information). A series of global penalty strengths  $\beta$  were used for the disc phantom to study the resolution and noise trade-off at different penalty strengths. The evaluated set of strengths started from 0.05 and increased by 0.05 to 0.3. The strength was fixed at  $10^{-3}$  for the noiseless XCAT datasets. A stronger  $\beta = 10^{-2}$  was chosen for XCAT data with noise and the real patient data. For reconstructions simulating the absence of the anatomical information, a uniform image was used to replace the anatomical image  $z$  while calculating PLS. For the 2D disc phantom, this is equivalent to using a uniform anatomical image with no lesion feature (except for the object boundary).

### C. Analysis

The potential benefits of using the modified PLS were investigated in terms of visual interpretation, noise, local perturbation and algorithm convergence rate. To be able to quantify the last two features, we adopted two metrics to measure the local contrast recovery ratio ( $CR$ ) and the distance between the converged image and current estimate ( $M$ ). The

latter was initially introduced in [28]. Given the reconstructed image with lesion(s)  $\hat{\mathbf{x}}^{\text{total}}$  and that without lesion(s)  $\hat{\mathbf{x}}^{\text{bg}}$ , the metrics are defined as:

$$CR = \frac{|\text{ROImean}(\hat{\mathbf{x}}_c^{\text{total}} - \hat{\mathbf{x}}_c^{\text{bg}})|}{\text{True Difference}} \times 100\%, \quad (30)$$

$$M_t = \sqrt{\frac{1}{N} \frac{\|\hat{\mathbf{x}}_t^{\text{total}} - \hat{\mathbf{x}}_c^{\text{total}}\|_2^2}{\text{mean}(\hat{\mathbf{x}}_c^{\text{total}})^2}}, \quad (31)$$

where  $\text{ROImean}(\cdot)$  is an operator that calculates the mean value of the ROI, ‘‘True Difference’’ is the assigned activity difference between the lesion and its background and  $N$  is the number of voxels. The subscript  $t$  indicates the iteration number and  $c$  denotes that the converged image of L-BFGS-B-PC (see section II-E for the definition of convergence of L-BFGS-B-PC). The  $CR$  value defined in (30) is essentially a normalized difference in intensity between the lesion and background. To prevent additional bias introduced by the variation of intensity in location, we drew the ROI in the same place on  $\hat{\mathbf{x}}^{\text{total}}$  and  $\hat{\mathbf{x}}^{\text{bg}}$  while calculating  $CR$ . The value is comparable to the common concept of lesion contrast in which a local background region in the same image is used. The ROIs were drawn in the center of the sphere or lesion with size of  $9 \times 9$  voxels for the disc phantom and  $3 \times 3 \times 3$  voxels for both the XCAT phantom and real patient data.

In the first part of the study, we used 100 noise realizations of the 2D disc phantom to study the bias and variance trade-off of the proposed weighting. In that case, the average  $CR$  value for the disc phantom and the standard deviation (STD) for the voxel at the image center over 100 realizations were used respectively to quantify the bias (or local resolution) and noise properties of the estimated image with respect to different global penalty strengths. For quantifying the algorithm convergence rate, we used plots of the convergence measure  $M$  (31) against the total number of projection operations. One projection operation means a forward or back-projection of the full dataset. We chose the number of projection operations instead of iterations as it takes into account the extra computational demand induced by the backtracking algorithm. Fast decrease of  $M$  values indicates fast convergence rate to the solution of the reconstruction problem. To further explore the performance of the algorithm with and without applying the modified PLS, the required number of projection operations and the corresponding iterations for achieving ‘‘effective’’ convergence were computed as well. The corresponding iteration number was determined by:

$$t_M^* = \min \{t : M_t \leq 0.01\}. \quad (32)$$

Note that the dependence of the local perturbation on location was evaluated using noiseless data.

## IV. RESULTS

### A. Bias & variance analysis (disc phantom)

The noise realizations for the disc phantom with either hot or cold surroundings were used to investigate the potential influence of applying  $\hat{\kappa}$  on the bias and noise trade-off. We defined  $\hat{\kappa}_0$  as the value at the center of  $\hat{\kappa}$  calculated with the corresponding noiseless dataset for the phantom

with hot background region. Fig. 2 shows the bias and noise trade-off curves for different surroundings and reconstruction conditions. Since similar local resolutions were enforced at the center of the image reconstructed with and without applying the modified penalty function for the phantom with hot surroundings, the bias and noise trade-off points for these two reconstruction conditions are nearly identical when the same  $\beta$  is considered, in contrast to the case with the cold surroundings. For each evaluated global penalty strength  $\beta$ , much more similar  $CR$  values are obtained for hot and cold surrounding activities when  $\hat{\kappa}$  is applied than without using  $\hat{\kappa}$ . Example reconstructions at convergence for one realization for different surroundings are shown in Fig. 3 for  $\beta = 0.2$ . Although the sphere has a relatively high visual contrast to cold surroundings when  $\hat{\kappa}$  is not applied (Fig. 3 (f) and (h)), a more consistent lesion contrast, insensitive to the change of the surrounding activity, is observed for images reconstructed using  $\hat{\kappa}$  (Fig. 3, first and third columns). Consistent with the trade-off curves, similar reconstructed images are obtained for the phantom with hot surroundings, regardless of the use of  $\hat{\kappa}$  or not (Fig. 3, (a) and (b) for PLS and (c) and (d) for smooth TV). The images for the phantom with cold surroundings reconstructed with and without using  $\hat{\kappa}$ , however, are quite different as the effective penalty strength is different for each condition (Fig. 3, (e) and (f) for PLS and (g) and (h) for smooth TV). In the absence of anatomical information, all curves for different surrounding activities and penalization schemes shift to the lower left (Fig. 2, bottom, as compared to Fig. 2, top), indicating lower  $CR$  and STD values, hence smoother reconstructed images (Fig 3, third and fourth columns).

### B. Dependence of contrast ratio on location (XCAT phantom)

We used noiseless datasets for the XCAT phantom to study the dependence of local contrast (quantified by  $CR$ ) on different lesion locations. We selected the data generated using the phantom with high liver uptake as the reference dataset. Fig. 4 shows the central coronal view of the converged images reconstructed with and without applying  $\hat{\kappa}$  for the phantom with high liver uptake. The corresponding difference images obtained by subtracting the image reconstructed from data without lesions from the one with lesions are also provided to assist the visual comparison. As shown in the figure, a relatively uniform visual contrast for lesions at different locations is observed in reconstructions using  $\hat{\kappa}$  (Fig. 4 (a) and (c)). When  $\hat{\kappa}$  is not applied (Fig. 4 (b) and (d)), the lesion near the end slices has a lower visual contrast (indicated with purple arrows). Although not shown, similar results were obtained for data simulated with low liver uptake.  $CR$  values for each lesion under different data simulation and reconstruction conditions are calculated and listed in Table II. Consistent with the visual comparison, the variation of the contrast recovery ratio in locations is reduced when  $\hat{\kappa}$  is used. However, the influence of surrounding activity is not obvious for the XCAT phantom as lesion 5 and 6 show similar  $CR$  values when the activity level of their surroundings (liver) is changed, regardless of  $\hat{\kappa}$  being applied or not.

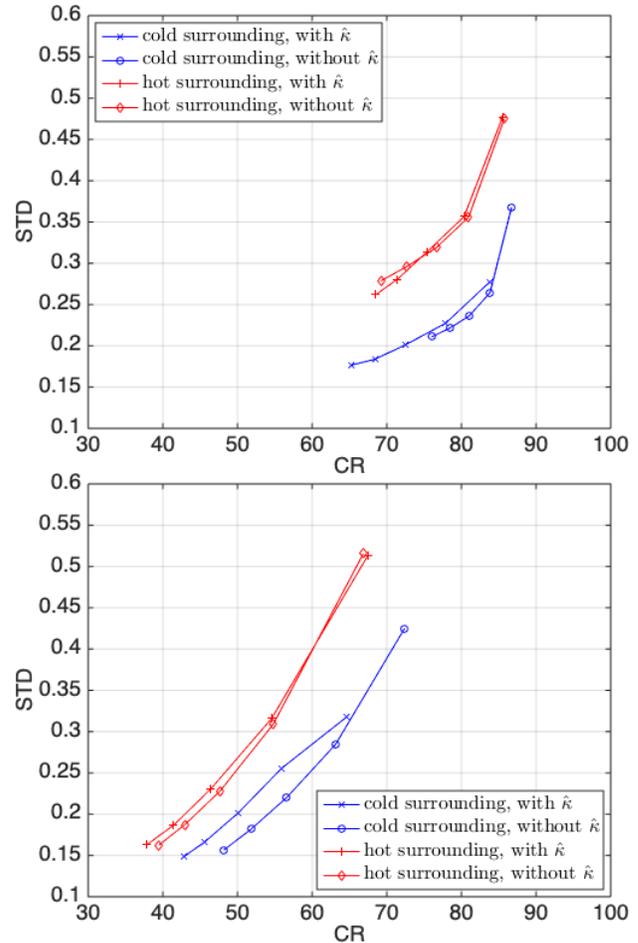


Fig. 2. Bias ( $CR$ ) and noise (STD) trade-off curves for PLS (top) and smooth TV (bottom) with and without applying  $\hat{\kappa}$ . The data were generated with the disc phantom with either hot or cold surrounding activity. Each marker indicates a value of the global penalty beta. The same values were used in sequence for all cases (see main text). When the anatomical information is available (top), the mean absolute differences in  $CR$  values between sphere in hot and cold surroundings are 2.68 and 4.16 for the reconstructions with and without using  $\hat{\kappa}$ . For the reconstructions without using the anatomical information (bottom), they are 3.43 and 8.08, respectively.

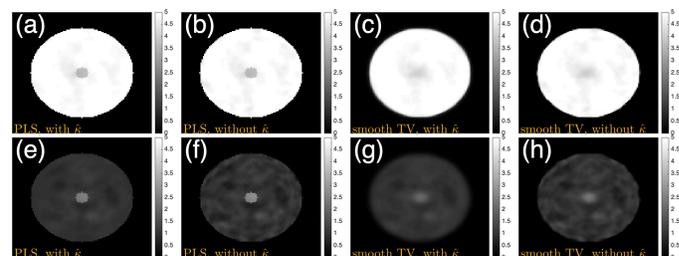


Fig. 3. Converged images for the disc phantom with high (top row) or low (bottom row) activity uptake in surroundings ( $\beta = 0.2$ ). Images reconstructed with and without using  $\hat{\kappa}$  are presented in odd and even columns, respectively. The first two columns are for PLS, while the last two are for smooth TV.

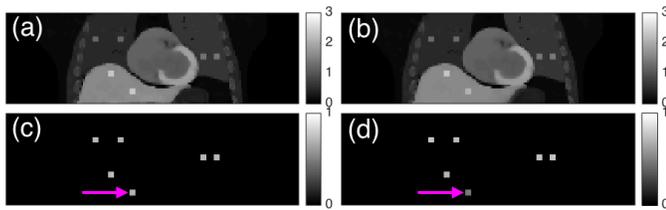


Fig. 4. Converged images reconstructed with (a) and without (b) using  $\hat{\kappa}$  for the XCAT phantom having high liver uptake. The corresponding difference images obtained by subtracting the image reconstructed from data without lesions from the one with lesions are given in (c) and (d), respectively.

TABLE II

CR VALUES (30) FOR LESIONS AT DIFFERENT LOCATIONS IN THE XCAT PHANTOM.

		1	2	3	4	5	6	mean $\pm$ std
with $\hat{\kappa}$	HL <sup>1</sup>	91	87	89	90	92	88	89 $\pm$ 1.78
	CL	91	88	89	90	88	86	
without $\hat{\kappa}$	HL	95	93	95	96	89	59	88 $\pm$ 13.35
	CL	96	93	95	96	88	61	

<sup>1</sup> HL and CL stand for hot and cold liver uptake, respectively.

TABLE III

REQUIRED NUMBER OF PROJECTIONS AND ITERATIONS (IN BRACKETS) FOR ACHIEVING EFFECTIVE CONVERGENCE OF  $M$  VALUES (31) FOR DATASETS REPRESENTING DIFFERENT NOISE LEVELS.

	with $\hat{\kappa}$	without $\hat{\kappa}$
$S_{\text{tot}} = 50$ M	82 (39)	412 (200)
$S_{\text{tot}} = 250$ M	92 (45)	686 (334)
$S_{\text{tot}} = 1252$ M	132 (65)	812 (398)

### C. Dependence of convergence rate on noise & background level

Simulated data with  $S_{\text{tot}} = 50$  M, 250 M and 1252 M, representing high, medium and low noise level, were reconstructed with and without using  $\hat{\kappa}$  to study the influence of the noise level. The  $\hat{\kappa}$  for the dataset with  $S_{\text{tot}} = 250$  M was chosen as the reference. Fig. 5 shows the central coronal view of the converged image of each reconstruction. Noisier end slices (the bottom two slices) of the reconstructed images are observed when the modified PLS is applied compared to those without using  $\hat{\kappa}$ . The effect becomes less apparent as the noise level of the data decreases. In addition to visual comparison, the performance of the algorithm was also evaluated by plotting convergence estimates  $M$  (31) against the total projection operations (Fig. 6) and by listing the required number of projection operations and iterations to achieve the effective convergence defined in (32) (Table III). The former illustrates the convergence rate in early iterations while the latter gives an insight to late iterations. The convergence rate is improved significantly when applying  $\hat{\kappa}$ . Moreover, it also reduces the difference in convergence rate for data with different noise levels.

The effect of the background level on the performance of L-BFGS-B-PC with and without using  $\hat{\kappa}$  was investigated with data in both groups of fixed  $S_{\text{tot}}$  (G1) and fixed  $S_{\text{true}}$  (G2). The results were compared to those from the data having medium noise and background level ( $S_{\text{tot}} = 250$  M,  $S_{\text{true}} = 94$  M,  $S_{\text{bg}} = 156$  M and TBR = 0.6). The reference strength

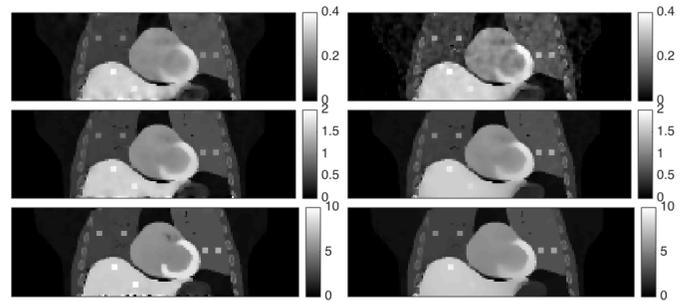


Fig. 5. Converged images reconstructed with (left column) and without (right column) using  $\hat{\kappa}$  for the XCAT phantom with high liver uptake. From top to bottom row are images for high, medium and low noise level.

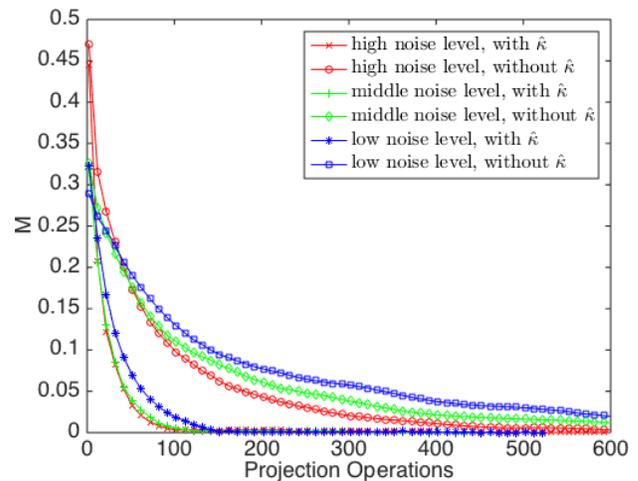


Fig. 6. Convergence estimates  $M$  (31) plotted against the total number of projection operations for datasets representing different noise levels.

TABLE IV

REQUIRED NUMBER OF PROJECTIONS AND ITERATIONS (IN BRACKETS) FOR ACHIEVING EFFECTIVE CONVERGENCE OF  $M$  VALUES (31) FOR DATASETS REPRESENTING DIFFERENT BACKGROUND LEVELS.

		with $\hat{\kappa}$	without $\hat{\kappa}$
G1	TBR = 3	82 (40)	644 (313)
	TBR = 0.12	92 (43)	522 (251)
G2	TBR = 3	82 (40)	512 (253)
	TBR = 0.12	142 (56)	624 (300)

$\hat{\kappa}_0$  was also calculated with this dataset. For both groups of data, the converged images are slightly more affected by noise when the modified PLS is used (Fig. 7 and 8). In terms of convergence rate, plots of  $M$  values against the total number of projection operations for each dataset in both groups are provided in Fig. 9. The required number of projection operations for achieving the effective convergence of  $M$  values are listed in Table IV. Based on the results, the use of the spatially-variant penalty strength shows the ability to improve not only the convergence speed in both early and late iterations but also the consistency of the convergence rate among data with different background levels.

### D. Demonstration with Patient Data

Fig. 10 shows the coronal plane through the inserted lesions for both reconstructions with and without using  $\hat{\kappa}$  for the

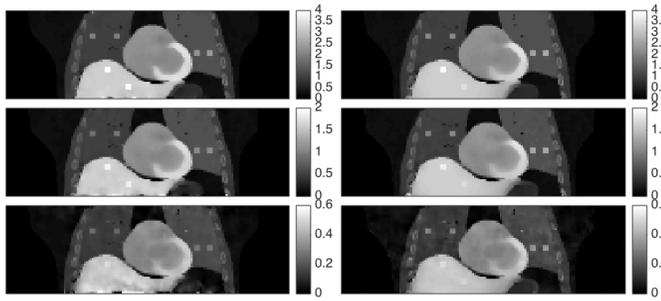


Fig. 7. Converged images reconstructed with (left column) and without (right column) using  $\hat{\kappa}$  for data in G1. From top to bottom row are images for high, medium and low TBR.

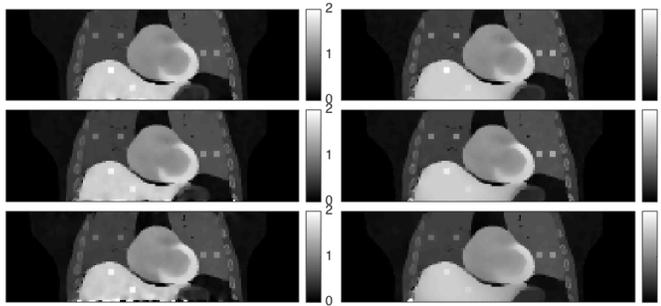


Fig. 8. Converged images reconstructed with (left column) and without (right column) using  $\hat{\kappa}$  for data in G2. From top to bottom row are images for high, medium and low TBR.

TABLE V  
CR VALUES (30) FOR EACH PSEUDO LESION.

	1	2	3	mean $\pm$ std
with $\hat{\kappa}$	41	53	44	$46 \pm 6.25$
without $\hat{\kappa}$	55	49	17	$40 \pm 20.43$

TABLE VI  
REQUIRED NUMBER OF PROJECTIONS AND ITERATIONS (IN PARENTHESES) FOR ACHIEVING EFFECTIVE CONVERGENCE OF  $M$  VALUES (31) FOR THE PATIENT DATASET WITH LESIONS.

	with $\hat{\kappa}$	without $\hat{\kappa}$
$S_{\text{tot}} = 219 M$	42 (20)	262 (129)

patient data. A relatively uniform visual contrast for lesions at different locations is observed when  $\hat{\kappa}$  is available.  $CR$  values for all lesions in both reconstruction conditions are listed in Table V. Consistent with the visual comparison, a smaller variation in the contrast recovery ratio is obtained for data reconstructed using  $\hat{\kappa}$ . The required number of projection operations for achieving the effective convergence (32) is also provided in Table VI. Again, the convergence rate of the reconstruction is substantially improved when using  $\hat{\kappa}$ .

## V. DISCUSSION

This study demonstrates the feasibility of using the spatially-variant penalty strength  $\hat{\kappa}$  with an anatomical penalty to reduce the dependence of local contrast (and LPR) on surroundings and location. Note that the contrast will still depend on the local features in the emission image [16] and the anatomical information. Potential benefits of applying  $\hat{\kappa}$

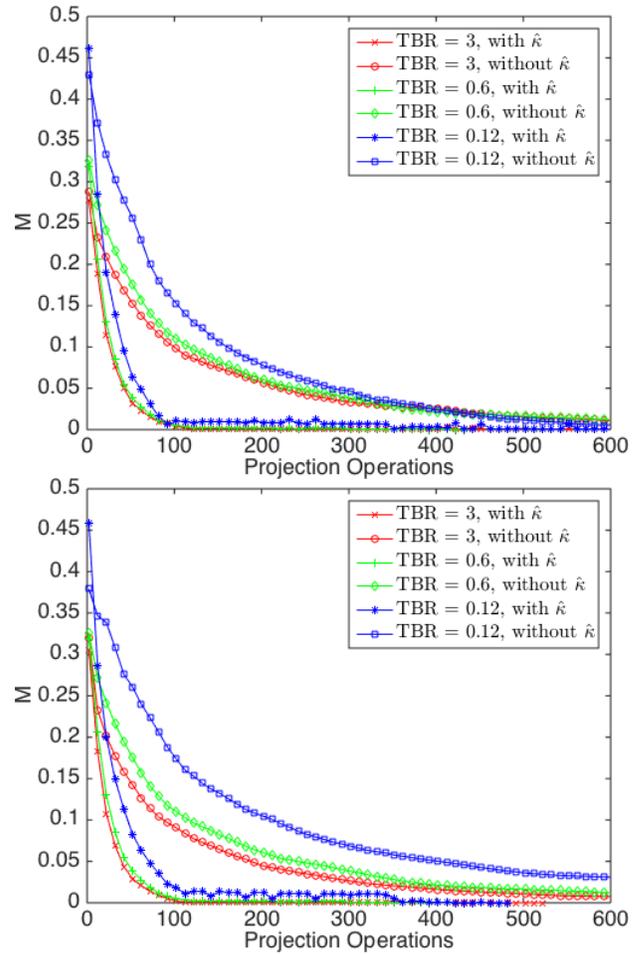


Fig. 9. Convergence estimates  $M$  (31) plotted against the total number of projection operations for data sets in G1 (top) and G2 (bottom).

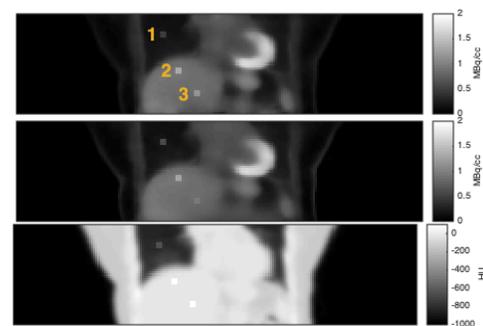


Fig. 10. Converged images reconstructed with (top) and without (middle) using  $\hat{\kappa}$  for the patient dataset with inserted pseudo lesions. The lesions are marked from 1 to 3 as illustrated in the top image. The corresponding coronal view of the enhanced CT is provided at the bottom.

to the convergence rate and convergence consistency are also explored with a previously proposed reconstruction algorithm, L-BFGS-B-PC.

As discussed in [16], the dependence of the local perturbation on surrounding activity can be explained using the analytical model defined in (17). For any voxel  $j$  in the image, the local perturbation around  $j$  depends on the strength of  $\kappa_j(\tilde{x}_b)$ . When the activity level of  $\tilde{x}_b$  is high around  $j$ , a small  $\kappa_j(\tilde{x}_b)$  hence a strong penalization around  $j$  is introduced in the original (unweighted) case. In this study, we used a 2D disc phantom and a 3D XCAT phantom representing  $^{18}\text{F}$ -FDG distribution in the thorax region to investigate the influences of surrounding activity and location on the local contrast, respectively. Different activity levels were assigned to the liver in the XCAT phantom to simulate the change of surroundings for lesions in the liver as well. For the disc phantom, the dependence of the local contrast on different surroundings is reduced by applying  $\hat{\kappa}$  (Fig. 2). Regardless of the use of  $\hat{\kappa}$ , the dependence of the contrast recovery on surroundings is not obvious for the XCAT phantom, judging from the small differences in  $CR$  values for lesions in different liver uptakes (lesion 5 and 6 in Table II). In contrast, without applying  $\hat{\kappa}$  there was a large difference in  $CR$  for lesions near the center of the FOV vs. the end slices (lesion 6 in Table II). This implies that the main reason for variation in local contrast when using spatially-invariant penalty weights is location for  $^{18}\text{F}$ -FDG studies using 3D PET. In fact, for a uniform activity distribution,  $\kappa_j(\tilde{x}_b)$  can be interpreted as an index of the spatial variations in sensitivity. Using the spatially-variant penalty strength is therefore roughly equivalent to applying sensitivity compensation across the FOV.

The results for the bias and variance trade-off indicated that the reconstructions with and without using  $\hat{\kappa}$  follow an essentially identical curve while the local resolution is matched (Fig. 2, curves for hot surroundings). This implies that one can achieve similar local properties for images reconstructed with and without applying  $\hat{\kappa}$  by tuning  $\beta$ . Similar observations were reported in [7] for quadratic penalization. In the case of applying  $\hat{\kappa}$ , there would therefore be a trade-off between the image noise level and the stability of LPR, especially for lesions near the end slices. Fortunately, as multiple bed positions are often used in clinical practice, the increased noise level of end slices can be compensated by averaging across neighboring bed positions. Although it seems likely that the bed position averaging would help with the stability of the lesion uptake values in the case without using  $\hat{\kappa}$ , the potential dependence on the surrounding activity and a slower and relatively inconsistent convergence rate among different datasets are still inevitable.

Regardless of the use of  $\hat{\kappa}$ , the LPR varies with the global strength  $\beta$  that controls the influence of the penalty over whole objective function and parameters that determine the weight of the edge information in the anatomical or functional image. For example, since  $\beta$  used for the patient dataset ( $\beta = 10^{-2}$ ) was larger than that for noiseless data generated using the XCAT phantom ( $\beta = 10^{-3}$ ), a lower  $CR$  value was observed for lesions in the patient dataset (Table II and V). When the same  $\beta$  is chosen, lesions in the patient

data are able to converge to similar  $CR$  values as those in the XCAT phantom (not shown). However, the reconstructed image becomes noisier due to the reduction of the penalty strength.

To be able to compare the results with and without  $\hat{\kappa}$ , we matched the local resolution at the center of the images reconstructed with a reference dataset. This means that different effective penalty strengths are introduced to lesions off the image center in different reconstruction conditions. It explains the relatively low  $CR$  values for lesions in the lung region of the XCAT phantom while  $\hat{\kappa}$  is used (lesion 1 to 4 in Table II). As the sphere is located at the image center for the 2D disc phantom, almost identical local contrasts are observed for reconstructions with and without using  $\hat{\kappa}$  when the reference dataset is considered (Fig. 2, curves for hot surroundings). Tuning the global penalty strength  $\beta$  as well as other parameters that control the local property of the penalty function would be necessary for particular application. The strength of  $\beta$  should be handled with caution for both reconstructions with and without using  $\hat{\kappa}$ . For unrealistically high  $\beta$ , the strong penalty across the boundaries can lead to a redistribution of the activity hence unpredictable  $CR$  values (results not shown).

The main motivation of rescaling  $\alpha$  and  $\eta$  in PLS according to the scale of the functional and anatomical images is to have a similar local influence from the penalty function for different datasets. In this way, the comparison of  $CR$  values depends mainly on the global parameter  $\beta$  and the use of the spatially-variant penalty strength  $\hat{\kappa}$ . As the smoothness enforced by PLS in a uniform region (or also across the boundary when  $\beta$  is large) is modulated by  $\alpha$  and  $\eta$ , the benefit of using  $\hat{\kappa}$  might be affected by these two parameters. This means that the redistribution of the activity across the boundaries for large  $\beta$  might be less severe with small  $\alpha$  such that using  $\hat{\kappa}$  would be less beneficial. However, as this study focuses on achieving consistent local contrast instead of accurate quantification, discussions on the dependence of the quantitative accuracy and image quality on parameters that control the global strength of the penalty or the strength of the local information are beyond the scope of this paper.

We have demonstrated that the convergence rate of the algorithm can be improved significantly when  $\hat{\kappa}$  is applied (Fig. 6 and 9). Moreover, the variability of the algorithm performance for different datasets is reduced. As shown in Table III and IV, all data representing different noise and background levels show the ability to achieve practical convergence in 150 projection operations. This implies that a more reliable quantitative comparison among different datasets can be obtained when a fixed number of projection operations is chosen. The reason for this improved convergence rate is not clear but we hypothesize that the optimization problem becomes better conditioned when using the spatially-variant weight. However, as here we used the previously proposed algorithm L-BFGS-B-PC with the same  $\hat{\kappa}$  used in the transformation matrix  $\mathbf{P}$  as for the penalty weight, investigating if the improved convergence rate is algorithm dependent remains future work. Note that when the matrix  $\mathbf{P}$  was constructed, we replaced the Hessian of the penalty function with a small constant

$\epsilon = 10^{-4}$ . This approximation of the Hessian of the objective function is suboptimal for large  $\beta$  and the convergence rate of the algorithm might become slow. For future investigation on convergence rate with large (but reasonable)  $\beta$ , it could be helpful to include the penalty while constructing  $\mathbf{P}$ .

In this study, we have proposed a new type of  $\kappa(\mathbf{x})$  that uses the square root of the row-sums of  $\mathbf{F}(\mathbf{x})$ . Since it was reused to construct the transformation matrix  $\mathbf{P}$  in L-BFGS-B-PC, no additional computation was required. The computational efficiency combined with reduced sensitivity to noise make the proposed  $\hat{\kappa}$  a practical substitution for  $\kappa(\mathbf{x})$ . A systematic comparison between the proposed and other types of  $\kappa(\mathbf{x})$ , such as the one defined in (15), in capturing the shift variance of the Fisher information matrix and improving the convergence rate as well as the convergence consistency of the reconstruction algorithm is left for future research. Although the performance of the proposed  $\hat{\kappa}$  can be influenced by the initial image, the results in the Appendix indicate that the proposed method with the suggested initial condition (1 full iteration of OS-EM with 35 subsets) outperforms the “plug-in” one in (24) and has no potential division-by-zero problem (2). However, as the SNR of the noisy dataset used for the comparison was relatively high, further investigations on the highest limit of data noise level for applying  $\hat{\kappa}$  to improve the uniformity of local contrast and algorithm performance are required.

For the XCAT phantom and patient dataset, we used simulated contrast enhanced CT images to provide anatomical information. This could have wide-ranging implications to the work-flow and economics of the procedure in clinical use. The overall radiation dose received by the patient can also be increased due to the additional contrast enhanced CT scan. Moreover, as the contrast agent does not accumulate uniformly in lesions in clinic, evaluations on the influence of the non-uniformity in local regions of the anatomical image to the proposed method are necessary. Since the application of the method is not limited to PET studies with contrast enhanced CT images, the anatomical information can be provided by other anatomical images, such as magnetic resonance (MR) images. In fact, for lesions involving density change, the attenuation map can simply be used to provide the corresponding anatomical information. The potential benefits and limitations of applying the proposed strategy with different anatomical resources should be assessed in future work.

## VI. CONCLUSION

In this study, we demonstrated the use of a spatially-variant penalty strength with a convex anatomical penalty function to reduce the dependence of local perturbation on surrounding activity and location. The proposed weighting scheme for the penalization can be precomputed. Moreover, when using L-BFGS-B-PC, the weights can be reused for constructing the preconditioner such that the overall computational demand remains unchanged. Based on the results for simulated data, the effective convergence rate of  $M$  values is considerably improved when the spatially-variant penalization is applied. Moreover, the variation of the convergence rate for different

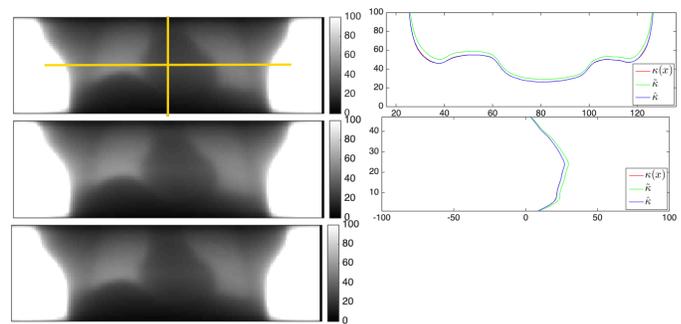


Fig. 11.  $\kappa(\mathbf{x})$ ,  $\tilde{\kappa}$  and  $\hat{\kappa}$  values (left column, top to bottom images) for the central coronal plane of the XCAT phantom with high liver uptake. Poisson noise was considered and the total counts of the data was 250 M. Horizontal and vertical profiles through the central point of each map are also provided (right column, top and bottom profiles, respectively).

data noise and background levels is considerably reduced when the weighting is taken into account. The idea is further demonstrated using a real patient dataset with inserted pseudo lesions in different slices. Consistent with the simulation results, significant improvement in quantitative consistency and algorithm performance is observed.

## APPENDIX A COMPARISON BETWEEN $\tilde{\kappa}$ & $\hat{\kappa}$

The comparison between  $\tilde{\kappa}$  (24) and  $\hat{\kappa}$  (25) was conducted with a simulated dataset for the XCAT phantom with high liver uptake in the presence of noise. The total counts of the data was 250 M, close to that of the patient data used in this study. The initial image was reconstructed by 1 full iteration of OS-EM with 35 subsets. For each approximation of  $\kappa(\mathbf{x})$ , the central coronal plane and profiles along its central point are provided in Fig. 11. Results were compared with those from  $\kappa(\mathbf{x})$  calculated with the corresponding noiseless dataset. Although the central coronal planes for both approximations are visually identical, profiles for  $\tilde{\kappa}$  and  $\hat{\kappa}$  are not the same. The former show higher values than those for  $\kappa(\mathbf{x})$ , while the latter are nearly indistinguishable from those for  $\kappa(\mathbf{x})$ . The results support the feasibility of using  $\hat{\kappa}$ , instead of  $\tilde{\kappa}$ , as an alternative to  $\kappa(\mathbf{x})$ .

## REFERENCES

- [1] J. M. Ollinger and J. A. Fessler, “Positron-emission tomography,” *IEEE Signal Process. Mag.*, vol. 14, pp. 43–55, 1997.
- [2] A. Alessio and P. Kinahan, *PET image reconstruction*, 2nd ed. Amsterdam Elsevier, 2006.
- [3] P. J. Green, “On use of the EM for penalized likelihood estimation,” *J. Roy. Stat. Soc. B*, vol. 52, no. 3, pp. 443–452, 1990.
- [4] L. Yang, J. Zhou, A. Ferrero, R. D. Badawi, and J. Qi, “Regularization design in penalized maximum-likelihood image reconstruction for lesion detection in 3D PET,” *Phys. Med. Biol.*, vol. 59, no. 2, pp. 403–19, 2013.
- [5] 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. 2194–204, 2012.
- [6] D. M. Higdon, J. E. Bowsher, V. E. Johnson, T. G. Turkington, D. R. Gilland, and R. J. Jaszczyk, “Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data,” *IEEE Trans. Med. Imag.*, vol. 16, no. 5, pp. 516–26, 1997.
- [7] J. A. Fessler and W. L. Rogers, “Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs,” *IEEE Trans. Im. Proc.*, vol. 5, no. 9, pp. 1346–58, 1996.

- [8] J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography," *IEEE Trans. Med. Imag.*, vol. 5, no. 3, pp. 493–506, 1996.
- [9] Q. Li, E. Asma, J. Qi, J. R. Bading, and R. Leahy, "Accurate estimation of the Fisher information matrix for the PET image reconstruction problem," *IEEE Trans. Image Processing*, vol. 23, no. 9, pp. 1346–58, 2004.
- [10] J. Qi and R. M. Leahy, "A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data," *IEEE Trans. Med. Imag.*, vol. 18, no. 4, pp. 293–305, 1999.
- [11] —, "Resolution and noise properties of MAP reconstruction for fully 3-D PET," *IEEE Trans. Med. Imag.*, vol. 19, no. 5, pp. 493–506, 2000.
- [12] J. W. Stayman and J. A. Fessler, "Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction," *IEEE Trans. Med. Imag.*, vol. 19, no. 6, pp. 601–15, 2000.
- [13] —, "Compensation for nonuniform resolution using penalized-likelihood reconstruction in space-variant imaging systems," *IEEE Trans. Med. Imag.*, vol. 23, no. 3, pp. 269–84, 2004.
- [14] J. Qi and R. H. Huesman, "Penalized maximum-likelihood image reconstruction for lesion detection," *Phys. Med. Biol.*, vol. 51, no. 16, pp. 4017–29, 2006.
- [15] J. Nuyts and J. A. Fessler, "A penalized-likelihood image reconstruction method for emission tomography, compared to postsmoothed maximum-likelihood with matched spatial resolution," *IEEE Trans. Med. Imag.*, vol. 22, no. 9, pp. 1042–52, 2003.
- [16] S. Ahn and R. M. Leahy, "Analysis of resolution and noise properties of nonquadratically regularized image reconstruction methods for PET," *IEEE Trans. Med. Imag.*, vol. 27, no. 3, pp. 413–24, 2008.
- [17] P. Cheebsumon, F. H. P. v. Velden, M. Yaqub, V. Frings, A. J. d. Langen, O. S. Hoekstra, A. A. Lammertsma, and R. Boellaard, "Effects of image characteristics on performance of tumor delineation methods: a test-retest assessment," *J. Nucl. Med.*, vol. 52, no. 10, pp. 1550–8, 2011.
- [18] A. Rahmim, J. Qi, and V. Sossi, "Resolution modeling in PET imaging: theory, practice, benefits, and pitfalls," *Med Phys.*, vol. 40, no. 6, pp. 064301–15, 2013.
- [19] V. Evilevitch, F. Eilber, W. Tap, M. Allen-Auerbach, M. Phelps, J. Czernin, and W. Weber, "Influence of ROI definition on the results of FDG-PET studies in patients with sarcomas," *J. Nucl. Med.*, vol. 48, no. Suppl 2, p. 145P, 2007.
- [20] M. Vanderhoek, S. B. Perlman, and R. Jeraj, "Impact of the definition of peak standardized uptake value on quantification of treatment response," *J. Nucl. Med.*, vol. 53, no. 1, pp. 4–11, 2012.
- [21] J. F. Eary, D. S. Hawkins, E. T. Rodler, and E. U. Conrad III, "<sup>18</sup>F-FDG PET in sarcoma treatment response imaging," *Am. J. Nucl. Med. Mol. Imaging*, vol. 1, no. 1, pp. 47–53, 2011.
- [22] M. Brändengen, K. Hansson, C. Radu, A. Siegbahn, H. Jacobsson, and B. Glimelius, "Delineation of gross tumor volume (GTV) for radiation treatment planning of locally advanced rectal cancer using information from MRI or FDG-PET/CT: a prospective study," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 81, no. 4, pp. 439–45, 2011.
- [23] D. E. Spratt, R. Diaz, J. McElmurray, I. Csiki, D. Duggan, B. Lu, and D. Delbeke, "Impact of FDG PET/CT on delineation of the gross tumor volume for radiation planning in non-small-cell lung cancer," *Clin. Nucl. Med.*, vol. 35, no. 4, pp. 237–43, 2010.
- [24] Y.-J. Tsai, G. Schramm, J. Nuyts, S. Ahn, C. W. Stearns, A. Bousse, S. Arridge, and K. Thielemans, "Spatially-variant strength for anatomical priors in PET reconstruction," *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. Rec.*, 2017.
- [25] J. E. Bowsher, H. Yuan *et al.*, "Utilizing MRI information to estimate F18-FDG distributions in rat flank tumors," in *IEEE Nucl. Sci. Symp. Conf. Rec.*, vol. 4, 2004, pp. 2488–2492.
- [26] M. J. Ehrhardt, P. Markiewicz, M. Liljeroth, A. Barnes, V. Kolehmainen, J. S. Duncan, L. Pizarro, D. Atkinson, B. F. Hutton, S. Ourselin, K. Thielemans, and S. R. Arridge, "PET reconstruction with an anatomical MRI prior using parallel level sets," *IEEE Trans. Med. Imag.*, vol. 35, no. 9, pp. 2189–99, 2016.
- [27] 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. 613–26, 2003.
- [28] Y.-J. Tsai, A. Bousse, M. J. Ehrhardt, C. W. Stearns, S. Ahn, B. F. Hutton, S. Arridge, and K. Thielemans, "Fast quasi-Newton algorithms for penalized reconstruction in emission tomography and further improvements via preconditioning," *IEEE Trans. Medical Imaging*, vol. 37, no. 4, pp. 1000–10, 2018.
- [29] Y.-J. Tsai, A. Bousse, M. J. Ehrhardt, B. F. Hutton, S. Arridge, and K. Thielemans, "Performance evaluation of MAP algorithms with different penalties, object geometries and noise levels," *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. Rec.*, 2015.
- [30] Y.-J. Tsai, A. Bousse, C. W. Stearns, S. Ahn, B. F. Hutton, S. Arridge, and K. Thielemans, "Performance improvement and validation of a new MAP reconstruction algorithm," *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. Rec.*, 2016.
- [31] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, 1995.
- [32] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Trans. Math. Softw.*, vol. 23, no. 4, pp. 550–560, 1997.
- [33] K. Thielemans, C. Tsoumpas, S. Mustafovic, T. Beisel, P. Aguiar, N. Dikaos, and M. W. Jacobson, "STIR: software for tomographic image reconstruction release 2," *Phys. Med. Biol.*, vol. 57, no. 4, pp. 867–83, 2012.
- [34] M. Teräs, T. Tolvanen, J. J. Johansson, J. J. Williams, and J. Knuuti, "Performance of the new generation of whole-body PET/CT scanners: Discovery STE and Discovery VCT," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 34, no. 10, pp. 1683–92, 2007.
- [35] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research," *Med. Phys.*, vol. 37, no. 9, pp. 4902–15, 2010.
- [36] Y. E. Chung, K. W. Kim, J. H. Kim, J. S. Lim, Y. T. Oh, J.-J. Chung, and M.-J. Kim, "Optimal delay time for the hepatic parenchymal enhancement at the multidetector CT examination," *J. Comput. Assist. Tomogr.*, vol. 30, no. 2, pp. 182–8, 2006.
- [37] F. Gao, M. Li, Y. Sun, L. Xiao, and Y. Hua, "Diagnostic value of contrast-enhanced CT scans in identifying lung adenocarcinomas manifesting as GGNs (ground glass nodules)," *Medicine (Baltimore)*, vol. 96, no. 43, pp. 1–6, 2017.
- [38] H. Ohkubo, Y. Kanemitsu, T. Uemura, O. Takakuwa, M. Takemura, K. Maeno, Y. Ito, T. Oguri, N. Kazawa, R. Mikami, and A. Niimi, "Normal lung quantification in usual interstitial pneumonia pattern: the impact of threshold-based volumetric CT analysis for the staging of idiopathic pulmonary fibrosis," *PLoS One*, vol. 11, no. 3, pp. 1–13, 2016.