ISWLS: Novel Algorithm for Image Reconstruction in PET

E. Karali, S. Pavlopoulos, S. Lambropoulou, and D. Koutsouris

Abstract—The purpose of this study is to introduce a novel empirical iterative algorithm for medical image reconstruction, under the short name ISWLS (image space weighted least squares), which is expected to have image space reconstruction algorithm (ISRA) properties in noise manipulation and weighted least-squares (WLS) acceleration of the reconstruction process. We used phantom data from a prototype small-animal positron emission tomography system and the methods presented here are applied to 2-D sinograms. Further, we assess the performance of the new algorithm by comparing it to the simultaneous version of algebraic reconstruction technique (ART), simultaneous algebraic reconstruction technique (SART), to expectation maximization maximum likelihood (EM-ML), ISRA, and WLS. All algorithms are compared in terms of cross-correlation coefficient, reconstruction time, and contrast-to-noise ratios (CNRs). As it turns out, ISWLS presents higher CNRs than EM-ML, ISRA, and SART for objects of different sizes. Also, ISWLS shows similar performance to WLS during the first iterations but it has better noise manipulation. Finally, ordered subsets ISWLS (OS-ISWLS), the OS version of ISWLS, shows its best performance between the first six-nine iterations. Its behavior seems to be a compromise between OS-ISRA and OS-WLS.

Index Terms—Image reconstruction, positron emission tomography (PET), small-animal imaging.

I. INTRODUCTION

I mage reconstruction in positron emission tomography (PET) uses the collected projection data of the object/patient under examination. Until recently, image reconstruction in commercial clinical PET systems has been performed with analytical reconstruction algorithms based on the filtered backprojection method (FBP). FBP is based on direct computation of the inverse Radon transformation that combines collected data with image pixels values and offers an immediate mathematical solution for image formation. Moreover, FBP uses the Fourier theorem or central section theorem, which connects 1-D Fourier transformation of projection data to 2-D image Fourier transformation, to avoid image blurring. Analytical techniques present high computational speed at low computational cost and in general do not require expensive, powerful computing systems. One of the major drawbacks of FBP methods is their lack to incorporate in the reconstruction process many of the different factors related to γ-ray production and detection, as well as other factors affecting image quality (e.g., system geometry, object and septa scatter, detector characteristics, positron range, photons non-collinearity, randoms, and photon attenuation). Moreover, FBP does not preserve image nonnegativity while the effect of missing data tends to produce streak artifacts in the reconstructed images [1], [2].

Iterative image reconstruction algorithms have been proposed as an alternative to conventional analytical methods. Despite their computational complexity, they become more and more popular, mostly because they can produce images with better contrast-to-noise (CNR) and signal-to-noise ratios at a given spatial resolution, compared to FBP techniques. Iterative methods are able to incorporate a model of all the physical phenomena during the acquisition process, including scanner characteristics. Based on predetermined criteria and after a series of successful iterations, they attempt to find the best approach to the true image of radioactivity spatial distribution. Yet, the high computational cost and the lack of an efficient termination criterion have prevented, in the previous years, their application in commercial systems [1]–[3].

Iterative techniques are divided into two main categories: algebraic and statistical. Algebraic techniques were presented as an alternative to FBP and were used by Hounsfield to reconstruct images from data collected from the first-generation current transformer systems. Algebraic methods rely on the assumption that projection data are connected linearly to image pixels. So, they try to solve iteratively a linear system of $M$ equations and $N$ unknowns, where $M$ is the total number of detector tubes and $N$ is the total number of image pixels. Simultaneous algebraic technique (SART) is an algebraic method where all image pixels are updated simultaneously using all projection data [4]. Statistical techniques are classified to maximum likelihood algorithms and least-squares methods. Maximum likelihood algorithms are very popular in the field of image reconstruction. These methods assume that collected data are independent Poisson variables, and in every iteration step they try to maximize the log-likelihood. The most famous maximum likelihood technique is the expectation maximization maximum likelihood (EM-ML) algorithm, which was first presented by Shepp and Vardi [5]. Least-squares methods suppose that noise in the reconstruction model is normally distributed with mean value zero and the same standard deviation $\sigma$, namely it has the characteristics of white noise. Image space reconstruction algorithm (ISRA) [6] is a least-squares reconstruction method introduced by Daube-Witherspoon and Muehllehner. ISRA guarantees positive approximation of the
true radiopharmaceutical distribution inside the object of interest if the initial solution is also positive. Moreover, it shows better noise manipulation than EM-ML. Another least-squares algorithm is the weighted least-squares technique (WLS), due to Anderson et al. [7]. WLS assumes that random independent noise variables present different standard deviations. The matrix of these standard deviations consists of the expected projection data. WLS accelerates the reconstruction process and results in reconstructed images of better spatial resolution.

Depending on the way the initial solution is updated in every iteration, different algorithmic schemes can be used. In an attempt to speed up known iterative techniques, ordered subsets (OS) methods were introduced. OS algorithms update simultaneously image pixels using a subset of the collected data in every iteration, without deteriorating reconstructed image quality [8]. The OS version of ISRA is OS-ISRA, of WLS is OS-WLS, and of EM-ML is ordered subsets expectation maximization (OSEM). Furthermore, OSEM, due to Hundson and Larkin, consists of the most popular reconstruction method [9].

The purpose of this study is on the one hand to introduce a new empirical image reconstruction algorithm, under the short name ISWLS (image space weighted least squares) [10], produced by the maximization of an objective function. To maximize the objective function, the Kuhn–Tucker condition must be satisfied. ISWLS is expected to have ISRA properties in noise manipulation and WLS acceleration of the reconstruction process. To assess the performance of the new iterative reconstruction method, we have used phantom data produced from simulating a prototype small-animal PET system. We compared reconstruction data with those from the simultaneous versions of known algorithms, OSEM, OS-ISRA, and OS-WLS. The methods presented here are applied to 2-D sinograms.

We note that simultaneous versions of known algorithms, that is, algorithms where all image pixels are simultaneously updated in every iteration are of great interest because of their ability to be implemented in parallel computing architectures, which decreases drastically the total reconstruction time [8].

II. THEORY

In general, every iterative method relies on the hypothesis that the projection data $y$ are linearly connected to the image $x$ of radiopharmaceutical spatial distribution, according to the equation

$$y = A^T x$$

where $A$ is a matrix that characterizes the PET system being used for data acquisition. In bibliography, this matrix is called the system or probability matrix and it projects image data to the sinogram domain (the term sinogram refers to the projection data matrix) [11]. Every element $a_{ij}$ of the system matrix $A$ represents the probability an annihilation event emitted in image pixel $i$ to be detected in line of response (LORi). The significance of the probability matrix lies on the valuable information related to the data acquisition process; that is it can contain, e.g., number of detector rings, number of detector elements in every ring, ring diameter, diameter of transaxial field of view, detector size, image size, and spatial and angular sampling.

The most commonly used algebraic simultaneous iterative method is SART, with updating scheme in the $k$th iteration:

$$x_i^k = x_i^{k-1} + \frac{\lambda_k}{M} \sum_{j=1}^{M} a_{ij} y_j - \frac{1}{N} \sum_{i'=1}^{N} a_{i'j} x_{i'}^{k-1}.$$  \hspace{1cm} (2)

The relaxation parameter in (2) $\lambda_k$ lies in (0,1].

On the other hand, the most commonly used least-squares algorithms that are based on simultaneous iterative schemes are ISRA and WLS with updating step in the $k$th iteration:

$$\text{ISRA} : x_i^k = x_i^{k-1} - \frac{1}{N} \sum_{j=1}^{N} a_{ij} y_j x_i^{k-1}.$$  \hspace{1cm} (3)

$$\text{WLS} : x_i^k = x_i^{k-1} - \frac{1}{N} \sum_{j=1}^{N} \left( \sum_{i'=1}^{N} a_{i'j} x_{i'}^{k-1} \right)^{-1} a_{ij} y_j.$$  \hspace{1cm} (4)

The EM-ML algorithm has an updating step in the $k$th iteration:

$$\text{EM - ML} : x_i^k = x_i^{k-1} - \frac{1}{N} \sum_{j=1}^{N} \left( \sum_{i'=1}^{N} a_{i'j} x_{i'}^{k-1} \right)^{-1} a_{ij} y_j.$$  \hspace{1cm} (5)

A. ISWLS Algorithm

In this study, we propose a new algorithm under the short name ISWLS. Consider an image discretized into $N$ pixels and the measured data $y$ collected by $M$ detector tubes. We propose the following ISWLS estimator of $x$ in (1):

$$\hat{x} = \arg \min_{x} \phi(x) \text{ subject to } x_i \geq 0, \ i = 1, 2, ..., N$$  \hspace{1cm} (6)

where

$$\phi(x) = \sum_{j=1}^{M} \left[ \frac{1}{3} \left( y_j - \sum_{i=1}^{N} a_{ij} x_i \right)^3 + \left( y_j - \frac{2}{3} \sum_{i=1}^{N} a_{ij} x_i \right) \left( \sum_{i=1}^{N} a_{ij} x_i \right)^2 \right].$$  \hspace{1cm} (7)

Under the conditions in problem (6), $\hat{x}$ is a solution if and only if the Kuhn–Tucker condition is satisfied, namely

$$x_i \frac{\partial \phi(x)}{\partial x_i} |_{\hat{x}} = 0$$  \hspace{1cm} (8)

where

$$\frac{\partial \phi(x)}{\partial x_i} = \sum_{j=1}^{M} \left[ \left( y_j - \sum_{i'=1}^{N} a_{i'j} x_i \right)^2 a_{ij} + 2a_{ij} y_j \left( \sum_{i'=1}^{N} a_{i'j} x_i \right) \right. - \left. 2a_{ij} \left( \sum_{i'=1}^{N} a_{i'j} x_i \right)^2 \right].$$  \hspace{1cm} (9)
According to (9), (8) is written as

$$x_i \sum_{j=1}^{M} \left( a_{ij} y_j^2 - a_{ij} \left( \sum_{i'=1}^{N} a_{i'j} x_{i'} \right)^2 \right) = 0. \quad (10)$$

So, we obtain the fixed point iterative formula for the $i$th pixel's update as follows:

**ISWLS:**

$$x_i^k = x_i^{k-1} \frac{\sum_{j=1}^{M} a_{ij} y_j^2}{\sum_{j=1}^{M} a_{ij} \left( \sum_{i'=1}^{N} a_{i'j} x_{i'}^{k-1} \right)^2}. \quad (11)$$

Moreover, we can derive the OS version of ISWLS, OS-ISWLS, with updating scheme in the $t$th iteration for subset $S_n$

**OS-ISWLS:**

$$x_i^k = x_i^{k-1} \frac{\sum_{j \in S_n} a_{ij} y_j^2}{\sum_{j \in S_n} a_{ij} \left( \sum_{i'=1}^{N} a_{i'j} x_{i'}^{k-1} \right)^2}. \quad (12)$$

### III. RESULTS

#### A. Comparative Evaluation of Simple Algorithms

For the evaluation of the iterative reconstruction methods presented in Section II, projection data of a Derenzo-type phantom have been used. The Derenzo-type phantom consists of sets of rods filled with F$^{18}$, with diameters 4.8, 4, 3.2, 2.4, 1.6, and 1.2 mm, and the same separation between surfaces in the corresponding sets. The rods were surrounded by plastic (polyethylene). Data were produced using Monte Carlo simulation of a small-animal PET scanner.

Further, $18 \times 10^6$ coincidence events were collected. Projection data were binned to a 2-D sinogram, 55 pixels $\times$ 170 pixels in size, which means that data from 55 tubes of response (TORs) per rotation angle were collected and 170 totally angular samples were used. Since the two detector heads rotate from $0^\circ$ to $180^\circ$, the angular step size was 1.0647$^\circ$.

The system matrix was derived from an analytical method and calculated once before reconstruction. Each element $a_{ij}$ was computed as the area of intersection $E_{ij}$ of TOR$^j$ with image pixel $i$, according to the equation

$$a_{ij} = \frac{E_{ij}}{\sum_{j=1}^{M} E_{ij}} \quad (13)$$

where $M$ is the number of sinogram elements ($M = 55 \times 170$). The calculated system matrix is a sparse matrix. It consists of zero-valued elements in majority that have no contribution during iterative reconstruction process. So, only the nonzero elements were stored, resulting in significant reduction in system matrix size and consequently in required storage. The reconstructed 2-D images were 128 pixels $\times$ 128 pixels in size; thus, the system matrix consisted of $55 \times 170 \times 128 \times 128$ elements with 4.33% sparsity.

The initial image estimate for EM-ML, ISRA, WLS, SART, and ISWLS was

$$x_{oi} = \frac{\sum_{j=1}^{M} y_j}{N}, \quad i = 1, 2, ..., N \quad (14)$$

where $y_j$ is the value of the $j$th sinogram element and $N$ represents the total number of image pixels ($N = 128 \times 128$ in this implementation).

Fig. 1 shows the reconstructed transaxial images with EM-ML, ISRA, WLS, ISWLS, and SART after 1, 10, and 50 iterations.

In Fig. 2, cross-correlation coefficient $c$ [11] of every iterative method is plotted versus the number of iterations. The cross-correlation coefficient $c$ was calculated according to the equation

$$c = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} (\text{Recon}_{ij} - \bar{\text{Recon}})(\text{Real}_{ij} - \bar{\text{Real}})}{\sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} (\text{Recon}_{ij} - \bar{\text{Recon}})^2} \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} (\text{Real}_{ij} - \bar{\text{Real}})^2}} \quad (15)$$

where $\bar{\text{Recon}}$ and $\bar{\text{Real}}$ are the reconstructed image and the true phantom activity image mean values, respectively. Cross-correlation coefficient is a similarity measure between reconstructed and real radiodistribution image. Its values are in the
range \([-1, 1]\). Value \(c = 1\) corresponds to fully correlated images.

Except for the cross-correlation coefficient that shows the average performance of the reconstruction methods, local CNRs [12] for rods with different diameters were calculated. CNRs for 4.8-, 3.2-, and 1.6–mm rod diameters were computed, using squared regions-of-interest (ROIs), 4.55, 3.85, and 2.15 mm in size, respectively. The ROIs were placed inside the corresponding objects. The number of selected ROIs was equal to the number of same sized objects. ROIs of the same sizes were positioned in three different background areas, each CNR ROI was defined as

\[
\text{CNR}_{\text{ROI}} = \frac{R_{\text{obj ROI}} - R_{\text{Backg ROI}}}{\sigma_{\text{Backg ROI}}}
\]

where \(R_{\text{obj ROI}}\) is the mean value of reconstructed objects in the corresponding ROIs, and \(R_{\text{Backg ROI}}\) is the mean value of the three background ROIs in each case.

Further, \(\sigma_{\text{Backg ROI}}\) is the standard deviation of background values in the corresponding ROIs. The graphs in Fig. 3 illustrate the variation of CNR_{ROI} with respect to the number of iterations for the three different object diameters.

In Fig. 4, the reconstruction time for every iterative algorithm is presented as a function of the number of iterations. Reconstruction time calculations were performed on a Pentium M processor 1400 MHz (Intel Corp., Santa Clara, CA) personal computer (RAM 1280MB) under Windows XP Professional.

B. Comparative Evaluation of OS Versions

For the evaluation of the OS iterative reconstruction methods, the same methodology as in part A, was used. The initial image estimate for OSEM, OS-ISRA, OS-WLS, and OS-ISWLS was the same as described in (14). The algorithms under evaluation were OSEM, OS-ISRA, OS-WLS, and OS-ISWLS.

Fig. 5 presents reconstructed images with OSEM, OS-ISRA, OS-WLS, and OS-ISWLS after 1, 10, and 50 iterations for 15 subsets.

In Fig. 6, cross-correlation coefficient \(c\) of every OS iterative method is plotted versus the number of iterations, and in Fig. 7, CNRs for two objects of different diameter are plotted versus the number of iterations. CNRs are derived with the same method as explained in (16).

In Fig. 8, the reconstruction time for every OS iterative algorithm is presented as a function of the number of iterations.

IV. DISCUSSION

EM-ML, ISRA, WLS, SART, and ISWLS achieve high cross-correlation values as the number of iterations increases.
Cross-correlation coefficients of EM-ML and ISRA present almost the same variance as functions of the number of iterations. Reconstructed images with SART, WLS, and ISWLS achieve high correlation levels with the real radiodistribution image after the first ten iterations. Moreover, ISWLS shows the same high values of cross-correlation coefficient as WLS. As the number of iterations outreaches 90 to 95, cross-correlation of all iterative algorithms converges to the same value.

Cross-correlation coefficient is a similarity metric between two images with pixel intensities linearly connected. If the two images are identical, the cross-correlation coefficient is 1. In our research cross-correlation coefficient reaches up to 0.77. This is because extra corrections prior or during the reconstruction should be made such as attenuation, scatter, and random corrections. Despite the fact that these corrections have not been made, the final result of the average performance of ISWLS is not altered. It is estimated that cross-correlation coefficient will approach 1 after these corrections are made.

As illustrated in Fig. 3(a), EM-ML converges slowly during the first 50 iterations to the best approximation of the true image. After the first 50 iterations the algorithm enhances more image detail for small-sized objects [see Fig. 3(c)]. However, after that point, image contrast decreases, while noise component starts to increase. On the contrary, ISRA, in comparison to EM-ML, presents better performance between 50–100 iterations, where it reaches relatively high and constant CNRs values for big- and small-sized image objects. WLS shows almost identical noise manipulation as EM-ML. Image contrast starts degrading after 50 iterations due to increasing noise. On the other hand, WLS reaches the same CNR values faster than EM-ML in the first iterations. SART’s performance is a compromise between EM-ML, ISRA, and WLS during the first 50 iterations. CNRs of both, small- and big-image objects, achieve adequate values.
Reconstruction time of EM-ML, WLS, and SART is almost the same as a function of the number of iterations (≈3.8 s/iteration). Although, it is not obvious from Fig. 4, ISRA and ISWLS are slower than EM-ML, WLS, and SART during the first nine iterations. Their reconstruction speed is gradually improving with increasing number of iterations. ISWLS and ISRA reconstruction time converges to the others’ reconstruction time after ten iterations. The reason for slow reconstruction process during the first iterations lies in the time needed for backprojection computations (\( \sum_{i=1}^{M} a_{ij}y_j \) for ISRA and \( \sum_{i=1}^{M} a_{ij}y_j^2 \) for ISWLS) in the first iteration.

According to Fig. 6, the OS versions achieve high cross-correlation coefficients during the first ten iterations. They reach the value of 0.75 earlier than the simple versions, which reach the same value of cross-correlation coefficient after 50 iterations. OS-ISWLS presents the best performance during the first six–nine iterations. During these iterations, OS-ISWLS achieves higher cross-correlation coefficient values than OSEM and shows similar behavior to OS-ISRA. OS-ISRA presents the highest cross-correlation values as the number of iteration increases beyond ten iterations. After the first ten iterations, the cross-correlation coefficient of OS-ISWLS decreases, which implies that OS-ISWLS is affected very much from the introduction of noise. OS-WLS does not show adequate cross-correlation coefficient values.

Fig. 7 shows that for objects 4.8 mm in diameter all OS algorithms present similar CNRs. Although it is not apparent from Fig. 7, OS-ISWLS is preferable for one iteration as far as CNRs are concerned. In the first iteration it reaches the value of 34.3 for 4.8 mm object diameter (where the others have CNRs’ values < 29) and 2.7 for 1.6-mm object diameter (where the others have CNRs values < 2.1). In general OS-ISWLS is comparable to OSEM and shows better performance than OS-WLS according to the CNRs graph (see Fig. 7), which indicates that OS-ISWLS shows better noise manipulation than OS-WLS. OS-ISRA reaches the highest CNRs for very small objects (1.6 mm in diameter). To sum up, OS-ISWLS’s performance seems to be a compromise between OS-ISRA and OS-WLS. We reached to the same conclusions using 3, 9, and 24 subsets (not shown).

Reconstruction time is the same for all OS algorithms under study. One iteration lasts 29 s for all OS methods.

The choice of 15 subsets was made after a comparative study of ISWLS and OS-ISWLS. We compared OS-ISWLS to ISWLS using 3, 9, 15, and 24 subsets. The comparative criterion was CNR, calculated as described in (16). Small number of subsets resulted in CNRs similar to ISWLS. Increasing the number of subsets resulted in images with high CNRs during the first ten iterations. Using 15 or 24 subsets the reconstructed images reaches the highest CNR values before ten iterations. ISWLS achieves the same highest CNRs after 50 iterations. We chose 15 subsets to 24 because this number of subsets presented smaller image degradation as the number of iteration increased.

In this study, data were corrected for scanner sensitivity prior to the reconstruction process. Such a correction could be incorporated during the reconstruction process, but that exceeds our research purpose. Recently sensitivity correction has been combined with FBP to build the weighted filtered backprojection algorithm and applied to optical tomography data [13].

Our future plan is to combine the new algorithms with penalized techniques that take into account a priori information of radiopharmaceutical spatial distribution. A priori information usually consists in smoothing factors or anatomical data in order to reduce reconstruction process dependence on noise component. We plan to use median root prior (MRP) [14] as a smoothing factor in combination to one-step-late [15] algorithm and to evaluate the new MRP algorithm, namely MRP-ISWLS with established penalized methods.

V. CONCLUSION

In this paper, different simultaneous iterative reconstruction schemes were applied to data acquired from a simulation of a small-animal PET scanner. A new iterative scheme was introduced, namely ISWLS. SART, EM-ML, ISRA, WLS, and ISWLS and their OS versions were implemented and evaluated, in terms of task-dependent measures for quantization and detection. In general, EM-ML’s and SART’s results are inferior to ISWLS’s results. ISWLS shows comparable results to WLS and ISRA, but in relation to them excels, because it combines WLS’s reconstruction acceleration with ISRA’s good noise manipulation. OS-ISWLS shows its best performance between the first six–nine iterations. Its behavior seems to be a compromise between OS-ISRA and OS-WLS. ISWLS could be preferable for use in 3-D reconstruction applications, where the precalculation of the factor \( \sum_{i=1}^{M} a_{ij}y_j^2 \), which is constant, will lessen the computational cost and demands for high computational memory.

REFERENCES

[8] E. Karali, “Development of medical data reconstruction techniques based on a small-animal PET system through optimising and comparing medical


Author’s photograph and biographies not available at the time of publication.