HAL
open science

Noniterative sequential weighted least squares algorithm for positron emission tomography reconstruction.

Jian Zhou, Jean-Louis Coatrieux, Limin M. Luo

To cite this version:

Jian Zhou, Jean-Louis Coatrieux, Limin M. Luo. Noniterative sequential weighted least squares algorithm for positron emission tomography reconstruction.. Computerized Medical Imaging and Graphics, 2008, 32 (8), pp.710-9. 10.1016/j.compmedimag.2008.08.008 . inserm-00337826

HAL Id: inserm-00337826

https://inserm.hal.science/inserm-00337826

Submitted on 12 Nov 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Noniterative Sequential Weighted Least Squares Algorithm for Positron Emission Tomography Reconstruction

Jian ZHOU[1,2,3,4], Jean-Louis COATRIEUX[2,3,4], and Limin LUO[1,4]

[1]Laboratory of Image Science and Technology, Southeast University, 210096, China

[2]INSERM U642, Rennes, F-35000, France

[3]Université de Rennes 1, LTSI, Rennes, F-35000, France

[4]Centre de Recherche en Information Biomedicale Sino-francais (CRIBs), Rennes, France

Abstract

This paper proposes a new sequential weighted least squares (SWLS) method for positron emission tomography (PET) reconstruction. The SWLS algorithm is noniterative and can be considered as equivalent to the penalized WLS method under certain initial conditions. However, a full implementation of SWLS is computationally intensive. To overcome this problem, we propose a simplified SWLS as a reasonable alternative to the SWLS. The performance of this SWLS method is evaluated in experiments using both simulated and clinical data. The results show that the method can be advantageously compared with the original SWLS both in computation time and reconstruction quality.

Keywords: image reconstruction, emission tomography, weighted least squares, sequential.

I. Introduction

In modern PET scan systems, data corrections, e.g., for the effects of accidental coincidence (AC) event, have become necessary to yield PET image with accurate quantification of radiotracer activity. The consequence is that the precorrected data are no longer Poisson. The exact likelihood function describing the resulting data is difficult to work with. Therefore, it has been suggested that minimization of a weighted least squares (WLS) functional can be used [1]. Nevertheless, it is well known that tomography reconstruction is an ill-posed problem. As a result, a regularization or penalization method is usually recommended to stabilize the above WLS estimate. The corresponding estimation is called the penalized WLS estimation (PWLS).

Many investigators approach the PWLS estimate using iterative algorithms, which include, for example, the expectation-maximization (EM) algorithm [2][3], the gradient-based algorithm [4]–[8] and the coordinate-descent based algorithm [1][9]–[11] (the latter two are similar to some extent). All of them start with an initial guess

and then perform image updating iteratively until reaching some specified stopping criterions. But how to setup stopping criterions in some circumstances is not trivial [12]. Moreover, how fast to approach the required estimate depends on the convergent rate of the used iterative algorithm. The coordinate-descent algorithm usually offers more convergent rate than the EM or gradient-based algorithms. However, as pointed out in [3][11], this is likely at the cost of computational time.

In this paper, we develop a noniterative algorithm for the PWLS estimation. In [13], we have reported a sequential WLS (SWLS) method in which the data acquisition has been assumed to be a time related procedure. Although it is treated as an iterative method, the basic principle of SWLS based on the temporal filtering is noniterative. The SWLS method computes the image estimate using a block recursion procedure, which processes one block of data for every recursion and then proceeds to the next until all blocks are explored. The strategy is similar to the ordered-subsets (OS) technique (or block iteration) [14] in which each iteration consists of several sub-iterations and each sub-iteration uses only partial information of the measured data. The difference is that the blocks used in SWLS can not be overlapped. If one views data blocks as a time series, SWLS is then equivalent to the special Kalman filtering in which radioactivities (or the *states* in context of the theory of Kalman filtering) are supposed to be temporally unchanged. Such temporal filtering also has a strong relation to the PWLS method. Here, we will explore this relation by comparing the SWLS estimate with the corresponding PWLS estimate.

Unfortunately, like the Kalman filtering, SWLS suffers from the so-called "curse of dimensionality" which is caused by a major increase in computational complexity duo to the high dimension state or observation space. This problem is known to be difficult and usually task dependent. There are mainly two strategies that can be adopted. The first uses efficient numerical algorithms: an excellent overview of some solutions can be found out in [15]. The second way is to seek a suboptimal solution. This relies on our *a priori* knowledge about the problem being addressed, in order to find out some tradeoffs between performance and computational cost. In this paper, we consider the latter option and we propose a general strategy for the required suboptimal solution. We also discuss in detail a particular simplified SWLS approach for the problem of PET reconstruction.

The remaining of this paper is organized as follows. Section 2 briefly reviews PWLS and SWLS methods, which is followed by the development and implementation of simplified SWLS. In section 3, experiments are conducted by comparing SWLS's using both simulated data and clinical raw data. The results are given in section 4. Finally, in section 5, we conclude the paper and sketch the works in progress.

II. Method

A. The PWLS and sequential WLS estimation

Without loss of generality, let us consider the problem of two-dimensional PET image reconstruction. Denote $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_p]' \in \mathbb{R}^p$ as the unknown radioactivity distribution image vector that is to be reconstructed from the precorrected measurement vector $\boldsymbol{Y} = [Y_1, \ldots, Y_N]' \in \mathbb{R}^N$, where the prime denotes vector or matrix transpose.

The WLS objective function is as follows:

$$\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{A}\boldsymbol{\lambda})'\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{A}\boldsymbol{\lambda}) \tag{1}$$

where $\boldsymbol{A}$ is the system matrix. The weighting matrix $\boldsymbol{W}$ is diagonal with ith entry σ_i^{-2}. Here, σ_i^2 is an estimate of the variance of the ith precorrected measurement Y_i.

A simple but efficient penalty function that is often used in PET is the quadratic penalty:

$$\mathcal{R}(\boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{\lambda}'\boldsymbol{R}\boldsymbol{\lambda} \tag{2}$$

in which $\boldsymbol{R}$ is a well selected weighting matrix. This penalty has proven adequate to yield PET image reconstruction with satisfactory quality (see [1] for example). The PWLS estimate is achieved by minimizing the following objective function:

$$\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{A}\boldsymbol{\lambda})'\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{A}\boldsymbol{\lambda}) + \beta\mathcal{R}(\boldsymbol{\lambda}). \tag{3}$$

where the adjustable parameter β controls the strength of penalization. It has the closed-form expression:

$$\hat{\boldsymbol{\lambda}}_{\mathrm{PWLS}} = (\boldsymbol{F}+\beta\boldsymbol{R})^{-1}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{Y} \tag{4}$$

with $\boldsymbol{F}$ the Fisher information matrix: $\boldsymbol{F} = \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$.

According to the SWLS method [13], we partition the measurement data into M nonoverlapped blocks and denote $\boldsymbol{Y}_m$ the mth block. Then $\boldsymbol{Y} = [\boldsymbol{Y}_1' \mid \boldsymbol{Y}_2' \mid \ldots \mid \boldsymbol{Y}_M']'$. The block size can be arbitrary chosen but in this paper we only consider the normal case where each block has the same size $\ell \times 1$ (where $\ell = N/M$). Using similar notation, we further define the partitioned system matrix: $\boldsymbol{A} = [\boldsymbol{A}_1' \mid \boldsymbol{A}_2' \mid \ldots \mid \boldsymbol{A}_M']'$ with $\boldsymbol{A}_m$ the $\ell \times p$ matrix representing the partial projection operator corresponding to each partial measurement data $\boldsymbol{Y}_m$. Also we can rewrite the weight matrixing as $\boldsymbol{W} = [\boldsymbol{W}_1' \mid \boldsymbol{W}_2' \mid \ldots \mid \boldsymbol{W}_M']'$in which $\boldsymbol{W}_m$ is a $\ell \times \ell$ diagonal weighing matrix of $\boldsymbol{Y}_m$.

The SWLS method assumes that there is a known prior covariance[1] $\boldsymbol{P}_0$ and a starting estimate $\hat{\boldsymbol{\lambda}}_0$. Then, it performs the recursion for $m = 1, 2, \ldots, M$:

$$\hat{\boldsymbol{\lambda}}_m = \hat{\boldsymbol{\lambda}}_{m-1} + \boldsymbol{K}_m(\boldsymbol{Y}_m - \boldsymbol{A}_m\hat{\boldsymbol{\lambda}}_{m-1}) \tag{5}$$

where $\boldsymbol{K}_m$ $(p \times \ell)$ is calculated by

$$\boldsymbol{K}_m = \boldsymbol{P}_{m-1}\boldsymbol{A}_m'(\boldsymbol{A}_m\boldsymbol{P}_{m-1}\boldsymbol{A}_m' + \boldsymbol{W}_m^{-1})^{-1} \tag{6}$$

and

$$\boldsymbol{P}_m = (\boldsymbol{I}_p - \boldsymbol{K}_m\boldsymbol{A}_m)\boldsymbol{P}_{m-1} \tag{7}$$

where $\boldsymbol{I}_p$ is the $p \times p$ identity matrix. The SWLS estimate, denoted by $\hat{\boldsymbol{\lambda}}_{\mathrm{SWLS}}$, can be obtained after a fully M-block recursion, i.e., $\hat{\boldsymbol{\lambda}}_{\mathrm{SWLS}} = \hat{\boldsymbol{\lambda}}_M$.

[1]Note that the covariance here, as we can show, does not mean actually the covariance of image estimate. We use this notation in order to be consistent with some early literatures, e.g., [16].

The properties of SWLS are derived in appendix. It can be shown that the SWLS estimate is equivalent to the PWLS estimate when we start with an appropriate initial guess of covariance $\boldsymbol{P}_0$.

B. Simplified SWLS estimation

When using the SWLS method, there are two bottlenecks worth to mention: the calculation of $\boldsymbol{K}_m$ and the updating of covariance $\boldsymbol{P}_m$. $\boldsymbol{K}_m$ is called the gain in Kalman filtering theory. To compute this gain, we can rewrite it as:

$$\boldsymbol{K}_m(\boldsymbol{A}_m\boldsymbol{P}_{m-1}\boldsymbol{A}_m' + \boldsymbol{W}_m^{-1}) = \boldsymbol{P}_{m-1}\boldsymbol{A}_m'. \tag{8}$$

Because $(\boldsymbol{A}_m\boldsymbol{P}_{m-1}\boldsymbol{A}_m' + \boldsymbol{W}_m^{-1})$ is always positive definite, the above equation can be solved efficiently using Gaussian elimination method. Moreover, notice that $(\boldsymbol{A}_m\boldsymbol{P}_{m-1}\boldsymbol{A}_m' + \boldsymbol{W}_m^{-1})$ is a $\ell \times \ell$ matrix which depends on the size of the $\boldsymbol{Y}_m$ blocks. Thus the smaller the block size, the lower the computational cost will be.

However, $\boldsymbol{P}_m$ is a full matrix for which either to compute or to save has turned out to be the real bottleneck of the SWLS algorithm (see the review in [15]). For example, the Carlson's square-root filter suggested using a Cholesky factorized covariance $\boldsymbol{P}_m$ but the storage of one of Cholesky triangular matrices costs half the storage of $\boldsymbol{P}_m$, which for large dimensional problem is still expensive. For this reason, we prefer seeking a suboptimal SWLS algorithm.

The property 3 (see Appendix) shows that the SWLS estimate has a closed-form solution which depends on the final covariance $\boldsymbol{P}_M$ (here we suppose that the initial image is zero so that SWLS is equivalent to PWLS). The key feature of our simplified SWLS estimation is to find out an approximation $\tilde{\boldsymbol{P}}_M$ in place of $\boldsymbol{P}_M$. This is similar to a pseudo-inversion method. Examining the structure of $\boldsymbol{P}_M$ is critical to design the required approximation. To visualize this, Fig. 1 shows the various matrices for a toy PET problem (refer to [17]). We see that due to the typical tomography response, $\boldsymbol{F}$ is mainly concentrated about its diagonal. This property plays a central role. In fact, $\boldsymbol{P}_M$, nearly diagonal, is called the deblurring matrix. One of its kernels is shown in Fig. 1(c) using a 2D visualization with (d) and (e) the corresponding horizontal and vertical profiles. A quick look at such $\boldsymbol{P}_M$ points out that we can possibly use another $\tilde{\boldsymbol{P}}_M$ with a local compact support kernel in place of $\boldsymbol{P}_M$, or more specifically, we seek $\tilde{\boldsymbol{P}}_M$ that satisfies:

$$\tilde{\boldsymbol{P}}_M[j,k] \approx \boldsymbol{P}_M[j,k], \quad \mathrm{dist}(j,k) \leq r \tag{9}$$

and $\tilde{\boldsymbol{P}}_M[j,k] = 0$ otherwise. Here, $\tilde{\boldsymbol{P}}_M[j,k]$ denotes the jth row and kth column element of $\tilde{\boldsymbol{P}}_M$; $\mathrm{dist}(j,k)$, the Euclidean distance between the jth and the kth pixel; $r \geq 0$ is the radius of the local support. Notice that when r is small, the resulting $\tilde{\boldsymbol{P}}_m$ is a band matrix that requires relatively lower storage cost.

In properties 1 and 2, we have shown that the covariance is accumulated when propagated recursively. Our simplified approach, thus, attempts to keep the band structure of each $\boldsymbol{P}_m$ throughout the progress of covariance propagation. To compute the corresponding estimate, we adapt the same image update (5) as the SWLS algorithm

but we compute another suboptimal $\tilde{\boldsymbol{K}}_m$ using

$$\tilde{\boldsymbol{K}}_m = \tilde{\boldsymbol{P}}_{m-1}\boldsymbol{A}_m'(\boldsymbol{A}_m\tilde{\boldsymbol{P}}_{m-1}\boldsymbol{A}_i' + \boldsymbol{W}_m^{-1})^{-1} \tag{10}$$

in which $\tilde{\boldsymbol{P}}_m$ is given by

$$\tilde{\boldsymbol{P}}_m = \mathcal{B}^r\left((\boldsymbol{I}_p - \tilde{\boldsymbol{K}}_m\boldsymbol{A}_m)\tilde{\boldsymbol{P}}_{m-1}\right) \tag{11}$$

where the operator $\mathcal{B}^r(\boldsymbol{X})$ outputs a band matrix with the same size as $\boldsymbol{X}$ while preserving only $\boldsymbol{X}$'s elements that fulfills $\mathrm{dist}(j,k) \leq r, \forall j,k$.

Since $\tilde{\boldsymbol{P}}_{m-1}$ is banded, the left hand side of (11) also leads to

$$\begin{aligned}\mathcal{B}^r\left((\boldsymbol{I}_p - \tilde{\boldsymbol{K}}_m\boldsymbol{A}_m)\tilde{\boldsymbol{P}}_{m-1}\right) &= \tilde{\boldsymbol{P}}_{m-1} - \mathcal{B}^r\left(\tilde{\boldsymbol{K}}_m\boldsymbol{A}_m\tilde{\boldsymbol{P}}_{m-1}\right) \\ &= \tilde{\boldsymbol{P}}_{m-1} - \mathcal{B}^r\left(\tilde{\boldsymbol{P}}_{m-1}\boldsymbol{A}_m'(\boldsymbol{A}_m\tilde{\boldsymbol{P}}_{m-1}\boldsymbol{A}_m' + \boldsymbol{W}_m^{-1})^{-1}\boldsymbol{A}_m\tilde{\boldsymbol{P}}_{m-1}\right).\end{aligned} \tag{12}$$

In this case, the implementation of $\mathcal{B}^r$ provides an efficient algorithm without need of inversion. Taking $r = 0$ as an example, we can use the following solution:

1) Compute $\boldsymbol{v}_j = \boldsymbol{A}_m\tilde{\boldsymbol{P}}_{m-1}\boldsymbol{e}_j$ where $\boldsymbol{e}_j$ denotes the jth unit vector of length p;
2) Solve the equation $(\boldsymbol{A}_m\tilde{\boldsymbol{P}}_{m-1}\boldsymbol{A}_m' + \boldsymbol{W}_m^{-1})\boldsymbol{u}_j = \boldsymbol{v}_j$ using, e.g., the Gaussian elimination method;
3) Output the value of the jth diagonal entry: $\boldsymbol{v}_j'\boldsymbol{u}_j$.

A slight modification of the above steps can be also used for $r > 0$.

In this paper, a particular case of simplified SWLS method can be obtained by letting $M = N$ together with $r = 0$. In that case, $\boldsymbol{Y}_m$ is a scalar, $\boldsymbol{A}_m$ becomes a $1 \times p$ row vector while $\tilde{\boldsymbol{P}}_m$ reduces to a diagonal matrix. Then, it can be shown that

$$\hat{\boldsymbol{\lambda}}_m = \hat{\boldsymbol{\lambda}}_{m-1} + \frac{1}{(\boldsymbol{A}_m\tilde{\boldsymbol{P}}_{m-1}\boldsymbol{A}_m' + \sigma_m^2)}\tilde{\boldsymbol{P}}_{m-1}\boldsymbol{A}_m'(\boldsymbol{Y}_m - \boldsymbol{A}_m\hat{\boldsymbol{\lambda}}_{m-1}) \tag{13}$$

with

$$\tilde{\boldsymbol{P}}_m = \tilde{\boldsymbol{P}}_{m-1} - \frac{1}{(\boldsymbol{A}_m\tilde{\boldsymbol{P}}_{m-1}\boldsymbol{A}_m' + \sigma_m^2)}\mathcal{B}^0\{\tilde{\boldsymbol{P}}_{m-1}\boldsymbol{A}_m'\boldsymbol{A}_m\tilde{\boldsymbol{P}}_{m-1}\} \tag{14}$$

or in a simple elementwise notation

$$\tilde{\boldsymbol{P}}_m[j,j] = \tilde{\boldsymbol{P}}_{m-1}[j,j] - \frac{\tilde{\boldsymbol{P}}_{m-1}^2[j,j]\boldsymbol{A}_m^2[1,j]}{\sum_{l=1}^p \boldsymbol{A}_m^2[1,l]\tilde{\boldsymbol{P}}_{m-1}[l,l] + \sigma_m^2}, \quad j = 1,\ldots,p. \tag{15}$$

We see that the image updating (13) in the simplified approach is performed in terms of single line of response (i.e., $\boldsymbol{Y}_m$). This is similar to the well-known ART (algebraic reconstruction technique) iteration method. In fact, assuming that the image is an homogeneous random field and σ_m^2 tends to be zero, we can show that

$$\hat{\boldsymbol{\lambda}}_m = \hat{\boldsymbol{\lambda}}_{m-1} + \frac{1}{\boldsymbol{A}_m\boldsymbol{A}_m'}\boldsymbol{A}_m'(\boldsymbol{Y}_m - \boldsymbol{A}_m\hat{\boldsymbol{\lambda}}_{m-1}) \tag{16}$$

which is exactly the iteration formula of ART. Thus, from this point of view, the proposed SWLS method can be an extension of ART, which takes the measurement statistics into account.

The computational complexity of the simplified SWLS is much lower than the original one. As for the covariance update, it can be shown that a full iteration of SWLS costs at least $(\frac{3}{2}p^2 + \frac{5}{2}p + 1) \times N$ flops while only $(4p+1) \times N$ flops are used in the simplified approach.

III. Experiments

In this section, we evaluate the performance of our proposed SWLS methods. Since SWLS and PWLS are theoretically equivalent, our comparisons are mainly conducted between the SWLS algorithm and the simplified SWLS algorithm. Both simulated phantom data and clinical real PET raw data have been considered.

A. Materials

The phantom used in our experiments is a 32×32 emission image (shown in Fig. 2(a)) and similar to the one used in [17] , which has intensities of 2, 5, and 10 in the cold disk (right), background ellipse, and hot disk (left), respectively. The sinogram data have 32 projection angles equally covering 180° and 34 projection bins per angle with a sampling distance of one pixel. We used pseudo-random log-normal variants with a standard deviation of 0.3 (based on empirical fits to the logarithm of measured efficiency normalization factors) to simulate detector pairs with nonuniform detector efficiencies. The sinogram was then globally scaled into three kinds of mean sum of true events: 100K, 1M and 10M, which corresponded to the low, medium and high level of photon counts respectively. For each sinogram, we generated $L = 1000$ realizations of AC precorrected sinogram measurements using pseudo-random Poisson variants according to Fessler's formula (see (13) in [1]):

$$Y_i \sim c_i \left\{ \text{Poisson} \left\{ c_i^{-1} (\bar{Y}_i + a_i) \right\} - \text{Poisson} \left\{ c_i^{-1} a_i \right\} \right\} \tag{17}$$

where c_i is the ith detector efficiency normalization factor, a_i is the mean AC contribution to the ith detector pair which corresponds to 5% random coincidences, and $\bar{Y}_i$ is the mean of the ith detector pair measurement. Other factors such as attenuation were not simulated.

In the clinical study, PET raw data were measured by the multi-ring PET/CT scanner (CTI Siemens Biograph Sensation 16). The data acquisition started with the arrival of the radioactivity in the brain of a human volunteer and lasted for about 30 minutes. A total number of 47 image planes was generated. The sinograms (192×192) of the emission rate were stored for each plane. A normalization of detector efficiencies was performed using measurements obtained from a rotating line source prior to image reconstruction. Random coincidences were measured using the delayed time window method. The random coincidences measured in every detector pair were subtracted from the total number of coincidences detected by that pair. This resulted in the required precorrected data.

B. Reconstruction

We chose $\boldsymbol{R}$ being an identity matrix. The required PWLS estimate can then be called the minimum norm solution. Note that this penalization does not take the local smoothness information into account, thus the reconstruction can be less smoothing. An improved version using the same penalty to preserve local structures of reconstruction has been well investigated and validated for PET reconstruction in [18][19]. The matrix $\boldsymbol{W}$ was computed using Fessler's "data-plugin" technique [1]: $\sigma_i^2 = \max\{Y_i, 1\}$. The initial covariance $\boldsymbol{P}_0$ was given by $(\beta \boldsymbol{R})^{-1}$ and $\hat{\boldsymbol{\lambda}}_0$ was chosen to be a zero vector so that the SWLS estimate (20) is the PWLS estimate (4) according to the property

3. We computed the required SWLS estimate using the algorithm shown in subsection 2.1. where M was set to $N = 32 \times 32$. The same configuration was also used for the proposed sub-SWLS algorithm.

For clinical data, due to the high data dimension, three iterative methods rather than the SWLS method were used to compute the penalized WLS image estimate: OSL (one-step-late) EM [20], conjugate-gradient (CG) and iterative coordinate-descent (ICD) [10]. They were compared with the noniterative algorithm in computational complexity. Note that the original OSL-EM algorithm was developed only for the Poisson data. To make it suitable for the non-Poisson data, the following modified iterative formula based on the shifted-Poisson model can be used [21]:

$$\hat{\lambda}_j^{k+1} = \frac{\hat{\lambda}_j^k}{\sum_{i=1}^N \boldsymbol{A}[i,j] + \beta \hat{\lambda}_l^k} \sum_{i=1}^N \boldsymbol{A}[i,j] \frac{Y_i + 2a_i}{\sum_{l=1}^p \boldsymbol{A}[i,l]\hat{\lambda}_l^k + 2a_i}, \quad k = 1, 2, \ldots \tag{18}$$

where k denotes the iteration number. All iterative methods were applied with the same initial image.

C. Evaluation

To quantitatively evaluate the reconstruction performance of these methods, several criteria were used: root mean square error (RMSE) [13], bias-variance tradeoff, spatial resolution, and computational cost. Here, to evaluate the spatial resolution, three local impulses (see Fig. 2(b) where each impulse is defined by a value of 150) sampling the hot, cold and center background were also projected and added to the replicate data sets. Reconstruction of the impulses was obtained by subtracting the average reconstruction of the impulse-free data from the average reconstruction of the data containing the added impulses. A profile of the estimated impulse fitted with a standard Gaussian function. The computational cost measures the computational complexity in terms of the CPU time cost used by different methods through a complete reconstruction.

IV. Results

Fig. 3 shows the RMSE changes generated by varying parameter β within the range $[10^{-5}, 10^5]$. As we can see, for relatively large β, curves are almost identical. This shows that both SWLS's have very close global performance. When β decreases, SWLS can be better than the simplified one but it degenerates quickly. This strange behavior is apparently due to the ill-posed nature of penalized WLS estimate that is in charge of the reconstruction when the strength of penalty becomes weak. Many investigators have early observed this character and discussed on how to obtain the optimal β offering the global best RMSE performance. This, however, is beyond the scope of this paper. In contrast, the simplified SWLS seems more robust than SWLS since it shows less sensitivity to the choice of β. In view of the global minimum RMSE, the simplified approach is close to SWLS when handling either the low- or medium-count PET raw data, while it seems likely suboptimal for high-count data as compared with the SWLS estimate.

Fig. 4 shows an example of reconstruction. Here the medium-count data set was used and β was set to 0.3. Notice that, in this case, SWLS is superior to the simplified SWLS in global RMSE performance. The results show no difference between two these SWLS's at least from the visual standpoint. The mean images are also close to each

other. A slight difference can be found out in the comparison of the central horizontal profiles of the mean images. However, two standard deviation images are apparently distinguished. This reveals that the simplified approach can produce even lower Std. Dev than SWLS.

Fig. 5 demonstrates the bias-variance tradeoff curves generated by SWLS and the simple version for three different noisy projection data sets. The first column in Fig. 5 shows the tradeoff for the hot region and the second column corresponds to the cold region. We see now that two SWLS's provide a similar tradeoff for region with high activity. The performance of the simplified SWLS is slightly better than SWLS since, for a given bias, the simplified SWLS offers relatively lower standard deviation than the original SWLS method (though there are some exceptions, this holds for a wide range of bias). For the cold region, there is a good agreement between the two tradeoff curves. This suggests that SWLS and the simplified SWLS are equivalent to a large extent, and that our suboptimal strategy for the covariance computation is likely an adequate approximation.

Fig. 6 shows central horizontal profiles of the local impulses estimated using different β values. Here, we only considered the medium activity data set since the spatial resolution property is usually characterized by the system (e.g., the strength of the employed smoothing penalty) instead of measurement. Again, the agreement between SWLS's is shown by a good overlap of the symbols. This reflects that two SWLS's are comparable in spatial resolution. Another worthy mention is that none of these methods can provide uniform resolution. Typically, regions with high activity have relatively lower resolution. This can be seen by comparing the magnitude of the local impulse profiles in Fig.6. Such resolution nonuniformity has been early noticed by others, and is usually caused by the specific nonstationary noise variances of the Poisson data. A modification to parameter β accounting for this noise character has proven useful to achieve relatively uniform resolution (see [17]). This however is not in the scope of this paper.

Fig. 7 shows the PET reconstructions using different methods with brain data. In this particular experiment, what we want to highlight are the quantitative performances as displayed in Tab. I. The number of iterations for OSL-EM and CG (20 iterations) has been defined in such a way that their computation time is close to that of the simplified SWLS. We used for ICD (which is converging faster but has higher complexity) only 5 iterations. Fig. 8 compares the final reconstructions from another real data example. In this case, we setup a stopping criterion: $\|\hat{\boldsymbol{\lambda}}^{k+1} - \hat{\boldsymbol{\lambda}}^{k}\|/\|\hat{\boldsymbol{\lambda}}^{k}\| < 10^{-10}$ for all iterative algorithms to ensure the effective convergence. As we can see, it costs about more than one hundred iterations for OSL-EM and CG to reach the stopping criterion while for ICD it requires 22 iterations. The reconstructions seem quite similar to each other at least from the visual point of view. Here we chose $\beta = 10.0$ so that all image reconstructions have relatively lower resolution (since the larger the β is, the smoother the image reconstruction would be). This indicates that our proposed simplified SWLS might have the same behavior as that of OSL-EM, CG and ICD. This comparison also shows the effectiveness of simplified SWLS and can be safely used as other methods for image reconstruction. In this paper, all codes were written in C++ on a personal computer with a 3.06GHz CPU and 512MBytes memory storage.

V. Discussion and Conclusion

We have considered a PWLS estimate for PET systems that precorrects for AC events. We proposed the SWLS algorithm for minimizing that objective, and demonstrated that SWLS is indeed a PWLS estimation under properly selected initial estimate and *a priori* covariance. Although this algorithm is noniterative, its computational costs can be expensive. Instead of using alternative implementations, we sought a suboptimal solution in place of such SWLS estimate. This has been done by first exploring the structure of the covariance and then replacing it with a band approximation. The aim was to maintain the characteristics of the covariance while reducing burdens for both saving and computation. We have discussed one particular SWLS algorithm which can be viewed as an extension of ART. Quantitative comparisons to the original SWLS on a simulated phantom with hot and cold regions demonstrated that the performance of the simplified SWLS is comparable to SWLS. Although this SWLS algorithm may be suboptimal in preserving the global RMSE performance, the method appears to have some quantitative and qualitative advantages over SWLS. The required computation time compared with iterative algorithms in a clinical data study is nearing the realm of being practical for routine use.

In this work, the simplest form of the penalty function has been studied. Its advantage is that it leads to an analytical solution. Other non-quadratic penalty functions are of interest for SWLS and the simplified version and will be further studied. Our ongoing study is also aimed at extending the simplified approach to 3D PET reconstruction, where faster algorithm would be more desirable.

Appendix

Four properties of SWLS are depicted as follows.

Property 1: The mth covariance $\boldsymbol{P}_m = (\boldsymbol{P}_0^{-1} + \sum_{j=1}^{m} \boldsymbol{A}_j' \boldsymbol{W}_j \boldsymbol{A}_j)^{-1}$.

Proof: To show this, we notice that:

$$\begin{aligned} \boldsymbol{P}_m &= \boldsymbol{P}_{m-1} - \boldsymbol{P}_{m-1}\boldsymbol{A}_m'(\boldsymbol{A}_m\boldsymbol{P}_{m-1}\boldsymbol{A}_m' + \boldsymbol{W}_m^{-1})^{-1}\boldsymbol{A}_m\boldsymbol{P}_{m-1} \\ &= (\boldsymbol{P}_{m-1}^{-1} + \boldsymbol{A}_m'\boldsymbol{W}_m\boldsymbol{A}_m)^{-1}. \end{aligned} \tag{19}$$

The second equality uses the Sherman-Morrison-Woodbury formula. Then, the property is derived by using (19) recursively. ■

Note that $\boldsymbol{A}_m'\boldsymbol{W}_m\boldsymbol{A}_m$ can be seen as the information matrix of estimating $\boldsymbol{\lambda}$ from the mth measurement block $\boldsymbol{Y}_m$. This property means that the information is cumulated during the period of recursion.

A direct conclusion based on this property is that

Property 2: The Mth covariance $\boldsymbol{P}_M = (\boldsymbol{F} + \boldsymbol{P}_0^{-1})^{-1}$.

Proof: It is easy to show that $\sum_{j=1}^{M} \boldsymbol{A}_j'\boldsymbol{W}\boldsymbol{A}_j = \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} = \boldsymbol{F}$. ■

Property 3: The SWLS estimate $\hat{\boldsymbol{\lambda}}_{\mathrm{SWLS}}$ has the closed-form expression:

$$\hat{\boldsymbol{\lambda}}_{\mathrm{SWLS}} = \boldsymbol{P}_M\boldsymbol{P}_0^{-1}\hat{\boldsymbol{\lambda}}_0 + \boldsymbol{P}_M\boldsymbol{A}'\boldsymbol{W}\boldsymbol{Y}. \tag{20}$$

Proof: From (5), we have

$$\hat{\boldsymbol{\lambda}}_m = (\boldsymbol{I}_p - \boldsymbol{K}_m \boldsymbol{A}_m)\hat{\boldsymbol{\lambda}}_m + \boldsymbol{K}_m \boldsymbol{Y}_m. \tag{21}$$

With some algebra, it can be shown that

$$\boldsymbol{K}_m = (\boldsymbol{I}_p - \boldsymbol{K}_m \boldsymbol{A}_m)\boldsymbol{P}_{m-1}\boldsymbol{A}_m'\boldsymbol{W}_m. \tag{22}$$

Substituting it into (21), we arrive at

$$\begin{aligned}
\hat{\boldsymbol{\lambda}}_m &= (\boldsymbol{I}_p - \boldsymbol{K}_m \boldsymbol{A}_m)\hat{\boldsymbol{\lambda}}_{m-1} + (\boldsymbol{I}_p - \boldsymbol{K}_m \boldsymbol{A}_m)\boldsymbol{P}_{m-1}\boldsymbol{A}_m'\boldsymbol{W}_m\boldsymbol{Y}_m \\
&= (\boldsymbol{I}_p - \boldsymbol{K}_m \boldsymbol{A}_m)(\hat{\boldsymbol{\lambda}}_{m-1} + \boldsymbol{P}_{m-1}\boldsymbol{A}_m'\boldsymbol{W}_m\boldsymbol{Y}_m) \\
&= (\boldsymbol{I}_p - \boldsymbol{K}_m \boldsymbol{A}_m)\boldsymbol{P}_{m-1}(\boldsymbol{P}_{m-1}^{-1}\hat{\boldsymbol{\lambda}}_{m-1} + \boldsymbol{A}_m'\boldsymbol{W}_m\boldsymbol{Y}_m) \\
&= \boldsymbol{P}_m(\boldsymbol{P}_{m-1}^{-1}\hat{\boldsymbol{\lambda}}_{m-1} + \boldsymbol{A}_m'\boldsymbol{W}_m\boldsymbol{Y}_m) \\
&= \boldsymbol{P}_m\boldsymbol{P}_{m-1}^{-1}\hat{\boldsymbol{\lambda}}_{m-1} + \boldsymbol{P}_m\boldsymbol{A}_m'\boldsymbol{W}_m\boldsymbol{Y}_m.
\end{aligned} \tag{23}$$

By using (23) recursively, we can deduce:

$$\hat{\boldsymbol{\lambda}}_{\mathrm{SWLS}} = \hat{\boldsymbol{\lambda}}_M = \boldsymbol{P}_M\boldsymbol{P}_0^{-1}\hat{\boldsymbol{\lambda}}_0 + \boldsymbol{P}_M\boldsymbol{A}'\boldsymbol{W}\boldsymbol{Y}. \tag{24}$$

■

Now we are able to explore the relations between $\hat{\boldsymbol{\lambda}}_{\mathrm{SWLS}}$ and $\hat{\boldsymbol{\lambda}}_{\mathrm{PWLS}}$.

Property 4: $\hat{\boldsymbol{\lambda}}_{\mathrm{SWLS}}$ is equal to $\hat{\boldsymbol{\lambda}}_{\mathrm{PWLS}}$ if $\hat{\boldsymbol{\lambda}}_0 = \boldsymbol{0}$ and $\boldsymbol{P}_0^{-1} = \beta\boldsymbol{R}$ where $\boldsymbol{0}$ denotes the zero vector.

Proof: This is easy to obtained by comparing the PWLS estimate (4) with the SWLS estimate (20). ■

REFERENCES

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

[2] J. A. Fessler, and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Trans. Image. Process.,* vol. 4, pp. 1417-1429, 1995.

[3] H. Q. Zhu, J. Zhou, H. Z. Shu, et al. "A novel weighted least squares PET image reconstruction method using adaptive variable index sets," *Digital Signal Processing,* vol. 16, pp. 106–119, 2006.

[4] N. H. Clinthorne, "Constrained least-squares vs. maximum likelihood reconstructions for positron data. in *Proc. of 1992 IEEE Nucl. Sci. Symp. and Med. Imag. Conf.* 1992, pp. 1237–1239.

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

[6] E. Mumcuogu, R. Leahy, S. R. Cherry, et al. "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images," *IEEE Trans. Med. Imag.,* vol. 13, pp. 687–701, 1994.

[7] J. Anderson, B. A. Mair, M. Rao, et al. "Weighted least-squares reconstruction methods for positron emission tomography," *IEEE Trans. Med. Imag.,* vol. 16 pp. 159–165, 1997.

[8] J. A. Fessler, and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Image. Process.,* vol. 8 pp. 688–99, 1999.

[9] K. Sauer, and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Sig. Proc.,* vol. 41, pp. 534–548, 1993.

[10] C. Bouman, and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. Image. Process.,* vol. 5, pp. 480–492, 1996.

[11] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, et al. "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction," *IEEE Trans. Med. Imag.,* vol. 16, pp. 166–175, 1997.

[12] J. Llacer, and E. Veklerov, "Stopping rule for the MLE algorithm based on statistical hypothesis testing," *IEEE Trans. Med. Imag.,* vol. 6, pp. 313–319, 1987.

[13] J. Zhou, and L. M. Luo, "Sequential weighted least squares algorithm for PET image reconstruction," *Digital Signal Processing,* vol. 16, pp. 735–745, 2006.

[14] H. M. Hudson, and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Tran. Med. Imag.,* vol. 13, pp. 601–609, 1994.

[15] M. S. Grewal, and A. P. Andrews, Kalman Filtering: Theory and Practice Using MATLAB, *2nd edition,* Wiley-Interscience Publication: John Wiley & Sons, Inc., NY, 2001.

[16] S. M. Kay, Fundamentals of Statistical Signal Processing. Volume I: Estimation Theory, Prentice Hall, Inc. 1993.

[17] J. A. Fessler, and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs," *IEEE Trans. Image. Process.,* vol. 5, no. 9, pp. 1346–1358, 1996.

[18] X. Ouyang, W. H. Wong, V. E. Johnson, et al. "Incorporation of correlated structural images in PET image reconstruction," *IEEE Trans. Med. Imag.,* vol. 13, pp. 627–640, 1994.

[19] H. S. Lu, C. M. Chen, I. H. Yang, "Cross-reference weighted least square estimates for positron emission tomography," *IEEE Trans. Med. Imag.,* vol. 17, pp.1–8, 1998.

[20] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.,* vol. 9, pp. 84–93, 1990.

[21] M. Yavuz, and J. A. Fessler, "Penalized-likelihookd estimators and noise analysis for randoms-precorrected PET transmission scans," *IEEE Trans. Med. Imag.,* vol. 18, pp. 665–674, 1999.

TABLE I

COMPARISONS OF COMPUTATIONAL COST (UNIT IN SECOND). NUMBER IN BRACKET INDICATES THE ITERATION NUMBER.

	32×32	64×64	128×128
OSL-EM (20)	0.69	2.16	8.35
CG (20)	0.94	3.03	9.72
ICD (5)	9.28	29.21	92.83
Simplified SWLS	0.52	2.72	9.41

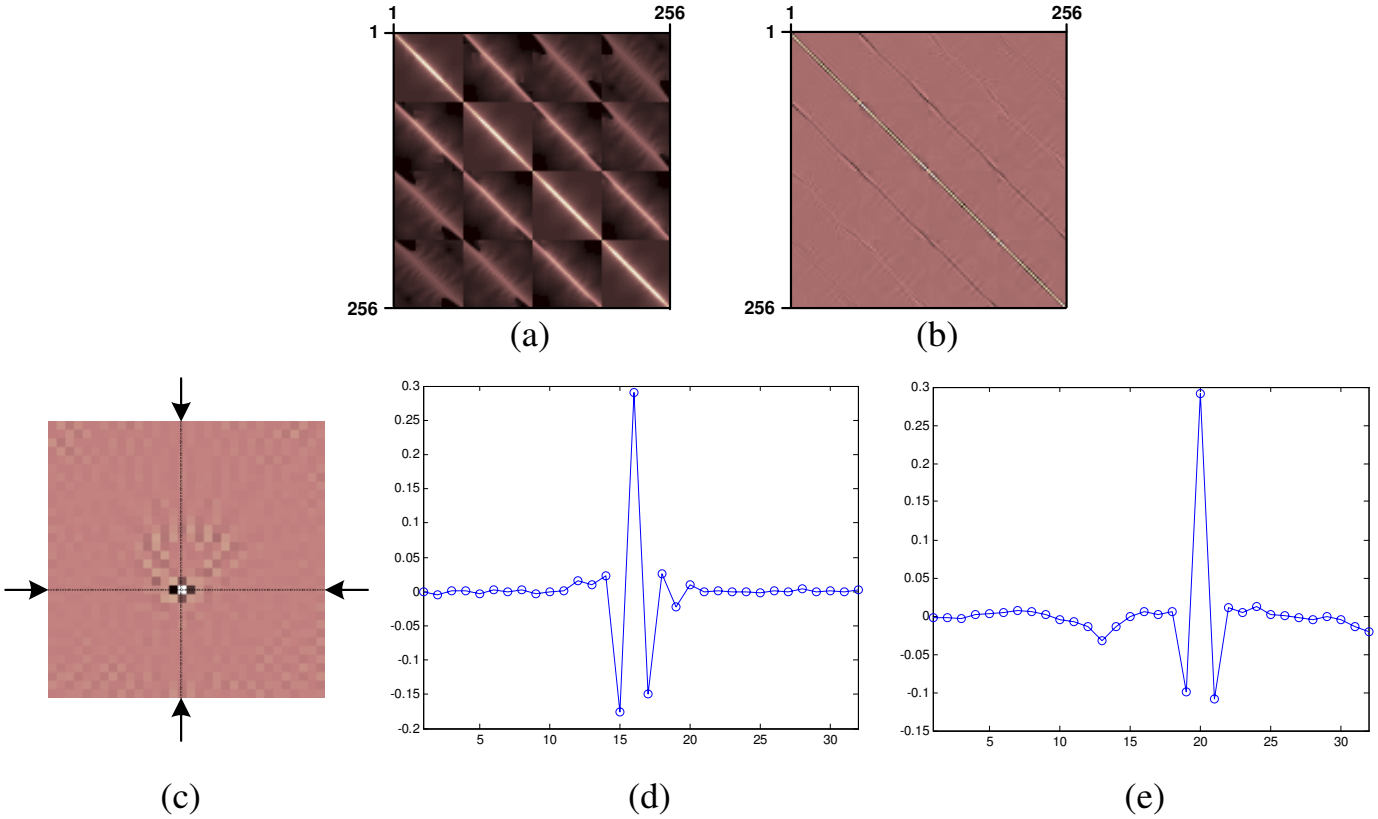
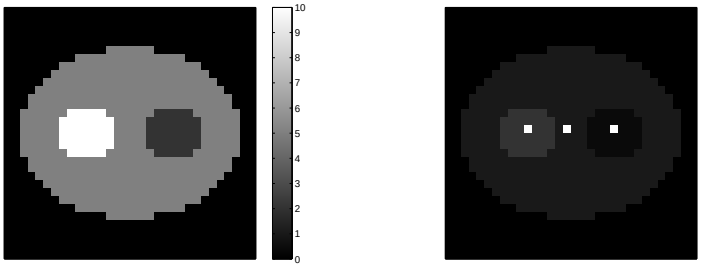
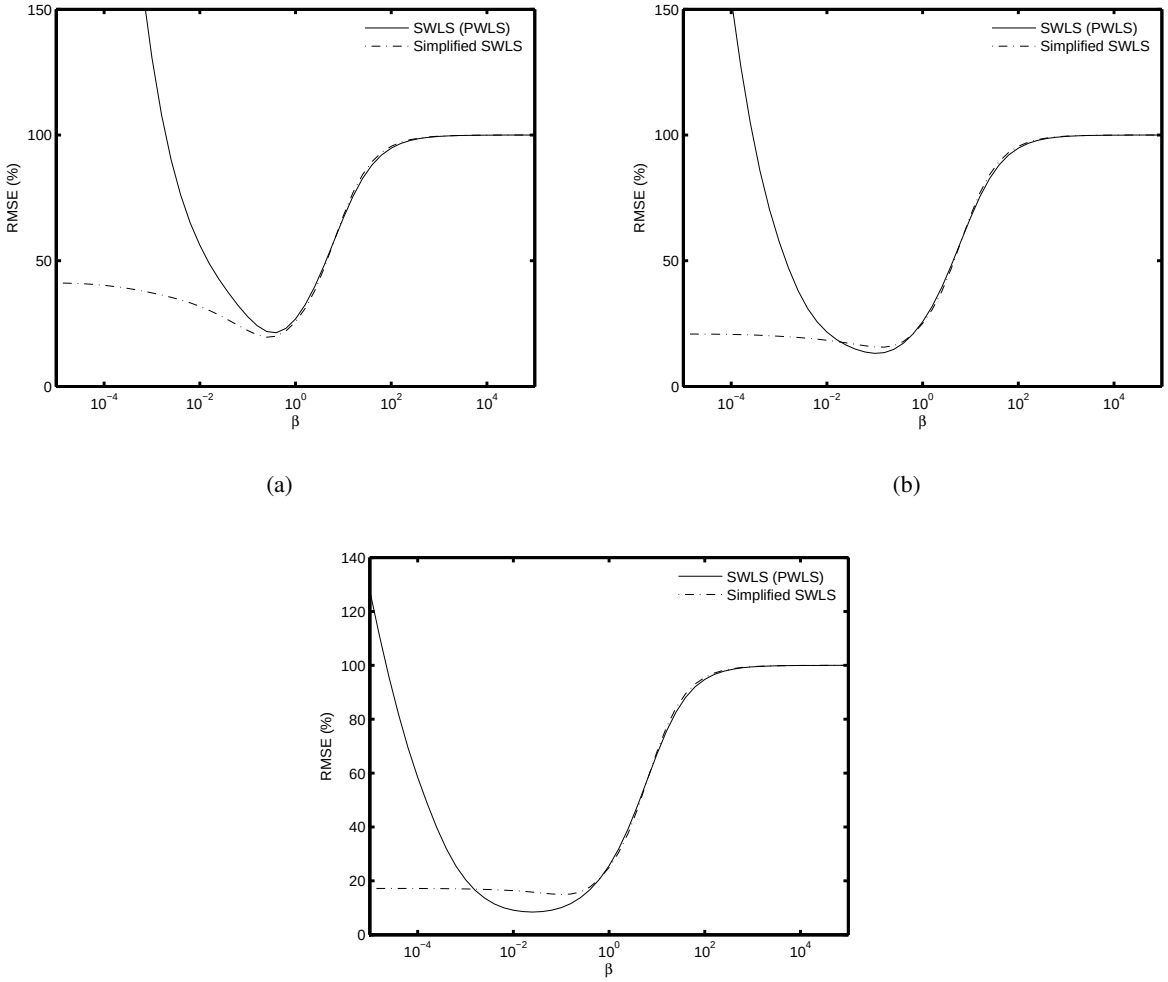


Fig. 1. A toy PET problem in which the system matrix $\boldsymbol{A}$ was computed by sampling a 32×32 image with 32 angles (uniformly covering $180°$) and 34 projections per angle. Both $\boldsymbol{R}$ and $\boldsymbol{W}$ were set to be identity matrices and $\beta = 1.0$. (a) the Fisher information matrix $\boldsymbol{F}$; (b) the matrix $\boldsymbol{P}_M$ as well as the inversion of $\boldsymbol{F} + \beta\boldsymbol{R}$; In order to visualize, the parts of matrix (the upper-left corner with size 256×256) are shown. (c) a kernel (32×32) extracted from the 500th row of $\boldsymbol{P}_M$; (c) and (d) the corresponding horizontal and vertical profiles across the kernel.



(a) (b)

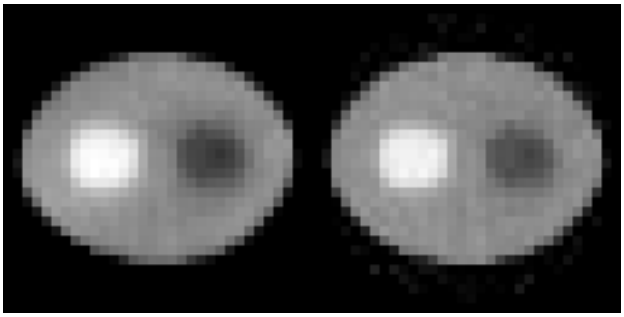
Fig. 2. (a) The computer simulated phantom. (b) The phantom with the three impulses. The bright dots indicate their positions.



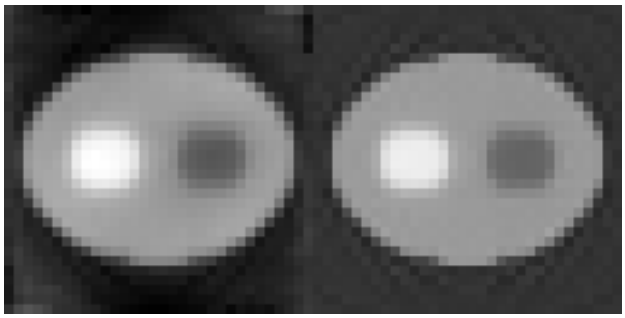
(a) (b)

(c)

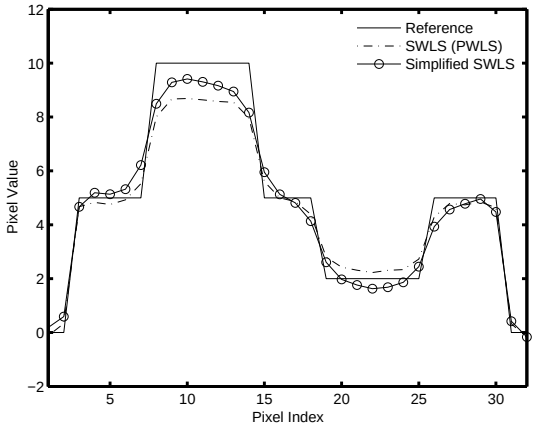
Fig. 3. RMSE changes as a function of β. (a), (b) and (c) are results corresponding to the lower, medium and high counts raw data, respectively.



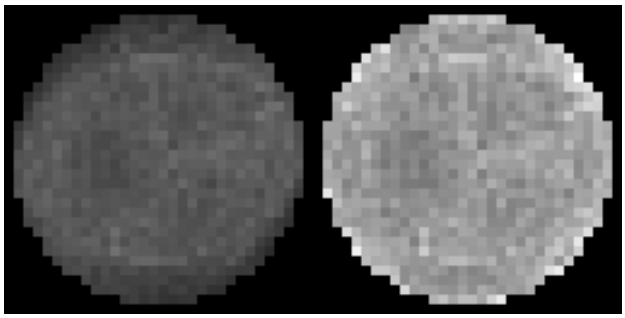
(a)



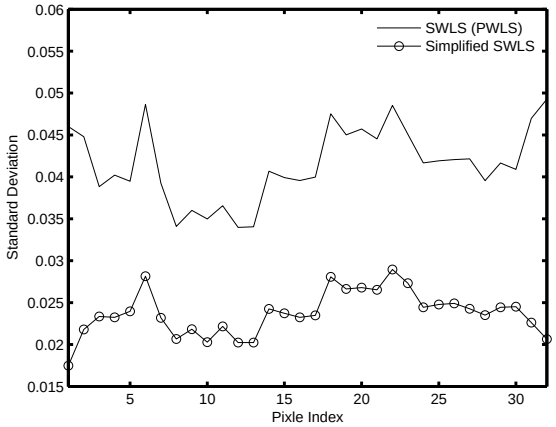
(b)



(c)



(d)



(e)

Fig. 4. (a) Reconstructions for medium-count data. left: the simplified SWLS; right: SWLS. (b) The mean images. left: the simplified SWLS; right: SWLS. (c) The central horizontal profile of the mean images. (d). The standard deviation images. left: the simplified SWLS, and right: SWLS. (e) The central horizontal profile of the standard deviation images. Note that images have been scaled to the same gray-level for a good comparison.

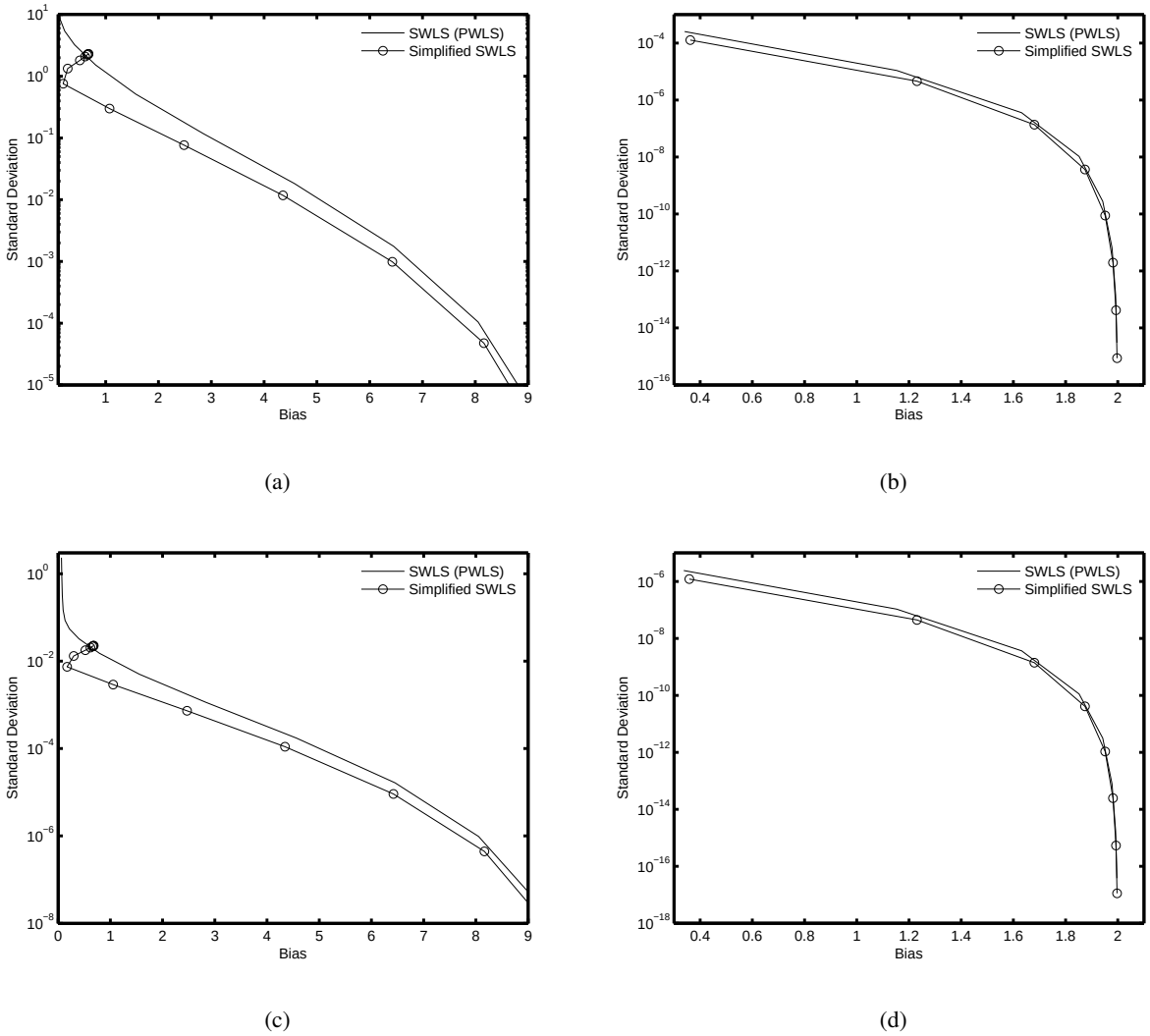


Fig. 5. Bias-Std. Dev. tradeoff curves for hot region (the first column) and the cold region (the second column). From top to bottom: results using low, high activity noisy measurement data.

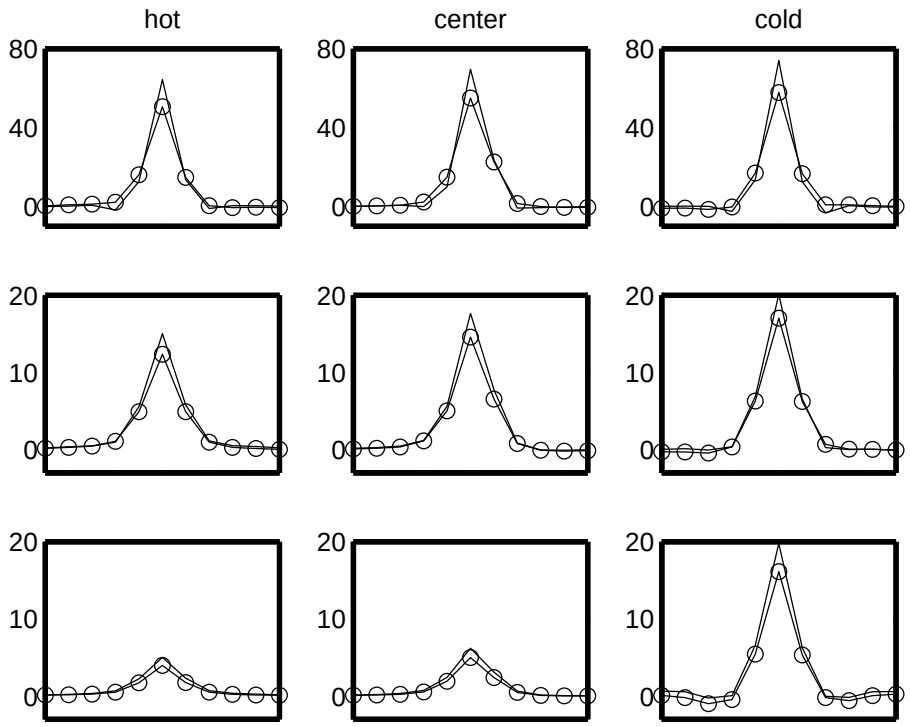


Fig. 6. Local impulse profiles estimated with different β. From top to bottom: results using $\beta = 0.05$, 0.5 and 1.5, respectively. ———: SWLS, —∘—: Simplified SWLS.

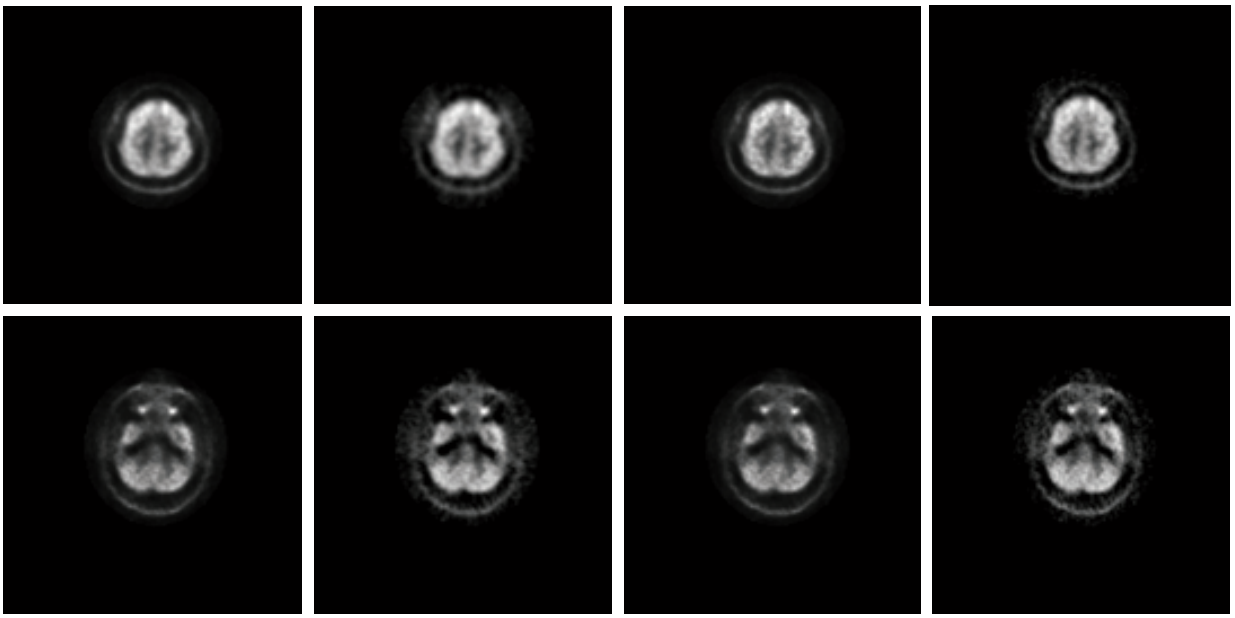
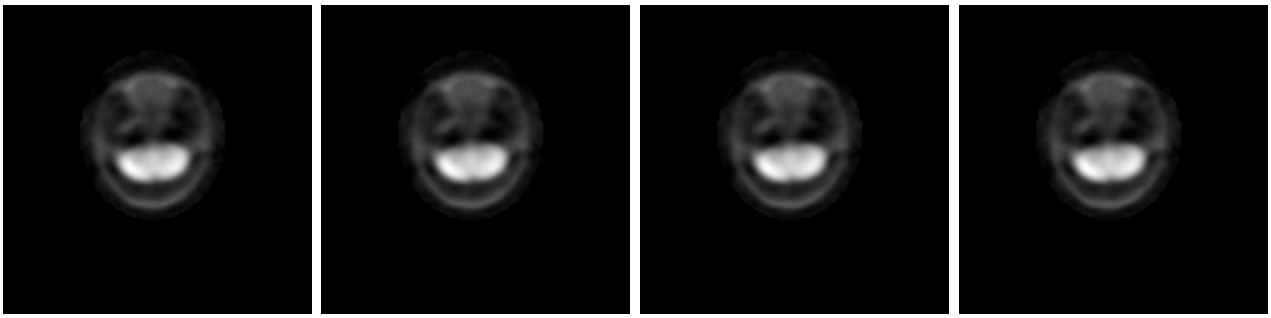


Fig. 7. Image reconstruction from clinical brain PET raw data using: (from left to right) OSL-EM (#20 ite.), CG (#20 ite.), ICD (#5 ite.) and the simplified SWLS (all algorithm use $\beta = 0.05$).



(a) OSL-EM (#169 ite.) (b) CG (#144 ite.) (c) ICD (#22 ite.) (d) Simplified SWLS

Fig. 8. Image reconstruction from clinical brain PET raw data using different algorithms with $\beta = 10.0$.