An Ordered-Subsets Proximal Preconditioned Gradient Algorithm for Total Variation Regularized PET Image Reconstruction

A. Mehranian, A. Rahmim, *Senior Member, IEEE*, M. R. Ay, *Member, IEEE* , F. A. Kotasidis, *Member, IEEE*, and H. Zaidi, *Senior Member, IEEE*

Abstract– **Statistical variability of the PET data pre-corrected for random coincidences or acquired in sufficiently high count rates can be approximated by a Gaussian distribution, which results in a penalized weighted least-squares (PWLS) cost function. In this study, a proximal preconditioned gradient algorithm accelerated with ordered subsets (PPG-OS) is proposed for the optimization of the PWLS function, while addressing its two challenges encountered by previous algorithms such as separable paraboloidal surrogates accelerated with ordered-subsets (SPS-OS) and preconditioned conjugate gradient. First, the penalty and the weighting matrix of this function make its Hessian matrix ill-conditioned; thereby surrogate functions end up with high-curvatures and preconditioners would poorly approximate the Hessian matrix. The second challenge arises when using non-smooth penalty functions such as total variation (TV), which makes the PWLS function not amenable to optimization using gradient-based algorithms. To deal with these challenges, we used a proximal point method to surrogate the PWLS function with a proxy, which is then split into a preconditioned gradient descent and a proximal mapping associated with the TV penalty. A dual formulation was used to obtain the proximal mapping the TV penalty and also its smoothed version, i.e. Huber penalty. The proposed algorithm was studied for three different diagonal preconditioners and compared with the SPS-OS algorithm. Using simulation studies, it was found that the proposed algorithm achieves a considerably improved convergence rate over the state-of-the-art SPS-OS algorithm. Bias-variance performance of the algorithm was th evaluated for the preconditioners. Finally, the proposed PPG-OS algorithm was assessment using clinical PET data.**

A. Mehranian is with Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva, Switzerland (e-mail: mehrani1@etu.unige.ch)

A. Rahmim is with the Department of Radiology, School of Medicine, and Department of Electrical & Computer Engineering, Johns Hopkins University, Baltimore, USA (e-mail: arahmim1@jhmi.edu)

M.R. Ay is with the Department of Medical Physics and Biomedical Engineering, and Research Center for Molecular and Cellular Imaging, Tehran University of Medical Sciences, Tehran, Iran (e-mail: mohammadreza_ay@sina.tums.ac.ir)

F. Kotasidis with Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva, Switzerland (e-mail: fotis.kotasidis@unige.ch)

H. Zaidi is with the Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva, Switzerland, Geneva, Neuroscience Center, Geneva University, Geneva, Switzerland, and Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, Netherlands (e-mail: habib.zaidi@hcuge.ch)

I. INTRODUCTION

MODEL-based iterative image reconstruction algorithms can substantially improve the quality and quantitative accuracy of Positron Emission Tomography (PET) images. In fact, model-based algorithms can explicitly account for the physical and statistical processes involved in emission tomography and thus allow for a rich description of the image formation process leading to diagnostic and quantitatively more reliable images. The inherent statistical variability of prompt PET measurements is best described by the Poisson probability distribution, however, when the measurements are corrected for random coincidences and/or acquired in a high count rate, they are no longer Poisson distributed and their statistics can be described by a shifted Poisson model [1] or for practical advantages by Gaussian distribution, which approximates the Poisson distribution to second order. The quadratic approximation of the Poisson distribution or equivalently making use of the Gaussian model leads to a weighted least squares data-fitting term, which when augmented by a penalty or regularizer, to penalize deviations from an a priori knowledge, results in a penalized weighted least-squares (PWLS) cost function [2].

Several efficient iterative algorithms have been proposed for optimizing this cost function such as separable paraboloidal surrogates accelerated with ordered-subsets (SPS-OS), block coordinate descent (BCD) and preconditioned conjugate gradient (PCG) methods [3, 4]. The PCG-type algorithms are of special interest, because they converge rapidly and update all pixels of the solution simultaneously, and hence can be used for parallelization on multi-core computers. However, there are two challenges encountered when developing optimization algorithms for the PWLS cost function. First, the ill-conditioning of its second-order derivative (Hessian), due to the weighting matrix (see eq. 5) and the penalty function. As a result of ill-conditioning, the surrogate functions end up with high curvatures [5] and gradient-based algorithms with small step-sizes, leading to slow convergence. In addition, the preconditioners, used to improve the condition number of the Hessian and thus to speed up the convergence, would poorly act on the resulting ill-conditioned Hessian matrix. The second challenge arises when using non-smooth penalty functions such as total variation (TV) and wavelet-based analysis penalties, which use the ℓ_1 norm of the gradient and the wavelet coefficients of the image solution. As these penalty functions are not continuously differentiable, the resulting PWLS cost functions are not amenable to optimization using conventional gradient-

based algorithms. To deal with these challenges, some special shift-invariant weighting matrices and preconditioners have been proposed in literature, however, with some assumptions and computational complexities [3, 6]. A conventional approach to address the non-differentiability of the TV penalty is to smooth the ℓ_1 norm with a Huber norm. However, these challenge has been recently addressed using the variable splitting and split Bregman-type schemes, which have emerged as computationally appealing approaches [7, 8]. In [9], the variable splitting and alternating direction minimization techniques are used to reduce the optimization problem into independent sub-problems toward addressing the above mentioned challenges.

To deal with the ill-conditioning of Hessian matrix due to the penalty function and the optimization issue arising from the non-smooth TV penalty function, in this study, we followed a proximal point technique to surrogate the PWLS cost function with a proxy function. The proxy was then split into a preconditioned gradient descent (accelerated with ordered subsets) and a proximal mapping associated with the TV regularizer. The performance of the proposed algorithm was studied for three different preconditioners and compared was the SPS-OS algorithm in a simulation study. Finally, the proposed algorithm was applied for image reconstruction of clinical PET measurements.

II. PROBLEM FORMULATION

Let $x \in \mathbb{R}^N$ be a 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 any corrections), the measured prompt coincidences, $y^p \in \mathbb{Z}^M$ is a Poisson-distributed random variable described by:

$$y_i^p \approx \text{Poisson}\{n_i a_i \sum_{j=1}^N G_{ij} x_j + \bar{r}_i + \bar{s}_i\}, \quad (1)$$

where n, a, $\bar{s}$ and $\bar{r} \in \mathbb{Z}^M$ account for detector efficiency, photon attenuation, and expected scatter and random coincidences, respectively. The negative Poisson log-likelihood for observing the prompts y^p is then given by:

$$\Phi(x) = \sum_{i=1}^M h_i([Gx]_i),$$
$$h_i(y_i) = (n_i a_i y_i + \bar{r}_i + \bar{s}_i) - y_i^p \log(n_i a_i y_i + \bar{r}_i + \bar{s}_i), \quad (2)$$

The true object distribution is then estimated by the minimization of the above log-likelihood. When the number of counts is sufficiently high or the measurements are pre-corrected for randoms and scatters, $h_i(y_i)$ can be approximated to its second order, resulting to weighted least squares cost function [10, 11]. Applying a second-order Taylor's expansion to $h_i(y_i)$ around the estimate of ith line of response, $\hat{y}_i$, yields:

$$h_i(y_i) \approx h_i(\hat{y}_i) + \dot{h}_i(\hat{y}_i)(y_i - \hat{y}_i) + \frac{1}{2}\ddot{h}_i(\hat{y}_i)(y_i - \hat{y}_i)^2, (3)$$

where $\dot{h}_i(y_i) = n_i a_i[1 - \frac{y_i^p}{n_i a_i y_i + \bar{r}_i + \bar{s}_i}]$, $\ddot{h}_i(y_i) = y_i^p \left(\frac{n_i a_i}{n_i a_i y_i + \bar{r}_i + \bar{s}_i}\right)^2$.

It is worth noting here that in the E-step of the standard OSEM reconstruction algorithm, the function $h_i(y_i)$ is also, in effect, approximated by an easy-to-optimize surrogate function [4]. Using the estimate $\hat{y}_i = (y_i - \bar{r}_i - \bar{s}_i)/n_i a_i$ in (3), we then arrive at:

$$h_i(y_i) \approx y_i^p - y_i^p \log(y_i^p) + \frac{1}{2}\frac{n_i^2 a_i^2}{y_i^p}(y_i - \hat{y}_i)^2 \quad (4)$$

where $\hat{y}$ is in fact a fully pre-corrected emission sinogram whose statistics no longer follows Poisson distribution due to randoms and scatters subtraction. Note that this second-order approximations of the $h_i(y_i)$ in effect results in a negative Gaussian log-likelihood. By dropping terms independent of y_i in (4), the Poisson log-likelihood therefore reads the following weighted least squares cost function:

$$\phi(x) = \frac{1}{2}\sum_{i=1}^M w_i([Gx]_i - \hat{y}_i)^2, \quad w_i = \frac{n_i^2 a_i^2}{y_i^p} \quad (5)$$

When the prompt emission sinogram is not available, the prompt coincidences can be estimated from the pre-corrected sinogram $\hat{y}$, thereby the weights w_i in (5) are obtained by:

$$w_i = n_i^2 a_i^2 / \max(1, n_i a_i \hat{y}_i + \bar{r}_i + \bar{s}_i). \quad (6)$$

In the above equation, we threshold the minimum of the estimated prompts to one count; this approach avoids division by zero, and also ensures the non-negativity of the resulting weights. To control noise in the estimation of x and generally to penalize estimates deviating from an a priori knowledge, the cost function in (5) can be augmented by a penalty or regularizer, $R(x)$, which results in the following penalized weighted least-squares (PWLS) cost function [12]:

$$\psi(x) = \frac{1}{2}\|Gx - \hat{y}\|_W^2 + \beta R(x), \quad (7)$$

where $W = \text{diag}\{w_i\}$ and $\beta > 0$ is a regularization parameter that controls the balance between the fidelity of Gx to measured data and the regularity of x to our prior knowledge.

In this study, we are interested in a three-dimensional isotropic total variation (TV) regularizer in order to impose a smoothness prior knowledge in the reconstruction of a volume image. The TV regularizer effectively suppresses noise while preserving sharp edges [13]. This feature makes it particularly attractive for PET-guided tumor volume delineation and shape identification. A 3D isotropic TV regularizer is defined as:

$$R_{\text{TV}}(x) = \|Dx\|_1 = \sum_{j=1}^N |[Dx]_j| \quad (8)$$

which is the ℓ_1 norm of the discrete gradient of x. $D = [D_v^T, D_h^T, D_a^T]^T \in \mathbb{R}^{3N \times N}$ is a derivative matrix composed of first-order finite difference matrices (in horizontal, vertical and axial directions) and $|[Dx]_j| = ([D_v x]_j^2 + [D_h x]_j^2 + [D_a x]_j^2)^{1/2}$, $[Dx]_j \in \mathbb{R}^3$. Generally, the ℓ_p norm of an N-length vector is defined as:

$$\|x\|_p = \begin{cases} (|x_1|^p + \cdots + |x_N|^p)^{1/p}, & 1 \leq p < \infty \\ \max(|x_1|, \ldots |x_N|), & p = \infty \end{cases} \quad (9)$$

As the TV regularizer is non-smooth and thus edge preserving, it sometimes results in artificially flat regions (stair-casing effect) in the reconstructed images. Furthermore, it is not differentiable at points where $Dx = \mathbf{0}$. As a result, the PWLS cost function is amenable to optimization using gradient-based approaches. Hence, the ℓ_1 norm in (8) is usually smoothed by the Huber potential function, φ_{H}, resulting in a Huber (or smoothed TV) regularizer as follows:

$$R_{\text{H}}(x) = \sum_{j=1}^N \varphi_{\text{H}}(|[Dx]_j|), \quad \varphi_{\text{H}}(t) = \begin{cases} t - \delta/2, & t \geq \delta \\ t^2/2\delta, & t < \delta \end{cases} \quad (10)$$

where $\delta > 0$ controls the edge-preserving effect of the resulting penalty function. Note that for $\delta = 0$, the Huber regularizer reduces to the TV one.

III. Proposed Algorithm

To optimize the PWLS function in (7), we follow a proximal point method [14, 15], through which one add a proximal penalty term to the original cost function at each iteration, thereby an easy-to-optimize proxy or surrogate function, $\psi(x, x^k)$, is obtained. The optimization of the proxy guarantees that of the original cost function. Hence, the solution can be iteratively estimated by:

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}_+^N} \psi(x, x^k),$$

$$\psi(x, x^k) \triangleq \psi(x) + \frac{1}{2}\|x - x^k\|_L^2 \quad (11)$$

where $L = \pi I_M - G^T W G$ is a positive definite matrix, in which I_M is an $M \times M$ identity matrix, $G^T W G$ is the Fisher information term of the Hessian of the PWLS functional, i.e. $\mathcal{H} = \nabla^2 \psi(x) = G^T W G + \beta \nabla^2 R(x)$ and π satisfies $\pi \geq \lambda_{\max}(G^T W G)/2$. $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue.

By some algebra we rewrite the proxy in Eq. (11) as follows:

$$\psi(x, x^k) = \phi(x^k) + (x - x^k)^T \nabla\phi(x^k) + \frac{\pi}{2}\|x - x^k\|^2 + \beta R(x) \quad (12)$$

where $\nabla\phi(x^k) = G^T W (Gx^k - \hat{y})$. Completing the square in Eq. (12), we then arrive at:

$$\psi(x, x^k) = \frac{\pi}{2}\left\|x - \left(x^k - \frac{1}{\pi}\nabla\phi(x^k)\right)\right\|^2 - \frac{1}{2\pi}\|\nabla\phi(x^k)\|^2 + \nabla\phi(x^k) + \beta R(x). \quad (13)$$

By this approach, one can split the optimization problem in Eq. (11) into two steps i) a gradient descent and ii) a proximal mapping associated with the regularizer $R(x)$:

$$\tilde{x} = x^k - \tau \nabla\phi(x^k), \quad (14)$$

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}_+^N} \frac{1}{2}\|x - \tilde{x}\|^2 + \tau\beta R(x) + C \quad (15)$$

where $\tau = 1/\pi$ is a step size and C is a constant. This is known as the proximal gradient algorithm in the literature [16, 17]. To improve the overall convergence rate of the algorithm, we upgrade the step (14) to a preconditioned gradient (PG) descent algorithm accelerated with ordered subsets; thereby the PPG-OS algorithm is derived.

A. Preconditioning and Optimal Step Size

Several preconditioners have been proposed in literature such as: $P_1 = \operatorname{diag}\{H_{ii}\}^{-1}$, which uses the inverse of the diagonal elements of the Hessian matrix [18], $P_2 = \operatorname{diag}\{H\mathbf{1}\}^{-1}$, where $\mathbf{1} \in \mathbb{R}^N$ and $P_3^k = \operatorname{diag}\{(x^k + \epsilon)/G^T\mathbf{1}\}^{-1}$, where $\epsilon > 0$ ensures the positivity of this EM-type preconditioner and $\mathbf{1} \in \mathbb{R}^M$[19]. The preconditioner P2 was inspired from the De Pierro's convexity method to the data fidelity term (see e.g. [20]). Other complicated preconditioners such as Fourier-based ones have also proposed in [3, 21]. In this study, we utilized the three above-mentioned diagonal preconditioners. As we extend step (14) to a preconditioned gradient descent, the step size τ should satisfy: $0 < \tau \leq 2/\lambda_{\max}(PG^T W G)$. Proof follows from Theorem 8.3 in [22]. To achieve the maximum amount of decrease of the cost function, the step size in Eq. (14) can be obtained by the following maximization:

$$\tau^k = \operatorname{argmax}_{\tau \geq 0} \phi(x^k + \tau P\nabla\phi(x^k)) \quad (16)$$

whereby, an optimal step length can be obtained by:

$$\tau^k = -\frac{(P\nabla\phi(x^k))^T \nabla\phi(x^k)}{(P\nabla\phi(x^k))^T G^T W G (P\nabla\phi(x^k))} \quad (17)$$

ALGORITHM 1: TV-PPG ALGORITHM

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

While $(\|x^{k+1} - x^k\|/\|x^k\|) \geq \eta$ **do**

1. $\nabla\Phi = G^T W (Gx^k - \hat{y})$
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. $k = k + 1$.
6. $\tilde{x} \leftarrow x^{k+1}$.
7. **While** $n \leq T$ **do**
   a. $x^n = [\tilde{x} - \beta D^T z^n]_+$
   b. $b = \frac{1}{\delta + \sigma}(\sigma z^n + Dx^n)$
   c. $z_i^{n+1} = \operatorname{proj}_{\{\|\cdot\|_\infty \leq 1\}}(b_i)$
   d. $n = n + 1$
8. $x^{k+1} \leftarrow x^n$

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

B. Proximal mapping

In order to solve the problem (15) for the non-continuously differentiable TV penalty, we make use of its Legendre-Fenchel dual formulation [23]. Generally, an ℓ_p norm can be equivalently defined as:

$$\|x\|_p = \max_{z \in D}\{z^T x\}, D = \{z: \|z\|_q \leq 1\} \quad (18)$$

where the dual feasible set D is the unit ball of an ℓ_q norm and q satisfies $1/q + 1/p = 1$ [24]. As the Huber penalty function generalizes the TV one, we re-define the Huber regularizer in duality context by [25]:

$$R_H(x) = \max_{z \in Q}\{z^T Dx\} - \frac{\delta}{2}z^T z,$$

$$Q = \{z: z \in \mathbb{R}^{3N}, \|z\|_\infty \leq 1\} \quad (19)$$

where Q is an ℓ_∞ unit ball and $\delta = 0$ reduces the Huber penalty to the TV one. The second term in the r.h.s. of (19) in fact accounts for the smoothness of the Huber penalty.

Substituting (19) in (15), we arrive at the following min-max problem:

$$\min_x \max_{z \in Q} S(x, z),$$

$$S \triangleq \frac{1}{2}\|x - \tilde{x}\|^2 + \beta(z^T Dx - \frac{\delta}{2}z^T z), \quad (20)$$

In brief, we then exchange the order of the min and max problems, minimize S with respect to x and substitute the resulting solution $x = \tilde{x} - \beta D^T z$ into the max problem. We address the non-negativity on this solution later within the body of the algorithm (Algorithm 1, step 7a). By multiplication with $-1/\beta^2$, we then cast the maximization into the following minimization problem:

$$\min_{z \in Q} M(\tilde{x}, z), \ M \triangleq \frac{1}{2}z^T DD^T z - \frac{1}{\beta}z^T D\tilde{x} - \frac{\delta}{2\beta}z^T z$$

$$= \frac{1}{2}\|D^T z - \tilde{x}/\beta\|_2^2 + \frac{\delta}{2\beta}\|z\|_2^2 + C \quad (21)$$

To find the solution of the above constrained minimization, we follow the proximal point method by adding the proximal penalty $\frac{1}{2}\|z - z^n\|_\Sigma^2$ to $M(\tilde{x}, z)$, where $\Sigma = (\alpha I_{3N})^{-1} - DD^T$, I_{3N} is a $3N \times 3N$ identity matrix and the parameter α should satisfy $0 < \alpha \leq 2/\lambda_{max}(DD^T)$. By some algebra, dropping constant terms and letting $\sigma = \beta/\alpha$, the minimizer of problem (11) then iteratively reads:

$$z^{n+1} = \text{argmin}_{z \in Q} \left\{ \frac{1}{2} z^T z - b^T z \right\},$$
$$b = \frac{1}{\delta + \sigma} \left(\sigma z^n + D(\tilde{x} - \beta D^T z^n) \right) \quad (22)$$

One can show that the solution of the above problem is the vector b projected into the constraint set Q, i.e.

$$z_j^{n+1} = \text{proj}_{\{\|\cdot\|_\infty \leq 1\}}(b_j). \quad (23)$$

To obtain the projection into an ℓ_∞ unit ball, we exploit the Moreau's decomposition defined for projectors as [26]:

$$b_j = \text{proj}_{\{\|\cdot\|_1 \leq 1\}}(b_j) + \text{proj}_{\{\|\cdot\|_\infty \leq 1\}}(b_j) \quad (24)$$

where the projection into an ℓ_1 unit ball is obtained by a soft-thresholding operator with a unit threshold [26]. Hence,

$$\text{proj}_{\{\|\cdot\|_\infty \leq 1\}}(b_j) = b_j - \begin{cases} b_j - \text{sign}(b_j), & |b_j| > 1 \\ 0, & |b_j| \leq 1 \end{cases} \quad (25)$$

Algorithm 1 summarizes the proposed algorithm. In this study, we made use of power iteration method to iteratively approximate the largest eigenvalue of the matrix DD^T [27], thereby one can declare $\alpha \leq 2/12$. To improve the convergence rate of the proposed PPG algorithm, we accelerated it with ordered subsets; thereby a PPG-OS algorithm was derived.

IV. RESULTS AND DISCUSSION

A. Phantom simulation

To qualitatively evaluate the performance of the proposed algorithm with respect to a ground truth, we performed PET acquisition in a simulated scanner with parallel strip-integral geometry (4.06 mm strip width, 130 radial samples, 204 angular samples uniformly spaced over 180 degrees). The system matrix describing this scanner was generated using the Image Reconstruction Toolbox (IRT) by Fessler *et. al.* [28]. A 2D 128×128 true object distribution together with its associated attenuation map was derived from the IRT's standard test image, which is based on the anthropomorphic Zubal phantom. The emission ground truth was reconstructed for 5 noise realizations in a simulated PET study for 5×10^6 total counts and 10% random and scatters coincidence rates. Fig. 1(a)–1(c) show the true image, the PET image of a give noise realization reconstructed by filtered back-projection and the attenuation map of emission data, respectively. The proposed PPG-OS algorithm was objectively compared with the state-of-the-art SPS-OS algorithm using 6 ordered subsets and for the preconditioners P_1, P_2 and P_3^k.
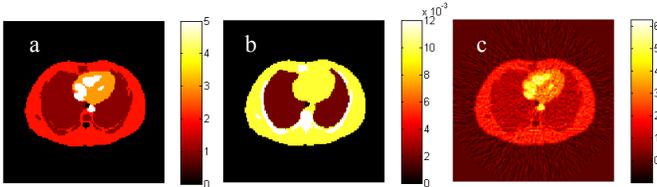
Fig. 1. (a) The true emission image simulating radiopharmaceutical uptake in a mediastinal lymphoma distribution. (b) The attenuation map used for attenuation correction of emission data, composed of linear attenuation coefficients of lung, soft tissue and bone (0.002, 0.0096 and 0.0120 mm^{-1}, respectively). (c) The image solution of a given noise realization reconstructed by filtered back-projection algorithm.

TABLE I. THE QUANTITATIVE PERFORMANCE OF THE STUDIED ALGORITHMS.

Algorithm	β	Five-trial average	Single-trial	
		RMSE (%)	RMSE (%)	Iterations
SPS-OS	20	13.43	13.68	77
	25	13.23	13.54	75
	30	13.25	13.52	75
	45	13.75	13.94	78
PPG-OS-P1	0.150	12.51	12.92	45
	0.020	12.45	12.79	44
	0.025	12.72	13.04	41
	0.030	13.09	13.37	40
PPG-OS-P2	0.150	12.96	12.88	50
	0.020	12.53	13.29	47
	0.025	13.49	13.76	45
	0.030	14.06	14.28	43
PPG-OS-P3	0.150	13.48	13.95	49
	0.020	13.25	13.62	44
	0.025	13.43	13.74	41
	0.030	13.77	14.08	38

All reconstructions were initialized with a zero initial image. Since the PWLS cost function with the non-smooth TV regularizer is not amenable to optimization using the SPS-OS algorithm, the Huber regularizer with $\delta = 0.02$ was considered instead in our comparative evaluations. The proximal mapping of the regularizer (step 7 in the Algorithm 1) was stopped after $T = 5$ iterations, which was found to be enough to perform the regularization step. As presented in Algorithm 1, a global convergence was declared when the relative change between two successive iterates fell below a tolerance of $\eta = 5 \times 10^{-4}$. Generally, the performance of the image reconstruction algorithms incorporating a regularizer (penalty function) strongly depends on the form of the penalty and regularization parameter, which controls the impact of regularizer on the solution. To access the performance of the algorithms, we heuristically chose four regularization parameters that result in the lowest root mean squared error (RMSE) between the true image, x^*, and the reconstructed image $\hat{x}$ after convergence. The RMSE is defined by:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(\hat{x}_i - x_i^*)^2}{N}} \quad (26)$$

Table 1 summarizes the average RMSE performance of the studied algorithms for the five noise realizations (five-trial average) and their performance for a single noise realization (single-trial).

As presented, the algorithms depicts nearly the same average RMSE performance for the chosen regularization parameters, however, PPG-OS-P1 algorithm has given rise to the lowest RMSE compared to other algorithms. The single trial results, on the other hand, show that the proposed PPG-OS algorithm achieves its RMSE performance in a considerably fewer iterations than the SPS-OS algorithm. The same results were found for the other noise realizations.

Fig 2 compares the single-trial RMSE performance of the algorithms as a function of the number of iterations. In this figure, the results of the PPG-OS and SPS-OS algorithms are respectively shown for $\beta = 0.02$ and $\beta = 30$, where the algorithms achieves the lowest RMSE. As seen, the proposed

Fig. 2 The root mean squared error (RMSE) of the studeid algorithms as a function of iteration. Note the the vertical axis is in logarithmic scale.

Fig. 3 The convergance of the studeid algorithms as a function of iteration. Note the the vertical axis is in logarithmic scale.

Fig. 4 The progression of the image estimated by the studied algorithms with k th iteration.

Fig. 4 shows the progression of the image solution with iteration toward the final solution. In consistent with Fig. 3, the results show the improved convergence rate of the proposed algorithm, particularly with the preconditioners P1 and P2, over the SPS-OS algorithm. It is noticeable that after 2 iterations the PPG-OS-P1 and PPG-OS-P2 algorithms have almost estimated the low frequency details. The quantitative performance of image reconstruction algorithms are often evaluated in terms of the mean and variance of the average intensity within a region of interest (ROI) over several noise realizations of projection data.

To assess the bias-vs.-variance performance of the proposed algorithm with the studied preconditioners, we followed a ROI-based quantitative approach to obtain the normalized mean squared error (NMSE), as a measure of bias, and the normalized standard deviation (NSD), as a measure of variance (noise). Generally, Two ROIs were chosen such that ROI 1 covered a hot spot and ROI 2 covered a background spot on soft tissue (see Fig. 5). The NMSE for each ROI was obtained by:

$$\text{NMSE} = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{\bar{x}^{(j)}-\bar{x}^*}{\bar{x}^*}\right)^2 \quad (27)$$

where $\bar{x}^{(j)} = \sum_i^n x_i^{(j)}$, $\bar{x}^* = \sum_i^n x_i^*$, m is the number of noise realizations and n is the number of voxels in the under-study ROI, $x^{(j)}$ is the image reconstructed from jth noise realization and x^* is the true object distribution. The NSD was also obtained by:

$$\text{NSD} = \frac{1}{n}\sum_{i=1}^{n}\frac{\sqrt{\frac{1}{m-1}\sum_{j=1}^{m}x_i^{(j)}-\bar{x}_i}}{\bar{x}_i} \quad (28)$$

where $\bar{x}_i = \frac{1}{m}\sum_{j=1}^{m}x_i^{(j)}$ which represents the ensemble mean value of voxel i. For each ROI, the NMSE values were plotted against NSD values as a function of iteration number. This approach in fact indicates the bias-variance trade off as an overall image quality index.

Figs. 5(a)–5(a) show the bias-variance performance of the proposed algorithm ($\beta = 0.02$) for ROI 1 and ROI 2, respectively. Each curve in the plots presents the NSD versus NMSE variation with the increased iteration number. In ROI 1, the PPG-OS-P1 algorithm shows lower bias and higher

algorithm reduces RMSE (shown in logarithmic scale) much faster than the SPS-OS algorithm, particularly in early iterations. Fig. 3 keeps track of the convergence rate of the algorithms by the log of the 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} (see Algorithm 1). Similar to Fig. 2, the results are shown for $\beta = 0.02$ for PPG-OS algorithms and $\beta = 30$ for SPS-OS algorithm. The results demonstrate the considerably improved convergence rate of the proposed algorithm over the SPS-OS.

As seen, the proposed algorithm with the preconditioner P1 depicts overall the fastest convergence. The PPG-OS-P1 and PPG-OS-P2 algorithms have nearly similar convergence rates, while PPG-OS-P3 lags behind them at early iterations and approaches them at last iterations. The same trend was almost observed for the other noise realizations and regularization parameters. It should be emphasized that the convergence behavior of PPG-OS with the preconditioner P3 depends on the parameter ϵ, used to ensure its positivity. In this study, we set $\epsilon = 0.01$. Larger values of ϵ improve the initial convergence of the resulting algorithm, however, they cannot ensure a conversance rate as fast as smaller values.

Fig. 5 The NSD versus NMSE for (a) ROI 1 and (b) ROI 2 as a function of iteration number.

Fig. 6 The convergance rate of the PPG-OS-P1 algorithm for optimal and fixed step lenghts as a function of iteration number.

variance, while the PPG-OS-P2 algorithm has a reciprocal behavior and shows lower variance and higher bias. In ROI 2, the PPG-OS-P2 outperforms the PPG-OS-P1 algorithm and show respectively the best and the worst bias-variance trade off. In both ROIs, the PPG-OS-P3 depicts the worst performance. Generally, the same behavior was observed for other regularization parameters. Note that due to its different convergence rate, the bias-variance performance of SPS-OS algorithm was found very different from the PPG-OS algorithm. In both ROIs, this algorithm showed lower variance and considerably higher bias for $\beta = 30$. The same performance was observed with other regularization parameters.

For the image reconstruction of clinical data, presented in the next section, we utilized the PPG-OS-P2, because i) according to simulation results, it showed a comparable convergence rate to the PPG-OS-P1 algorithms and it outperformed the PPG-OS-P1 in terms of bias-variance trade off, ii) the preconditioner P1 needs direct access to the Hessian matrix, which in turn necessitates the pre-computation of system matrix. However, for clinical datasets, system matrix is a very large-sized matrix that raises memory shortage issues. Contrary to P1, the preconditioner P2 does not need system matrix pre-computation and can be pre-computed using on-the-fly forward and back-projections.

Generally, the computation of an optimal step length (Eq. 17) for large-sized clinical data remarkably increases the overall reconstruction time of the proposed subset-ized algorithm, because of an additional forward projection of each descent direction, $P\nabla\phi(x^k)$, in each subset. To reduce computation time, we set up an experiment to find a fixed, near-to-optimal step length for the PPG-OS-P1 algorithm according to the condition $0 < \tau^k \leq 2/\lambda_{\max}(PH)$. Using the power iteration method, it was found that the largest eigenvalue of the P_2H matrix is near to unity for all noise realizations and for a strip- or line-integral geometry, i.e. $\lambda_{\max}(P_2H) \cong 1$. As a result, one can declare $0 < \tau \leq 2$. Fig. 6 compares the convergence rate of the PPG-OS-P2 algorithms for optimal and fixed step lengths. The results show that as the step length increases up to 2, the convergence rate of the resulting algorithm approaches to the one with optimal step length. Therefore, using $\tau = 2$, one can considerably save computation time, yet exploit a favorable convergence rate.

B. Experimental PET data

To qualitatively evaluate the proposed PPG-OS-P1 algorithm, the NEMA IEC image quality phantom was scanned and reconstructed as defined in the NU 2-2001 PET performance measurements [29]. The background in the phantom and the six spheres (10, 13, 17, 22, 28 and 37 mm in diameter) were filled with 50 MBq of [^{18}F] with a 4:1 sphere-to-background activity concentration, and data were acquired on a Biograph 6 Barrel-HiRez PET/CT scanner (Siemens Molecular Imaging Inc., TN, USA) for 60 minutes (~350 million prompts). The B-HiRez scanner acquires the emission data in 32-bit list-mode and operates in 3D mode, resulting in sinograms of size 336×336×313 within 5 segments. Data were histogrammed into a static sinogram using an in-house MATLAB-based list-mode histogrammer, thereby the prompts, randoms and net-true sinograms were generated. Standard corrections were applied on the net-true sinograms (attenuation, normalization and scatter), whereby the fully corrected sinogram $\hat{y}$ was obtained. To realize the impact of regularization on noise reduction, the data were reconstructed in 2D mode (i.e. from 81 direct and cross planes within segment 0). The Siddon's

algorithm [30] was used to drive a line-integral based geometric system matrix for the B-HiRez scanner. To have a fast and memory efficient image reconstruction, a hybrid of system matrix pre-computation and in-plane and axial symmetry translations was utilized. At the end, the preconditioner P2 was pre-computed and an image volume of the size of 336×336×81 was reconstructed by the proposed algorithm using 4 iterations and 8 subsets.

Fig. 7(a)–7(c) compares a representative slice of the reconstructed image without regularization ($\beta = 0$), with those reconstructed with 3D Huber regularization ($\delta = 0.02$) using the parameters $\beta = 0.005$ and $\beta = 0.01$, respectively. As seen, the regularized reconstructions have noticeably reduced noise and have successfully located the 10 and 13 mm hot spots, which are of importance in the diagnosis of small tumors. In Fig. 7(d), the quantitatively performance of the resulting algorithms has been compared using an intensity profile along the line shown in Fig. 7(c). The profiles show that regularizations have effectively reduced background noise, however, with a compromise on the estimated activity of the 22 mm hot spot. In general, the results showed that 3D regularization can effectively improve the diagnostic quality of the reconstructed images. The future work would be a more complete qualitative evaluation of the proposed algorithm in clinical datasets terms of bias-variance and contrast recovery.
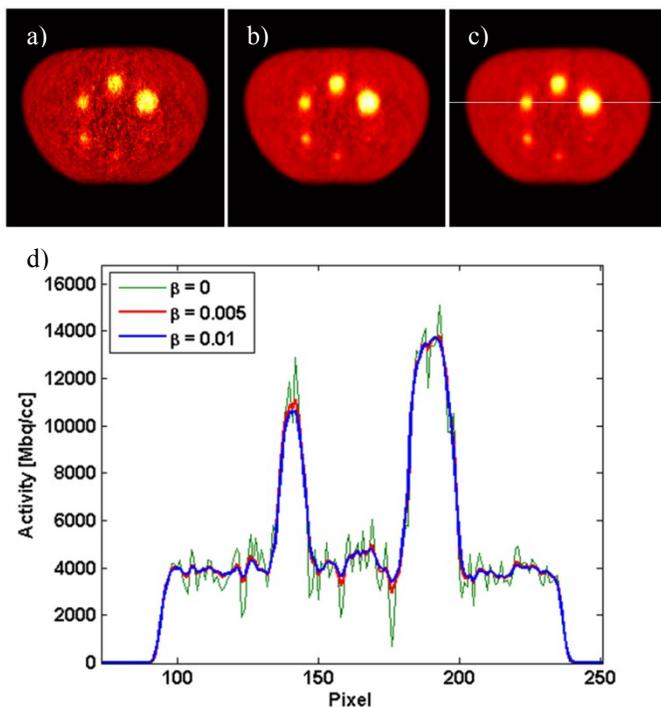
Fig. 7 The images reconstructed by the PPG-OS-P1 algorithm with the regularization parameters of (a) $\beta = 0$, (no regularization), (b) $\beta = 0.005$ and (c) $\beta = 0.01$. (d) The intensity profiles of the images in (a)–(c) along the line shown in (c). The display window is the same in all images.

V. CONCLUSION

We proposed a splitting-based preconditioned conjugate gradient algorithm, accelerated with ordered-subsets, for TV and Huber regularization in PWLS image reconstruction of PET data. The splitting allowed for decoupling of the regularizer from the PWLS cost function, thereby improving the ill-conditioning of its Hessian and thus convergence rate. In conclusion, the proposed algorithm proved efficient in image quality and algorithmic complexity and outperformed the SPS-OS algorithm.

REFERENCES

[1] M. Yavuz and J. A. Fessler, "Statistical image reconstruction methods for randoms-precorrected PET scans," *Medical Image Analysis,* vol. 2, pp. 369-378, 1998.

[2] J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. Medical Imaging,* vol. 13, pp. 290-300, 1994.

[3] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Im. Proc.,* vol. 8, pp. 688-699, 1999.

[4] J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography," *Phys. Med. Biol.,* vol. 51, pp. R541-R578, 2006.

[5] H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.,* vol. 44, pp. 2835-2851, 1999.

[6] A. H. Delaney and Y. Bresler, "A fast and accurate Fourier algorithm for iterative parallel-beam tomography," *IEEE Trans. Im. Proc.,* vol. 5, pp. 740-753, 1996.

[7] M. Afonso, *et al.*, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Im. Proc.,* vol. 19, pp. 2345-2356, 2010.

[8] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imaging Sci.,* vol. 2, pp. 323-43, 2009.

[9] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction," *IEEE Trans. Med. Imag.,* 2011. Submitted.

[10] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans. Med. Imag.,* vol. 21, pp. 89-99, 2002.

[11] K. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Processing,* vol. 41, pp. 553-548, 1993.

[12] J. A. Fessler, "Penalized Weighted Least-Squares Image Reconstruction for Positron Emission Tomography," *IEEE Trans. Med. Imag.,* vol. 13, pp. 290-300, 1994.

[13] L. Rudin, *et al.*, "Nonlinear total variation based noise removal algorithms," *Physica D,* vol. 60, pp. 259-268, 1992.

[14] J. Eckstein and D. Bertsekas, "On the DouglasRachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming,* vol. 5, pp. 293-318, 1992.

[15] S. Setzer, "Operator Splittings, Bregman Methods and Frame Shrinkage in Image Processing," *Int J Comput Vis,* vol. 92, pp. 265-280, 2011.

[16] A. Beck and M. Teboulle, "Gradient-Based Algorithms with Applications in Signal Recovery Problems," in *Convex Optimization in Signal Processing and Communications*, D. Palomar and Y. Eldar, Eds., ed: Cambribge University Press, 2010, pp. 33--88.

[17] W. Zuo and Z. Lin, "A generalized accelerated proximal gradient approach for total-variation-based image restoration," *IEEE Trans Image Process,* vol. 20, pp. 2748-2759, 2011.

[18] C. A. Johnson, *et al.*, "Interior-point methodology for 3-D PET reconstruction," *IEEE Trans. Med. Imaging,* vol. 19, pp. 271-85, 2000.

[19] L. Kaufman, " Maximum likelihood, least squares, and penalized least squares for PET," *IEEE Trans. Med. Imaging,* vol. 12, pp. 200-214, 1993.

[20] D. Kim and J. A. Fessler, "Accelerated ordered-subsets algorithm based on separable quadratic surrogates for regularized image reconstruction

in X-ray CT," in *Proc. IEEE Intl. Symp. Biomed. Imag.*, 2011, pp. 1132-7.

[21] N. H. Clinthorne*, et al.*, "Preconditioning methods for improved convergence rates in iterative reconstructions," *IEEE Trans. Med. Imag.,* vol. 12, pp. 78-83, 1993.

[22] E. P. Chong and S. H. Zak, *An Introduction to Optimization*, 2 ed.: John Wiley & Sons, 2001.

[23] A. Chambolle, "An algorithm for total variation minimization and applications," *J. of Math. Imaging and Vision,* vol. 20, pp. 89-97, 2004.

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.

[25] S. Becker*, et al.*, "NESTA: A fast and accurate first-order method for sparse recovery," California Institute of Technology,2009.

[26] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Multiscale Model. Simul.,* vol. 4, pp. 1168-1200, 2005.

[27] J. Kiusalaas, *Numerical Methods in Engineering with MATLAB*, 2 ed.: Cambridge University Press, 2010.

[28] J. A. Fessler. *Image Reconstruction Toolbox, http://www.eecs.umich.edu/~fessler/code/index.html*.

[29] M. E. Daube-Witherspoon*, et al.*, "PET performance measurements using the NEMA NU 2-2001 standard," *J. Nucl. Med.,* vol. 43, pp. 1398-1409, 2002.

[30] F. Jacobs*, et al.*, "A fast algorithm to calculate the exact radiological path through a pixel or voxel space," *J. Comput. Inform. Technol. ,* vol. 6, pp. 89-94, 1998.