King's Research Portal

[Link to publication record in King's Research Portal](#)

Patch-based image reconstruction for PET using prior-image derived dictionaries

You may also be interested in:

Anatomically-aided PET reconstruction using the kernel method
Will Hutchcroft, Guobao Wang, Kevin T Chen et al.

Sparsity-constrained PET image reconstruction with learned dictionaries
Jing Tang, Bao Yang, Yanhua Wang et al.

Anatomy-guided brain PET imaging incorporating a joint prior model
Lijun Lu, Jianhua Ma, Qianjin Feng et al.

4D image reconstruction for emission tomography
Andrew J Reader and Jeroen Verhaeghe

Sparse representation and dictionary learning penalized image reconstruction for positron emission tomography
Shuhang Chen, Huafeng Liu, Pengcheng Shi et al.

PET image reconstruction with anatomical edge guided level set prior
Jinxiu Cheng-Liao and Jinyi Qi

MR-guided dynamic PET reconstruction with the kernel method and spectral temporal basis functions
Philip Novosad and Andrew J Reader

Patch-based image reconstruction for PET using prior-image derived dictionaries

Marzieh S Tahaei[1] and Andrew J Reader[2]

[1] McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada
[2] Division of Imaging Sciences and Biomedical Engineering, Department of Biomedical Engineering, King's College London, St. Thomas' Hospital, London, UK

E-mail: marzieh.tahaei@mail.mcgill.ca

Abstract

In PET image reconstruction, regularization is often needed to reduce the noise in the resulting images. Patch-based image processing techniques have recently been successfully used for regularization in medical image reconstruction through a penalized likelihood framework. Re-parameterization within reconstruction is another powerful regularization technique in which the object in the scanner is re-parameterized using coefficients for spatially-extensive basis vectors. In this work, a method for extracting patch-based basis vectors from the subject's MR image is proposed. The coefficients for these basis vectors are then estimated using the conventional MLEM algorithm. Furthermore, using the alternating direction method of multipliers, an algorithm for optimizing the Poisson log-likelihood while imposing sparsity on the parameters is also proposed. This novel method is then utilized to find sparse coefficients for the patch-based basis vectors extracted from the MR image. The results indicate the superiority of the proposed methods to patch-based regularization using the penalized likelihood framework.

Keywords: positron emission tomography, image reconstruction, anatomical prior, hybrid PET-MR, patch representation, sparsity, re-parameterization

(Some figures may appear in colour only in the online journal)

1. Introduction

PET image reconstruction seeks to estimate a 3D image (i.e. object representation) of the radiotracer concentration from line of response measurements. Statistical algorithms for PET

image reconstruction have become increasingly popular due to their ability to model the noise in the measurement data and also their flexibility in modelling the physical characteristics of the scanning process. Among iterative methods, the maximum likelihood expectation maximization (MLEM) method and its variations (e.g OSEM (Green 1990)) are now widely being used in clinical practice as they provide a visually improved image quality and can have better performance in lesion detection than conventional methods (Wells *et al* 2000).

PET reconstruction is an ill-conditioned inverse problem. Hence, a regularization approach is required to restrict the space of feasible solutions. During the last decade many efforts have been taken to reduce the noise of the reconstructed images, spanning from the use of simple spatial smoothness priors, through more robust edge preserving smoothing priors to the recent methods of patch-based priors.

The literature on regularization in PET reconstruction can be divided into two main categories: (i) including a prior function in the objective using a maximum *a posteriori* (MAP) framework and (ii) representing the radioactive distribution as a superposition of spatial basis vectors such as blobs (re-parameterization).

Most regularization methods in the PET reconstruction literature focus on designing appropriate prior functions to be utilized in a MAP framework (Fessler 1997, Nuyts *et al* 2002, 2003, Wang and Qi 2012, Ahn *et al* 2015). Many of these MAP-based approaches benefit from the boundary information from magnetic resonance imaging (MRI) and computed tomography (CT) images in order to encourage smoothness only within anatomical structures and thus preserve edges (Ardekani *et al* 1996, Bowsher *et al* 1996, Rangarajan *et al* 2000, Somayajula *et al* 2005, Tang and Rahmim 2009, Cheng-Liao and Qi 2011).

Patch-based image processing techniques have become increasingly popular in the last decade. They leverage two interesting characteristics of natural images: the intrinsic redundancy and the locally structured patterns in images. These characteristics allow image patches to be sparsely represented (very few non-zero coefficients are needed to represent their content) as a superposition of some learned over-complete set of basis vectors. Sparse representation of image patches has proven to achieve good performance in a variety of image processing tasks like denoising and restoration (Buades *et al* 2005, Dabov *et al* 2007). Their robustness to noise soon motivated the medical imaging community to use them for medical image reconstruction.

The authors in Chen *et al* (2015) and Tang *et al* (2014) have proposed PET image reconstruction methods that use sparse representation of image patches using a dictionary learned from anatomical images. These methods are based on a Bayesian (or MAP) framework where the regularized objective function is composed of a data fidelity term and a sparse representation error term. A maximum *a posteriori* expectation maximization (MAPEM) or an approximate one-step-late MAPEM (Green 1990) method is utilized to estimate the image.

On the other hand, some of the recent anatomical-based regularization methods such as kernel-based regularization in Wang and Qi (2015) and Novosad and Reader *et al* (2016) and the recent super-voxel method presented in Jiao *et al* (2015) fall into the re-parameterization category and hence the conventional MLEM algorithm can be used to reconstruct the image. These methods have shown promising results in PET image reconstruction. However these methods work best in situations where the functional activity within tissues is homogeneous. This work tries to address this issue using a patch-based re-parameterization framework. To do so, the radioactivity distribution in the field of view is modelled as a composition of overlapping patches, where each patch is represented as a superposition of learned dictionary atoms.

Motivated by non-local patch-based denoising techniques, a novel patch-based basis function extraction method from a prior images is proposed. In this method, the patches from the prior image are first clustered into C sets and for each cluster a dictionary is learned from the

patches in that cluster. The resulting dictionaries are then used as basis vectors to represent patches in the PET image. In order to estimate the coefficients for these basis vectors, only a re-parameterization of the Poisson log likelihood is required and hence the image can be reconstructed using the conventional MLEM algorithm.

Furthermore, another important contribution of this work is to propose an algorithm for optimizing the re-parameterized Poisson log-likelihood objective function while imposing sparsity on the coefficients. An ADMM method is utilized to solve this l_1 penalized likelihood function. The proposed algorithm is then used to impose a sparsity penalty on the coefficients of the patch-based basis vectors extracted from the subject's T1-weighted MR image and the effect of the sparsity parameter is explored using simulation data.

In order to enhance readability, throughout the paper, bold capital letters are used to indicate matrices, bold lower-case letters represent column vectors and all non-bold letters denote scalar values.

2. Background

2.1. Sparse patch-based representation of images

Let us assume that $\mathbf{D} \in \mathbb{R}^{M \times N}$ is a dictionary of N atoms (basis vectors) learned from a database of overlapping patches of size M (i.e. $\sqrt{M} \times \sqrt{M}$ for a 2D image) taken from the image under consideration. The dictionary is learned in such a way that each patch $\mathbf{p}_k$ within the database can be represented as a sparse superposition of the dictionary atoms, i.e. $\mathbf{p}_k \approx \mathbf{D}\boldsymbol{\alpha}_k$, where $\boldsymbol{\alpha}_k \in \mathbb{R}^N$ is the coefficient vector representing the patch $\mathbf{p}_k$.

For a given dictionary, finding the sparse representation (coding) of the image $\mathbf{x}$ firstly involves extracting overlapping patches from the image. Let $\mathbf{R}_k \in \mathbb{R}^{M \times J}$ be a binary matrix operator that extracts the patch $\mathbf{p}_k$ at location k from the image $\mathbf{x}$:

$$\mathbf{p}_k = \mathbf{R}_k \mathbf{x} \tag{1}$$

The transpose of the matrix $\mathbf{R}_k$ operating on the patch $\mathbf{p}_k$, gives an image in which $\mathbf{p}_k$ is put back into the position of patch k and zero-pads the image elsewhere. This procedure is illustrated in figure 1.

Once all the patches are extracted, for each patch $\mathbf{p}_k$, the sparse coding algorithm finds the set of coefficients $\boldsymbol{\alpha}_k$ for the dictionary atoms which minimizes the representation error while imposing sparsity on the coefficient vector. This is usually done using the following objective function for each patch:

$$\arg\min_{\boldsymbol{\alpha}_k} \|\mathbf{p}_k - \mathbf{D}\boldsymbol{\alpha}_k\|_2^2, \text{such that } \|\boldsymbol{\alpha}_k\|_0 < T \tag{2}$$

where T is the sparsity parameter, indicating that fewer than T atoms must be used. Each pixel in the reconstructed image $\mathbf{x}$ is then obtained by averaging reconstructed patches that cover it together:

$$\mathbf{x} \approx \sum_{k=1}^{K} \frac{\mathbf{R}_k^T(\mathbf{D}\boldsymbol{\alpha}_k)}{\mathbf{q}} \tag{3}$$

where $\mathbf{q} = \sum_{k=1}^{K} \mathbf{R}_k^T \vec{\mathbf{1}}$ is a vector of size J in which each elements q_j indicates the number of patches contributing to pixel j. Here and throughout the paper, *vector* by *vector* multiplication and division should be interpreted as element-wise multiplication and division respectively.
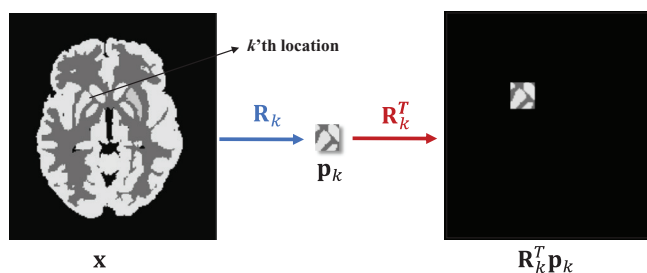
Figure 1. Schematic illustration of the patch-extraction operator and its transpose. Note that the order of pixels arranged in vector **x** depends on the method used to store the 2D/3D matrices.

2.2. MLEM

Let $\mathbf{m} = \{m_1 \dots m_I | m_i \in \mathbb{Z}^+ \cup \{0\}\}$ be a vector in which each element m_i contains the number of counts accumulated in bin i during a PET scan. Because of the counting nature of the annihilation events, m_i can be well modelled by a Poisson distribution with mean $\bar{m}_i$:

$$Pr(m_i|\bar{m}_i) = \frac{\mathrm{e}^{-\bar{m}_i}\bar{m}_i^{m_i}}{m_i!} \tag{4}$$

The mean of the measurement data can be modelled as an affine transform of the modelled mean activity distribution $\mathbf{x} \in \mathbb{R}^J$ within the scanner through the system matrix $\mathbf{A} = \{a_{i,j}\} \in \mathbb{R}^{I \times J}$ as :

$$\bar{\mathbf{m}} \approx \mathbf{Ax} + \mathbf{r}, \tag{5}$$

where I and J are the number of PET measurements and image pixels respectively. $\mathbf{r}$ is the model of the random and scatter events. Then the log likelihood of the true activity distribution of $\mathbf{x}$ given $\mathbf{m}$ can be written as:

$$L(\mathbf{x}|\mathbf{m}) = \sum_{i=1}^{I} (m_i \ln(\sum_{j=1}^{J} a_{i,j}x_j) - \sum_{j=1}^{J} a_{i,j}x_j). \quad \text{such that} \quad \mathbf{x} \geqslant 0 \tag{6}$$

$L(\mathbf{x}|\mathbf{m})$ can be maximized using the iterative MLEM algorithm. MLEM starts with a non-zero initialization of $\mathbf{x}$ and converges to the solution using the following update rule:

$$\mathbf{x}^{n+1} = \frac{\mathbf{x}^n}{\mathbf{s}} \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{Ax}^n + \mathbf{r}} \tag{7}$$

where n is the iteration number and $\mathbf{s} = \mathbf{A}^T \vec{\mathbf{1}}$ is the sensitivity image. From an optimization transfer perspective, in each iteration the MLEM method maximizes the following surrogate function for $L(\mathbf{x}|\mathbf{m})$ (Reader and Verhaeghe 2014):

$$G(\mathbf{x}|\mathbf{x}^n, \mathbf{m}) = \sum_{j=1}^{J} s_j(x_j^{\mathrm{EM}} \ln(x_j) - x_j), \tag{8}$$

where $\mathbf{x}^{\mathrm{EM}}$ is equal to $\mathbf{x}^{n+1}$ from the EM update in equation (7), therefore being a function of $\mathbf{x}^n$ and $\mathbf{m}$.

2.3. MAPEM

The MAP formulation for PET reconstruction can be written as:

$$L_{a\ posteriori}(\mathbf{x}|\mathbf{m}) = L(\mathbf{x}|\mathbf{m}) + \ln(Prior(\mathbf{x})), \tag{9}$$

Assuming that the prior function for image $\mathbf{x}$ follows a Gibbs distribution $\frac{1}{Z}\exp(-\beta U(\mathbf{x}))$, one can write the MAP formulation as:

$$L_{a\ posteriori}(\mathbf{x}|\mathbf{m}) = L(\mathbf{x}|\mathbf{m}) - \beta U(\mathbf{x}) \tag{10}$$

2.3.1. MAP with quadratic penalty (Q-MAP).
In this method, a weighted quadratic function is used as $U(\mathbf{x})$ to penalize large differences in the intensities between the neighbouring pixels:

$$U(\mathbf{x}) = \sum_{j=1}^{J} \sum_{k \in Neib(j)} w_{jk}(x_j - x_k)^2 \tag{11}$$

The resulting MAP objective function ($L_{a\ posteriori}$) can be solved using an optimization transfer framework (De Pierro's decoupling rule) (De Pierro 1995). In this work a Gaussian function is used to determine the weights. See Tahaei and Reader (2014) for a full derivation of the algorithm.

2.3.2. MAP with relative difference penalty (RD-MAP).
The problem with the quadratic penalty is that it can strongly smooth edges in the reconstructed image. One simple way to address this problem is to use the following relative difference penalty function (Nuyts *et al* 2002) :

$$U(\mathbf{x}) = \sum_{j=1}^{J} \sum_{k \in \text{Neib}(j)} \frac{(x_j - x_k)^2}{(x_j + x_k) + \gamma|x_j - x_k|} \tag{12}$$

where γ is a parameter that controls the degree of edge-preservation. In our implementation, the resulting MAP objective function is maximized using the one-step-late approach (Green 1990).

2.3.3. Sparse patch-based MAPEM (sparse PB-MAPEM).
One way to incorporate a patch-based representation for regularization in PET reconstruction is through the MAPEM framework. The penalty function in this model can be defined as minimizing representation error of patches in the image using a fixed dictionary while imposing sparsity on the coefficients. This leads to minimizing the following negative penalized log-likelihood objective function (Chen *et al* 2015):

$$\begin{aligned} &\min && -\lambda L(\mathbf{x}|\mathbf{m}) + R_{\text{sparsity}}([\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_K]|\mathbf{x}) \\ &= \min && -\lambda \sum_{i=1}^{I}(m_i \ln(\sum_{j=1}^{J} a_{i,j}x_j) - \sum_{j=1}^{J} a_{i,j}x_j) + \sum_{k=1}^{K}(\|\mathbf{R}_k\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}_k\|_2^2 + \mu\,\|\boldsymbol{\alpha}_k\|_0) \end{aligned} \tag{13}$$

where μ controls the sparsity. The above optimization can be solved in an alternating manner by (i) fixing $\mathbf{x}$ and solving for the $\boldsymbol{\alpha}_k$'s and then (ii) fixing the $\boldsymbol{\alpha}_k$'s and solving for $\mathbf{x}$ iteratively. In Chen *et al* (2015), problem (i) is solved using the orthogonal matching pursuit (OMP) algorithm (Tropp *et al* 2007). OMP is a fast greedy algorithm. It starts with an empty set of selected atoms and at each iteration finds an unselected atom from $\mathbf{D}$ that best matches the current residual. This atom is then added to the set, and the residual as well as the coefficients
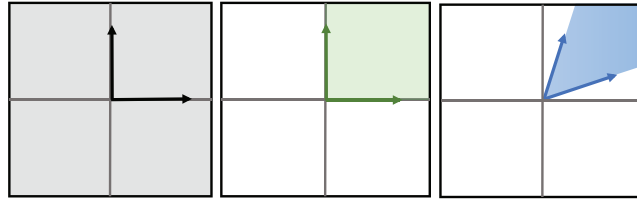
Figure 2. The span of two basis vectors in a 2D space. Left: the span of two orthogonal basis vectors. Middle: the span of two orthogonal basis vectors with non-negative coefficients. Right: the span of two non-orthogonal basis vectors with non-negative coefficients.

are updated using this updated set of dictionary atoms. The iterations are continued until the residual is less than a specified error tolerance called ϵ.

To solve problem (ii), an optimization transfer framework is used (De Pierro 1995). See Chen *et al* (2015) for a detailed derivation of the algorithm. In Chen *et al* (2015), the dictionary is learned from patches extracted from CT images. In this study however, for the sparse PB-MAPEM method the dictionary is learned from the subject's registered MR image using a Matlab implementation of the K-SVD algorithm (www.cs.technion.ac.il/~elad/software/).

2.4. Re-parameterized MLEM

Considering that the radioactivity distribution is modelled as a vector $\mathbf{x} \in \mathbb{R}^J$, one can represent the vector $\mathbf{x}$ as a superposition of some basis vectors:

$$\mathbf{x} = \sum_{l=1}^{L} \theta_l \phi_l = \mathbf{\Phi}\boldsymbol{\theta} \tag{14}$$

where $\mathbf{\Phi} \in R^{J \times L}$ is a matrix in which the l'th column is the l'th basis vector: $\mathbf{\Phi} = \{\phi_1, ..., \phi_L\}$ and $\boldsymbol{\theta}$ is the coefficient vector in which each θ_l is the coefficient for the l'th basis vector.

The are infinitely many choices of basis vectors. For a given set of basis vectors the coefficient vectors are considered as the representation of the radioactivity distribution in the given basis. In this setting, an obvious way to achieve regularization is to impose constraints on the coefficient vector and hence limit the space of possible representations of $\mathbf{x}$ for the given basis set $\mathbf{\Phi}$. This is usually achieved through a Bayesian approach.

Now, let us assume that for a given non-negative set $\mathbf{\Phi}$, one is only allowed to use non-negative coefficients to represent $\mathbf{x}$. With this new non-negativity constraint the span of the basis vectors is dependent on the shape of the basis vectors. Hence one can achieve different levels of regularization depending on the shape of the basis vectors. For a fixed number of basis vectors, the less the orthogonality between the basis vectors, the higher the regularization (the more limited the span of the possible radioactivity distributions). Figure 2 illustrates how the use of non-orthogonal basis vectors with non-negative coefficients can restrict the space of solutions.

This is very important for PET image reconstruction since the Poisson log-likelihood objective of the radioactivity distribution given the measurement data inherently imposes a non-negativity constraint on the mean of the radioactivity distribution. Therefore, one can achieve regularization by choosing non-negative basis vectors and using the traditional MLEM algorithm. This use of non-trivial basis vectors is considered as one of the categories of re-parameterization in the PET reconstruction literature. Different choices of overlapping and non-overlapping basis vectors have been explored in a number of studies. The use of blobs,

Gaussians and anatomical regions of interest (ROIs) as well as the recent kernel-based method are among these methods (Lewitt 1990, 1992, Jacobs *et al* 1998, Jiao *et al* 2015, Wang and Qi 2015). In the re-parameterization framework the MLEM update rule becomes:

$$\boldsymbol{\theta}^{n+1} = \frac{\boldsymbol{\theta}^n}{\boldsymbol{\Phi}^T \mathbf{A}^T \vec{\mathbf{1}}} \boldsymbol{\Phi}^T \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{A}\boldsymbol{\Phi}\boldsymbol{\theta}^n + \mathbf{r}} \tag{15}$$

3. Proposed method

In the following section, a framework for incorporating a patch-based representation as the re-parameterization of the whole image using matrix multiplication is presented. In the remainder of this section, novel methods for estimating the coefficients and extracting the basis vectors for this patch-based re-parameterization framework are proposed.

3.1. General framework: patch-based representation as image re-parameterization

Consider a general case in which each patch is represented using a dictionary associated with it. This dictionary can be identical for all patches or it can be chosen according to the structure or location of the patch. Let's assume that $\mathbf{D}_k = [[\mathbf{D}_k]_1...[\mathbf{D}_k]_N]$ is the dictionary matrix with $[\mathbf{D}_k]_d$ representing the d'th column vector (or dictionary atom) used for the representation of patch k (we will talk about how this dictionary can be learned later on), i.e. $\mathbf{p}_k \approx \mathbf{D}_k \boldsymbol{\alpha}_k$. The basis set $\boldsymbol{\Phi}_k$ can be defined by replacing each atom $[\mathbf{D}_k]_d$ in $\mathbf{D}_k$ by $\mathbf{R}_k^T[\mathbf{D}_k]_d$:

$$\boldsymbol{\Phi}_k = [\mathbf{R}_k^T[\mathbf{D}_k]_1, \mathbf{R}_k^T[\mathbf{D}_k]_2 \ldots \mathbf{R}_k^T[\mathbf{D}_k]_N] \tag{16}$$

By generating $\boldsymbol{\Phi}_1$ to $\boldsymbol{\Phi}_K$ from $\mathbf{D}_1$ to $\mathbf{D}_K$ respectively and concatenating them together we obtain a huge sparse matrix called $\boldsymbol{\Phi}$:

$$\boldsymbol{\Phi} = [[\boldsymbol{\Phi}_1], ..., [\boldsymbol{\Phi}_K]]. \tag{17}$$

Equivalently, the vector $\boldsymbol{\theta}$ is obtained by concatenating $\boldsymbol{\alpha}_1$ to $\boldsymbol{\alpha}_k$'s together :

$$\boldsymbol{\theta}^T = [[\boldsymbol{\alpha}_1]^T, ..., [\boldsymbol{\alpha}_k]^T]^T \tag{18}$$

where $\boldsymbol{\alpha}_k$ is the sparse coefficient vector for representing the patch $\mathbf{p}_k$.

As mentioned before, the final image is formed by combining sparsely represented patches together. One can easily show that $\boldsymbol{\Phi}\boldsymbol{\theta} = \sum_{k=1}^{K} \mathbf{R}_k^T(\mathbf{D}_k \boldsymbol{\alpha}_k)$. Hence, by normalizing for overlapping patches, the reconstructed image can be written in matrix multiplication form as:

$$\mathbf{x} \approx \mathbf{Q}^{-1}\boldsymbol{\Phi}\boldsymbol{\theta}, \tag{19}$$

where $\mathbf{Q} \in R^{J \times J}$ is a diagonal normalization matrix with diagonal elements identical to those of $\mathbf{q}$ in equation (3). For notational convenience, hereafter we refer to $\mathbf{Q}^{-1}\boldsymbol{\Phi}$ as $\boldsymbol{\Phi}'$.

$$\mathbf{x} \approx \boldsymbol{\Phi}'\boldsymbol{\theta}, \tag{20}$$

3.2. Extracting the basis vectors from a prior image

In this section a new basis extraction method which uses patches from a prior image to represent a patch in the PET image is proposed. This method is motivated by many novel non-local denoising techniques in the image processing literature including the non-local means and

BM3D algorithms (Kervrann and Boulanger 2006, Dabov *et al* 2007, Chatterjee and Milanfar 2012).

In the proposed method, instead of learning one large dictionary from the prior image to represent every patch in the PET image, one can first cluster similar patches together and then learn a separate dictionary from each cluster. The idea of clustering prior to dictionary learning has been previously used in image denoising (Chatterjee and Milanfar 2009) and shown to improve the performance. Using this strategy in the aforementioned re-parameterization framework, can educe the number of basis vectors needed to represent each patch in the PET image. This in turn reduces the number of parameters to be estimated within the MLEM algorithm.

Once the dictionaries are learned, for every patch $\mathbf{p}_k$ in the PET image, the cluster to which the corresponding patch in the anatomical image (i.e. the patch in the prior image, such as an MR image, at location k) belongs to is selected and the dictionary learned from this cluster is used to represent $\mathbf{p}_k$.

The problem with this approach, i.e. using a dictionary associated with the same location in the prior image, is that in order to be able to use the boundary information from the prior image the intensity values on the sides of that boundary cannot have opposite relative contrast compared to those of the selected patch in the PET image. This is due to the implicit non-negativity constraint imposed by the MLEM algorithm on the basis vectors and the coefficients.

An example is a T1-weighted MR image for reconstructing fluorodeoxyglucose (FDG) images. Since the contrast in PET grey matter and white matter are inverted a dictionary associated with cortical boundary regions in a T1-weighted MR image cannot represent a patch in these regions in the PET image. One simple way to address this issue is to modify the prior image beforehand. For the case of a T1-weighted MR image used in this study, this is done by segmenting the MR image into white matter and grey matter and changing the intensity of grey matter to a value greater than the maximum intensity in the white matter using the following:

$$\forall j \in \text{grey-matter region} \qquad y_j = aw^*, a > 1 \tag{21}$$

where y is the modified MR image from which the patches are extracted for dictionary learning and w^* is the maximum intensity in the white-matter.

Once the opposite contrast issue is resolved, the basis vectors for PET reconstruction can be extracted from the image. To do so, first the resulting modified T1-weighted MR image y is decomposed to overlapping patches of size $\sqrt{M} \times \sqrt{M}$. The patch decomposition performed on the PET and the modified MR image should be identical so every patch $\mathbf{p}_k$ from the PET image has one and only one correspondence at the same location and of the same size in the MR image. Every patch is then normalized by subtracting the minimum and dividing by the new maximum. Let us call the normalized patch from the modified MR image at location k, $\mathbf{w}_k$. The normalized patches are then clustered into C distinct clusters $\{\mathbf{U}_1, ..\mathbf{U}_c, ...\mathbf{U}_C\}$ using the k-means algorithm. For each cluster index c, a dictionary $\mathbf{D}_c$ is learned from the patches in the cluster using the non-negative matrix factorization algorithm (described in the following section). Then for every patch $\mathbf{p}_k$ in the PET image the dictionary atoms learned from the cluster to which the normalized patch from location k in the MR image belongs is used to represent the patch.

Moreover, in order to ensure that the set of basis vectors is able to represent structures not present in the prior image, an all-one vector rescaled to unit length is added to the dictionary:

$$\mathbf{p}_k \approx \left[\mathbf{D}_{c'}, \frac{\vec{\mathbf{1}}}{\sqrt{M}} \right] \alpha_k \quad c' \text{ is the index of the cluster containing patch } \mathbf{w}_k \tag{22}$$
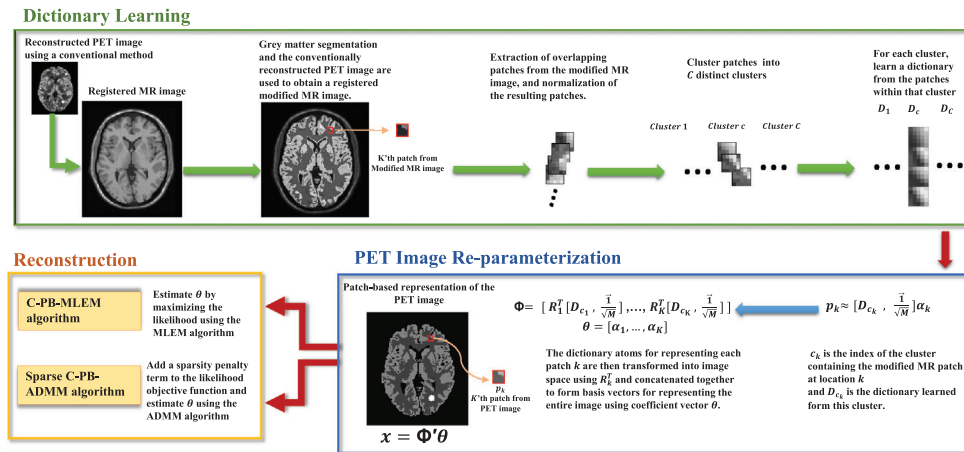
Figure 3. Illustration of the proposed method.

Note that adding this constant atom to the set of basis vectors, is expected to work well for relatively small patches and is empirically confirmed by simulating lesions in the PET data (see Simulation Studies section).

It is also worth noting that because of the normalization of the patches before basis extraction, the relative contrast imposed by the prior image (e.g. the modified MR image) will not force a similar contrast on the PET image. Therefore for the case of using the modified T1-weighted MR, as an example, the algorithm is not sensitive to the value of *a* in equation (21), as long as *a* is greater than 1. In our implementation *a* was set to 2. The procedure of extracting patch-based basis vectors from the subject's MR image is illustrated in figure 3.

3.2.1. Dictionary learning using sparse non-negative matrix factorization. Because of the implicit non-negativity constraint within the Poisson log likelihood objective function, a non-negative matrix factorization (NMF) is used for dictionary learning to ensure the non-negativity of the re-parameterized image. Let $\mathbf{U}_c$ be a matrix where each column is a patch of the modified MR image associated with cluster c. The goal of NMF is to factorize $\mathbf{U}_c$ into two non-negative matrices $\mathbf{D}$ and $\mathbf{E}$: $\mathbf{U}_c \approx \mathbf{DE}$. $\mathbf{D}$ can be viewed as the dictionary and columns of $\mathbf{E}$ can be considered as the coefficient vectors for patches in cluster c.

$$\underset{\mathbf{D},\mathbf{E}}{\text{minimize}} = \|\mathbf{U} - \mathbf{DE}\|_{\mathrm{F}}$$
$$\text{such that} \quad \mathbf{D} \geqslant 0, \quad \mathbf{E} \geqslant 0 \tag{23}$$

where $\|.\|_{\mathrm{F}}$ denotes the Frobenius norm. While the non-negativity constraint in NMF naturally favours sparsity, several studies in recent years have tried to impose a sparsity penalty on the data fidelity objective to enhance sparsity in a controllable manner (Hoyer 2004, Peharz and Pernkopf 2012) . For the purpose of this study, the NMF with the l_0 constraint, named NMFL^0H, proposed in Peharz and Pernkopf (2012) is used. In Peharz and Pernkopf (2012), multiplicative update rules are used to update the dictionary matrix in each step. In this study this algorithm is referred to as NMFl_0. The authors have also made a Matlab implementation of this method available online (www.spsc.tugraz.at/tools/). We have utilized this implementation for learning dictionaries from each cluster.

3.3. Estimating the coefficient vector

In this section, two methods for estimating the coefficient vector are presented. The first is the conventional re-parameterized MLEM method, and the latter is a novel method for finding a sparse coefficient vector using ADMM.

3.3.1. Clustered patch-based MLEM (C-PB-MLEM). By substituting $\mathbf{x}$ with $\boldsymbol{\Phi}'\boldsymbol{\theta}$ in the likelihood function L (which can only be done if $\boldsymbol{\Phi}'\boldsymbol{\theta} \geqslant 0$), a re-parameterized ML objective is obtained which can be maximized using the MLEM algorithm. We call this method clustered patch-based MLEM (C-PB-MLEM) which is simply a re-parameterization of the MLEM algorithm with clustered patch-based basis vectors extracted from a prior image. Therefore the update rule for C-PB-MLEM can be written as:

$$\boldsymbol{\theta}^{n+1} = \frac{\boldsymbol{\theta}^n}{\boldsymbol{\Phi}'^T \mathbf{A}^T \vec{\mathbf{1}}} \boldsymbol{\Phi}'^T \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{A}\boldsymbol{\Phi}'\boldsymbol{\theta}^n + \mathbf{r}} \tag{24}$$

Note that when a large number of clusters is used, the patches belonging to each cluster are very similar and hence only a few dictionary atoms are needed to represent the patches belonging to that cluster. On the other hand, when a small number of clusters is used, each cluster may contain patches with diverse structures and hence a larger dictionary may be required to represent the patches belonging to each cluster. For the latter case, imposing sparsity constraint on the coefficient vector can lead to a more stable representation of the object. An algorithm for imposing a sparsity in a log likelihood objective function is proposed in the following section.

3.3.2. Sparse clustered patch-based ADMM (sparse C-PB-ADMM). For over-complete dictionaries, the non-negativity of $\boldsymbol{\Phi}'$ and $\boldsymbol{\theta}$ along with the over-completeness of the representation naturally favours sparse coefficient vectors. However, in order to enhance sparsity of the coefficient vector, one can add a sparsity penalty term to the negative log-likelihood objective. By setting the penalty function to l_1 (as a convex approximation of l_0) , the solution can be achieved by minimizing the following objective function:

$$\arg\min_{\boldsymbol{\theta}} - L(\boldsymbol{\theta}|\mathbf{m}) + \beta \|\boldsymbol{\theta}\|_1,$$
$$\text{such that} \quad \boldsymbol{\theta} \geqslant 0 \tag{25}$$

Using equation (8), the surrogate function for this objective function can be written as:

$$M(\boldsymbol{\theta}|\boldsymbol{\theta}^n, \mathbf{m}) = -\sum_{l=1}^{L} s'_l(\theta_l^{\text{EM}} \ln(\theta_l) - \theta_l) + \beta \|\boldsymbol{\theta}\|_1,$$
$$\mathbf{s}' = \boldsymbol{\Phi}'^T \mathbf{A}^T \vec{\mathbf{1}},$$
$$\boldsymbol{\theta}^{\text{EM}} = \frac{\boldsymbol{\theta}^n}{\mathbf{s}'} \boldsymbol{\Phi}'^T \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{A}\boldsymbol{\Phi}'\boldsymbol{\theta}^n + \mathbf{r}} \tag{26}$$

3.3.3. The algorithm for sparse re-parameterized MAP. To benefit from ADMM, we split the variable in the data fidelity term and the sparsity penalty term by introducing an auxiliary variable $\mathbf{z}$ and writing the optimization in equation (25) in the following constrained setting:

$$\min \quad -L(\boldsymbol{\theta}|\mathbf{m}) + \beta \|\mathbf{z}\|_1,$$
$$\text{subject to} \quad \mathbf{z} = \boldsymbol{\theta}, \boldsymbol{\theta} \geqslant 0. \tag{27}$$

Algorithm 1. ADMM for sparse ML

Initialize: $n = 0, \boldsymbol{\theta}^0 = \vec{\mathbf{1}}, \mathbf{y}^0 = \vec{\mathbf{0}}, \mathbf{z}^0 = \vec{\mathbf{0}}, s = \boldsymbol{\Phi}'^T \mathbf{A}^T \vec{\mathbf{1}}, \rho^0 = 1;$

while *not converged* **do**

> **while** *not converged* **do**
>
>> $\boldsymbol{\theta}^{\mathrm{EM}} = \frac{\boldsymbol{\theta}^n}{\mathbf{s}} \boldsymbol{\Phi}'^T \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{A}\boldsymbol{\Phi}'\boldsymbol{\theta}^n + \mathbf{r}}$ $\mathbf{b} = \mathbf{s} - \rho^n(\mathbf{z}^n - \mathbf{y}^n),$
>>
>> $\boldsymbol{\theta}^{n+1} = \frac{2\mathbf{s}\boldsymbol{\theta}^{\mathrm{EM}}}{\mathbf{b} + \sqrt{\mathbf{b}^2 + 4\rho^n \mathbf{s}\boldsymbol{\theta}^{\mathrm{EM}}}},$
>
> **end**
>
> $\mathbf{z}^{n+1} = \max((\boldsymbol{\theta}^{n+1} + \mathbf{y}^n) - \frac{\beta}{\rho}, \mathbf{0})$
>
> $\mathbf{y}^{n+1} = \mathbf{y}^n + (\boldsymbol{\theta}^{n+1} - \mathbf{z}^{n+1})$
>
> ρ^{n+1} is updated according to equation (36)
>
> $n = n + 1$

end

The augmented Lagrangian for the above equation can be written as:

$$\min_{\mathbf{z} \in \mathbb{R}^+} - L(\boldsymbol{\theta}|\mathbf{m}) + \beta\|\mathbf{z}\|_1 + \frac{\rho}{2}\|\boldsymbol{\theta} - \mathbf{z} + \mathbf{y}\|_2^2 \tag{28}$$

where $\mathbf{y}$ is the Lagrange multiplier vector and ρ is a scalar penalty parameter. ADMM is used to solve the above optimization using the following iterative procedure:

$$\boldsymbol{\theta}^{n+1} = \arg\min_{\boldsymbol{\theta}} - L(\boldsymbol{\theta}|\mathbf{m}) + \frac{\rho}{2}\|\boldsymbol{\theta} - \mathbf{z}^n + \mathbf{y}^n\|_2^2 \tag{29}$$

$$\mathbf{z}^{n+1} = \arg\min_{\mathbf{z}} \beta\|\mathbf{z}\|_1 + \frac{\rho}{2}\|\boldsymbol{\theta}^{n+1} - \mathbf{z} + \mathbf{y}^n\|_2^2 \tag{30}$$

$$\mathbf{y}^{n+1} = \mathbf{y}^n + (\boldsymbol{\theta}^{n+1} - \mathbf{z}^{n+1}) \tag{31}$$

Solving subproblem (29): the surrogate function used for ML estimation is based on the EM algorithm and can be formulated as:

$$\boldsymbol{\theta}^{n+1} = \arg\min_{\boldsymbol{\theta}} \sum_{l=1}^{L} s_l'(\theta_l - \theta_l^{\mathrm{EM}} \ln(\theta_l)) + \frac{\rho}{2}\|\boldsymbol{\theta} - \mathbf{z}^n + \mathbf{y}^n\|_2^2$$

$$= \sum_{l=1}^{L} s_l'(\theta_l - \theta_l^{\mathrm{EM}} \ln(\theta_l)) + \frac{\rho}{2}\sum_{l=1}^{L}(\theta_l - (\mathbf{z}_l^n - \mathbf{y}_l^n))^2 \tag{32}$$

Taking the derivative with respect to θ_l and setting it to zero leads to the following equation;

$$\rho\theta_l^2 + \theta(s_l' - \rho(z_l^n - y_l^n)) - s_l'\theta_l^{\mathrm{EM}} = 0 \tag{33}$$

Solving this equation with respect to θ_l results in the following update:

$$\theta_l^{n+1} = \frac{2s_l'\theta_l^{\mathrm{EM}}}{b + \sqrt{b^2 + 4\rho s_l'\theta_l^{\mathrm{EM}}}}$$

$$\text{where} \quad b = s_l' - \rho(z_l^n - y_l^n) \tag{34}$$

Solving subproblem (30): this is an l_1-regularized non-negative least squares minimization which can be solved using the non-negative variant of the soft-thresholding operator (Daubechies *et al* 2004) defined as:

$$S_\beta^+(\boldsymbol{\theta}^{n+1} + \mathbf{y}^n, \frac{1}{\rho}) = \max((\boldsymbol{\theta}^{n+1} + \mathbf{y}^n) - \frac{\beta}{\rho}, 0) \tag{35}$$

The penalty parameter ρ of the ADMM algorithm is adaptively changed for each iteration based on the method proposed in Boyd *et al* (2011):

$$\rho^{n+1} = \begin{cases} T\rho^n & \text{if } \|\mathbf{e}_{\text{primal}}^n\|_2 > \mu \|\mathbf{e}_{\text{dual}}^n\|_2 \\ \rho^n/T' & \text{if } \|\mathbf{e}_{\text{dual}}^n\|_2 > \mu \|\mathbf{e}_{\text{primal}}^n\|_2 \\ \rho^n & \text{otherwise} \end{cases} \tag{36}$$

where,

$$\mathbf{e}_{\text{primal}}^n = (\boldsymbol{\theta}^n - \mathbf{z}^n), \quad \mathbf{e}_{\text{dual}}^n = -\rho(\mathbf{z}^n - \mathbf{z}^{n-1}) \tag{37}$$

are the primal and dual residuals.

3.4. Implementation

3.4.1. Obtaining the modified MR image.
The T1-weighted MR image is non-rigidly registered to a template space to perform the white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF) segmentation. The resulting segmentations are then registered back to the native space of the MR image. The WM and GM segmentations are then used to modify the T1-weighted MR image using equation (21). Then mutual information based rigid registration is used to register the modified T1-weighted MR image to the PET image reconstructed using MLEM (or OSEM). The resulting MR image is then re-sampled to the PET space to have the same voxel size. The registrations and segmentations are performed using the MINC toolkit (www.bic.mni.mcgill.ca/ServicesSoftware/ ServicesSoftwareMincToolKit).

3.4.2. Parameter selection.
One can note that when increasing the number of clusters, patches in each cluster become more similar and hence fewer dictionary atoms are needed to represent the patches associated with that cluster. In other words, the number of dictionary atoms needed to represent a cluster is inversely related to the total number of clusters.

On the other hand, the number of dictionary atoms needed to describe the underlying structure within a database of patches is directly related to the size of each patch. For simplicity, in our implementation, the number of atoms learned from every cluster is identical. Using this simplification, and considering the aforementioned information one can estimate the number of atoms to be learned from each cluster as a function of patch size and the total number of clusters:

$$N = \frac{M}{C}\delta \tag{38}$$

where N is the number of atoms learned per cluster, M is the number of pixels or voxels in the patch (patch size), C is the total number of clusters and δ is a constant parameter. We found that setting δ to 20 works well for the experiments performed in this study.

4. Simulation studies

For quantitative evaluation of the proposed methods the ensemble-based normalized root mean squared error (n-RMSE) is computed for each pixel within an ROI across the realizations and the results are averaged within the ROI.

$$\text{Mean pixel-based n-RMSE within ROI} = \frac{1}{|\text{ROI}|} \sum_{j \in \text{ROI}} \frac{\sqrt{\sum_{r=1}^{R}(x_j^r - x_j^{\text{true}})^2/R}}{x_j^{\text{true}}} \tag{39}$$

where R is the number of realizations, $\mathbf{x}^r$ is the reconstructed image of the r'th realization of the measurement data and $\mathbf{x}^{\text{true}}$ is the ground truth. $|\text{ROI}|$ is the number of pixels in the ROI.

The mean contrast recovery coefficient (CRC) of the lesions embedded in the simulation data is used as another figure of merit. The following shows the formula used for obtaining the CRC of the r'th realization :

$$\text{CRC}^r = \frac{\left| \frac{1}{|L|} \sum_{j \in L} x_j^r - \frac{1}{|B|} \sum_{j \in B} x_j^r \right|}{\left| \frac{1}{|L|} \sum_{j \in L} x_j^{\text{true}} - \frac{1}{|B|} \sum_{j \in B} x_j^{\text{true}} \right|} \tag{40}$$

where L is the lesion and B is the background region.

4.1. 2D phantom

4.1.1. Setup. A simulation study was performed using the BrainWeb phantom (http://brainweb.bic.mni.mcgill.ca/brainweb). One transverse slice of a realistically simulated T1-weighted MR image was re-sampled to PET image space.

For PET data simulation, the corresponding segmented brain was also re-sampled to the PET space. Then radioactivity distribution values from a PET SORTEO (Reilhac *et al* 2006) (http://sorteo.cermep.fr/home.php) [^{18}F]fludeoxyglucose (FDG) simulation were used to find a realistic count rate in grey matter and white matter of the corresponding transverse slice.

In order to investigate the performance of the proposed method in regions where the PET and MR images do not agree, three lesions were embedded in the PET ground truth. More specifically, two hot lesions with 50% activity increase relative to their background (WM and GM) were embedded in the ground truth. The area of the hot lesion in the GM was 404 pixels and the area of the lesion in WM was 505 pixels. One cold lesion with 70% activity decrease relative to its background (GM) and total area of 395 pixels was also added to the image. The exact known boundaries of the lesions were used for calculation of the CRC values. The MR slice used for dictionary learning as well as the ground truth PET with embedded lesions are shown in figure 5.

The ground truth was then forward projected using a 2D scanner model with 256 radial bins and 288 azimuthal angles. Linear attenuation coefficients equal to $0.099\,\text{cm}^{-1}$ for GM and WM, $0.095\,\text{cm}^{-1}$ for fat, CSF, skin and connective tissue, $0.14\,\text{cm}^{-1}$ for skull and $0\,\text{cm}^{-1}$ for the air were also modelled in the forward projection. A heavily smoothed sinogram with total activity equal to 25 percent of the forward projected true image was added to the forward projected image to provide a simplified simulation of scatter and randoms estimates. Poisson noise was then introduced to generate 20 realizations each having an expected total number of events equal to 300 000. Corrections for attenuation, scatter and random events are performed using their exact mean values in the reconstruction algorithm. All images were represented by 256×256 pixels of side length $1.219\,\text{mm}$.
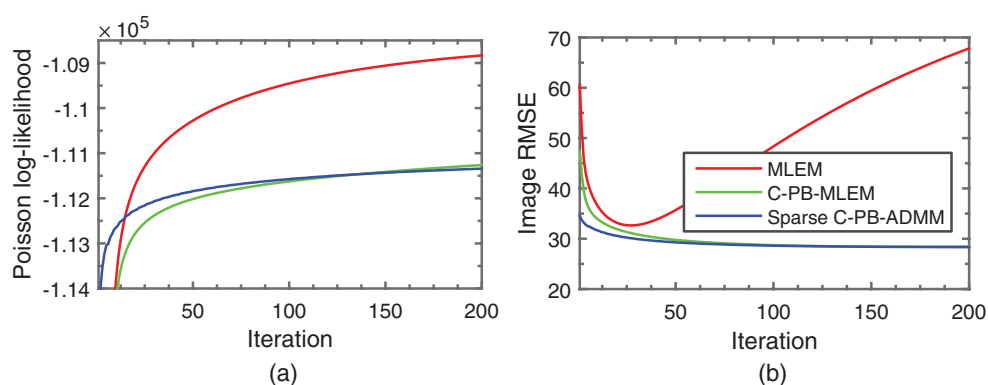
Figure 4. The log-likelihood of the estimated parameters given a noisy realization of the 2D phantom sinogram data for MLEM, C-PB-MLEM and sparse C-PB-ADMM as a function of iteration. (b) The RMSE of the estimated parameters as a function of iteration.

The same T1-weighted MR image used to generate the PET data was used to extract basis functions. Before that, MINC tools (https://github.com/BIC-MNI) were utilized in order to obtain tissue segmentation.

4.1.2. Results. The proposed methods were tested for different patch sizes. The patch size of 6 by 6 produced images with lowest total RMSE and thus was used in our experiments.

Figure 4 left shows the log likelihood of the parameters of interest given a noisy simulation of sinogram data as a function of iteration. For MLEM the parameters of interest are the image values x whereas for the C-PB-MLEM and sparse C-PB-ADMM methods the parameters of interest are the coefficients in θ. The number of clusters and hence the number of dictionaries learned from the modified MR image is 15. The β parameter for sparse C-PB-ADMM is set to 0.003.

This figure shows that the convergence rate of sparse C-PB-ADMM is higher than that of C-PB-MLEM. Once the dictionary is learned, the time needed to perform 50 iterations of C-PB-MLEM, C-PB-ADMM and MLEM on a machine with two 2.80 GHz 10-core processors (Intel Xenon 2.80) and 64 GB system memory is 7, 9.3 and 2.1 min respectively. Note that even though the computational cost of each iteration of sparse C-PB-ADMM is higher compared to C-PB-MLEM because of its faster convergence rate, fewer iterations are needed to achieve the minimum RMSE value (as shown in figure 4 right)

In this work, the performance of the proposed methods is compared to Q-MAP, RD-MAP and sparse PB-MAPEM methods. n-RMSE and mean CRC values are computed across 20 realizations. Figure 5 shows the effect of different parameters on these methods. To enhance readability, only a few of the tested parameters including the ones leading to lowest mean pixel-based n-RMSE within the brain are shown in the figure.

Figure 5 top left shows the effect of the σ value of a 7 by 7 truncated Gaussian kernel used for weight calculation and the β penalty parameters in Q-MAP. In order to tune the parameters, the σ value was ranged from 0.05 to 3 and the β value was varied from 0.01 to 5.

Figure 5 top right shows the effect of β and γ parameters for RD-MAP method. Both β and γ parameter were varied from 0.2 to 2. Figure 5 middle row shows the effect of λ and ϵ (error tolerance for the OMP method) parameters used in sparse PB-MAPEM on the accuracy of the reconstructed images. For sparse PB-MAPEM, as suggested by Chen *et al* (2015), the number
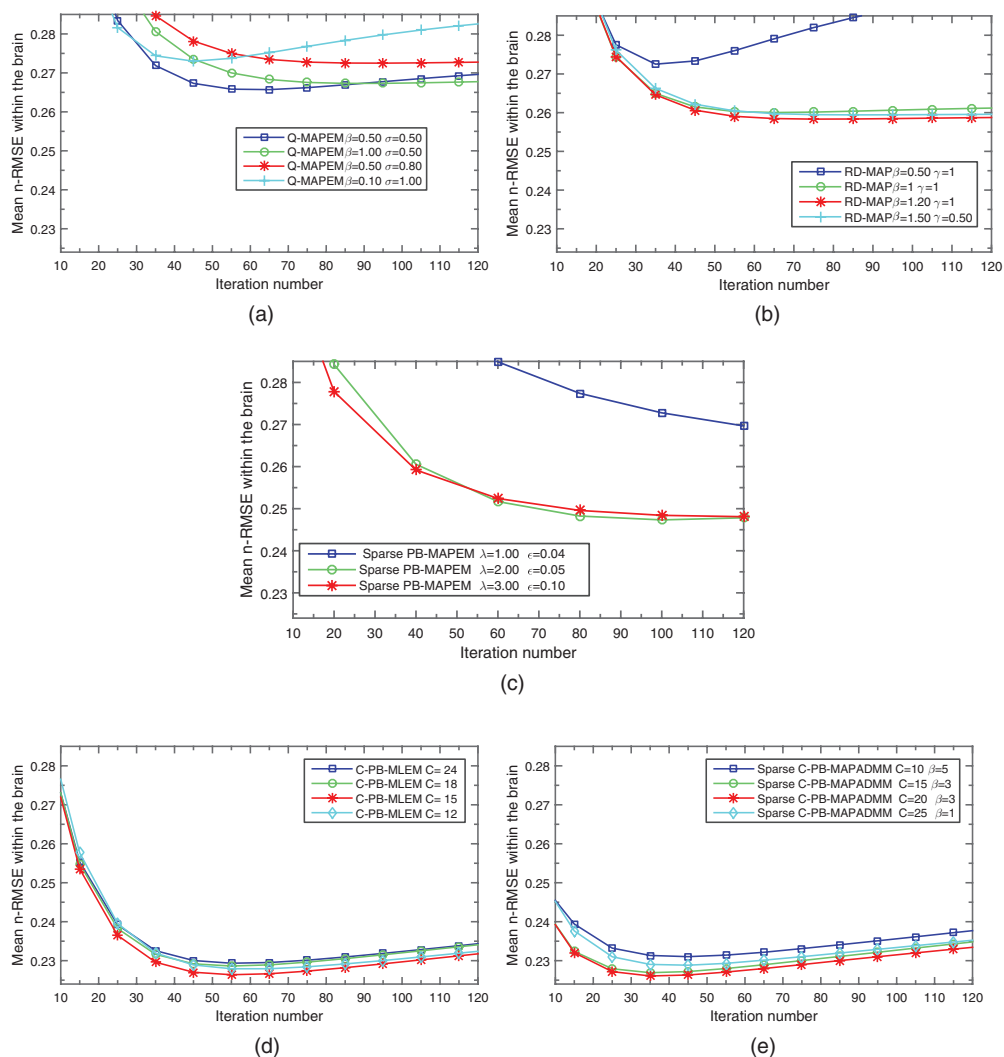
Figure 5. Mean pixel-based n-RMSE as a function of iteration within the brain for different parameters of five different methods: (a) Q-MAP, (b) RD-MAP, (c) sparse PB-MAPEM, (d) C-PB-MLEM and (d) sparse C-PB-ADMM.

of dictionary atoms is set to 144 and the patch size is set to 7 b 7. The λ value was tested for values of 1, 1.5, 2,2.5 and 3. The ϵ parameter was varied from 0.01 to 0.5.

In the bottom row, the right figure shows the effect of the number of clusters on the C-PB-MLEM algorithm. The left figure shows the influence of the number of clusters and the sparsity parameter β on the accuracy of the sparse-PB-ADMM method.

For both C-PB-MLEM and sparse C-PB-ADMM, the number of dictionary atoms learned per cluster is set according to the equation (38). As mentioned earlier, the sparsity parameter in the NMF-l_0 algorithm was set to 1/10th of the number of atoms learned per cluster. Both methods were tested for number of clusters C equal to 5, 10, 15, 20 and 25. The tested β value for sparse C-PB-ADMM ranged from 0.001 to 0.05.

Figure 6 shows the reconstruction results of a noisy realization using MLEM after 20 iterations (leading to minimum mean n-RMSE within the brain), Q-MAP, RD-MAP, sparse
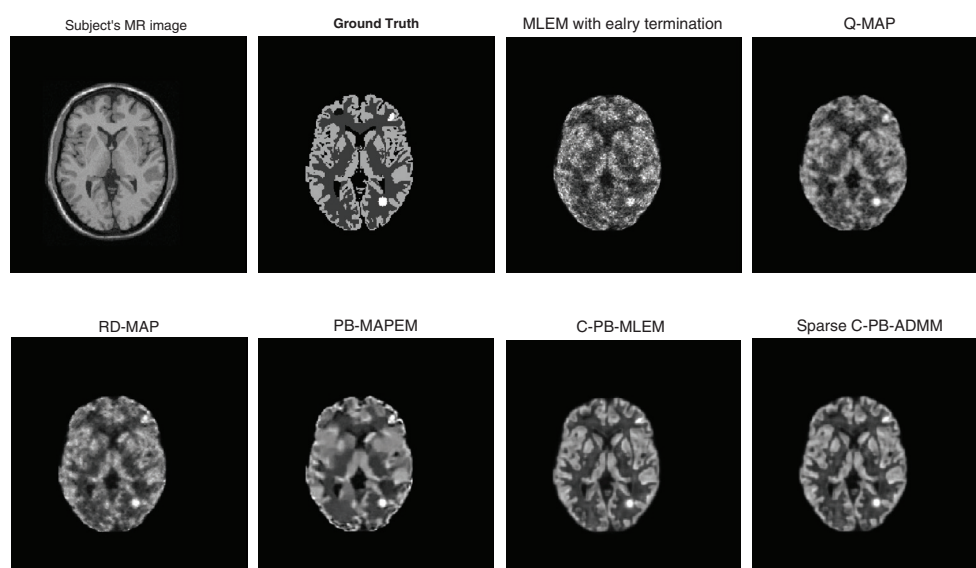
Figure 6. Sample reconstructed images using different methods at their best iteration and with tuned parameters.
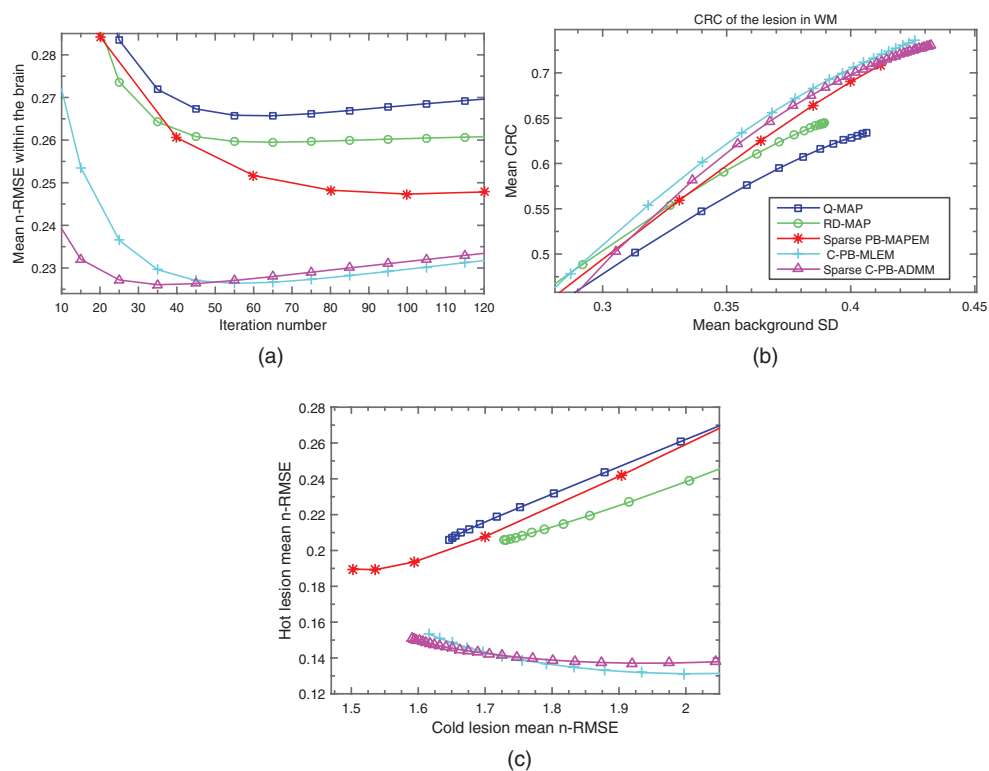
PB-MAPEM, the proposed C-PB-MLEM and sparse C-PB-ADMM methods with their best parameters and at their best iteration found based on the lowest mean pixel-based n-RMSE within the brain. We can see that using the proposed methods substantially improved the image quality over the conventional MLEM, Q-MAP, RD-MAP and sparse PB-MAPEM.

Figure 7, shows the performance of different regularization methods with their best parameters based on the minimum mean pixel-based n-RMSE within the brain. The top left figure shows the mean pixel-based n-RMSE within the brain as a function of iteration. The top right figure shows the mean CRC of the lesion in white matter versus background noise across multiple realizations. The bottom figure shows the mean pixel-based n-RMSE within the hot lesion in the grey matter versus the mean pixel-based n-RMSE within the cold lesion in the grey matter.

The results indicate that both C-PB-MLEM and sparse C-PB-ADMM outperform the other regularization techniques. Sparse C-PB-ADMM gives a slightly lower best mean pixel-based n-RMSE value within the brain at the best iteration number compared to C-PB-MLEM, while the opposite is the case for mean CRC of the hot lesion in white matter. The mean pixel-based n-RMSE of the hot versus cold lesion shows that sparse PB-MAPEM provides a lower error within the cold lesion while the error in hot lesion is higher than that of C-PB-MLEM. Sparse C-PB-ADMM seems to provide a slightly better n-RMSE in the hot versus cold lesion compared C-PB-MLEM method. In summary C-PB-MLEM and sparse C-PB-ADMM seem to have a similar performance.

4.2. 3D phantom

4.2.1. Setup. In order to explore the extendibility of the proposed method to 3D data, the same T1-weighted MR image used in the previous section is utilized. This time however, instead of one slice, the whole 3D volume is used to generate a 3D FDG phantom. Four lesions are simulated in the FDG PET ground truth (shown in figure 8). Lesions 2, 3 and 4 are embedded in the GM and follow the cortical boundaries while lesion 1 is embedded in the WM. In

(a)

(b)

(c)

Figure 7. Different figures of merit for Q-MAP, RD-MAP, sparse PB-MAPEM, C-PB-MLEM and sparse C-PB-ADMM with their best parameters, leading to minimum n-RMSE within the brain. (a) Mean n-RMSE within the brain as a function of iteration. (b) Mean CRC of the lesion in WM as a function of mean standard deviation in WM. (c) Mean n-RMSE of the hot lesion in GM versus mean n-RMSE of the cold lesion in GM. Mean pixel-based figures of merit for Q-MAP, RD-MAP, sparse PB-MAPEM, C-PB-MLEM and sparse C-PB-ADMM with their best parameters.
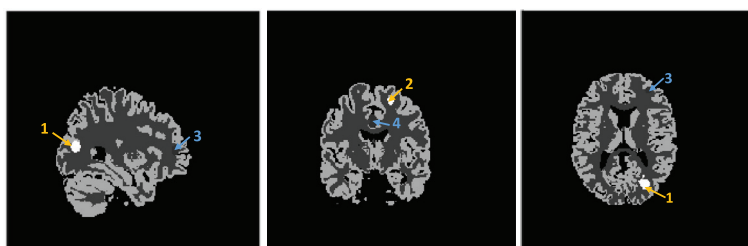
Figure 8. 3D FDG phantom with 4 embedded lesions.

lesions 1 and 2 the activity is increased by 50% relative to their background (WM and GM respectively). Lesions 3 and 4 have 70% activity reduction relative to the activity in the GM. The total volumes of lesions 1, 2, 3 and 4 are 515, 260, 283 and 553 voxels respectively.

The same procedure described in the previous section but in 3D is used to generate highly noisy sinogram data of this 3D phantom with 10 million counts. The size of the image is 256 by 256 by 207 voxels of 1.219 mm length. The forward projection is based on the HRRT
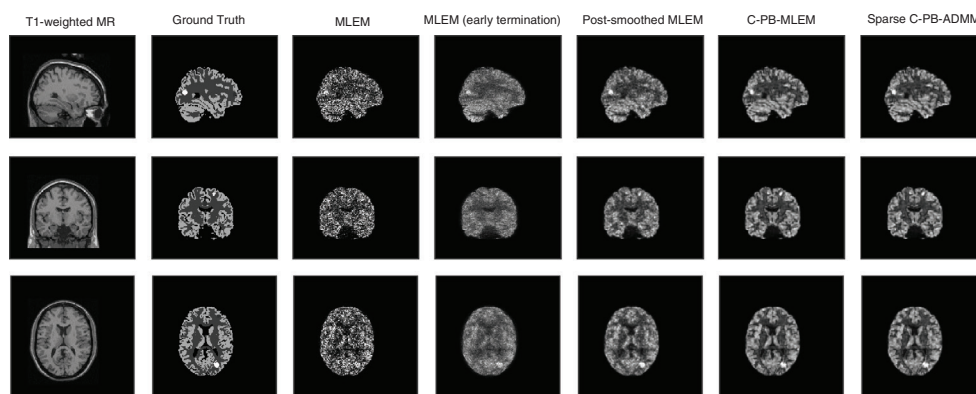
Figure 9. One example realization of the simulated 3D FDG data reconstructed using different methods shown for sagittal, coronal and transverse views. All methods are shown at their best iteration leading to minimum RMSE value.

scanner geometry consisting of 2209 sinograms with 288 view angles and 256 projection bins per view.

The patch size is set to 4 by 4 by 4 voxels. In order to decrease the number of patches extracted from the image and hence reduce the computational complexity, the distance between corresponding voxels in the neighbouring patches is set to two voxels.

Figure 9 shows the transverse, sagittal, and coronal views of the reconstructed image of the noisy simulation using different methods. MLEM is shown after 50 iterations as well as after 10 iterations leading to minimum RMSE. Post-smoothed MLEM with a Gaussian kernel of size 5 by 5 by 5 voxels is tested for different σ values ranging from 0.5 to 3 and number of iterations ranging from 10 to 80. The one leading to minimum RMSE is shown here (σ equal to 1 voxel at iteration 60). The number of clusters used for C-PB-MLEM and sparse C-PB-ADMM is 20 and 15 respectively. The number of atoms learned per cluster is found based on equation (38). C-PB-MLEM is shown after 35 iterations and sparse C-PB-ADMM is shown after 20 iterations. These iteration numbers were selected as they achieve minimum mean n-RMSE within the brain. Both C-PB-MLEM and sparse C-PB-ADMM significantly improve the image quality and provide a good contrast in the lesions.

To further evaluate the ability of the methods in improving the contrast in the lesions, the CRC values of the 4 lesions at the mentioned parameters and number of iterations (leading to minimum RMSE) are show in figure 10. For all four lesions the proposed methods outperform post-smoothed MLEM. Using sparse C-PB-ADMM leads to a greater CRC value compared to C-PB-MLEM for all lesions except for lesion 3, where the CRC value is slightly smaller.

5. Application to real subject data

In this section, the MLEM, C-PB-MLEM and sparse C-PB-ADMM methods are used to reconstruct images from real data. The HRRT list-mode data is obtained from a 370 MBq injection of [^{11}C]raclopride to a healthy human participant. The inter-frame motion was negligible and so not corrected for. Attenuation and normalization factors as well as scatter and randoms estimates obtained from the manufacturer's software were incorporated within the reconstruction. The duration of the reconstructed frame is 40 min starting 20 min after injection.

For patch-based basis vector extraction, the process described in the implementation section is used to obtain a registered re-sampled modified MR image from the subject's
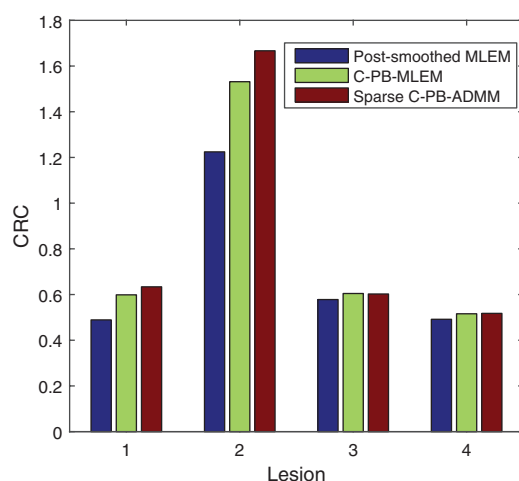
Figure 10. The CRC values of the 4 lesions indicated in figure 8 using post-smoothed MLEM after 60 iteration, C-PB-MLEM and sparse C-PB-ADMM. All methods are shown at their best iteration leading to minimum RMSE value.
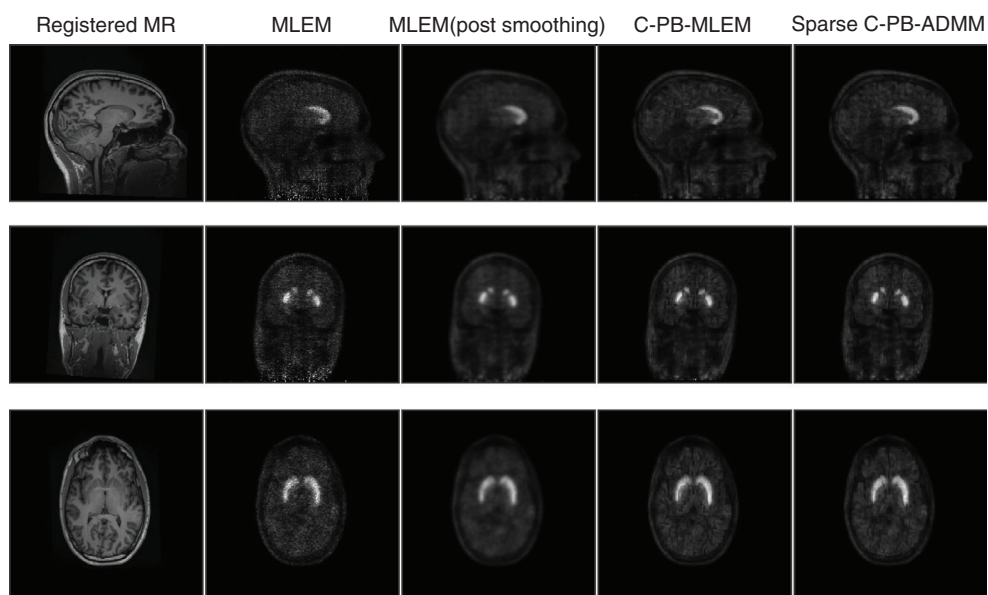
Figure 11. Reconstructed images of a 40 min frame [^{11}C]raclopride scan of the real subject by different algorithms. Each column shows the resulting slices of a 3D image of an algorithm in transverse, sagittal, and coronal views.

T1-weighted MR image. Then overlapping patches of size 4 by 4 by 4 with corresponding voxels in the neighbouring patches having a distance of 2 voxels, are extracted and used to obtain basis vectors. The number of clusters in C-PB-MLEM and C-PB-ADMM is set to 20 and 15 respectively. The β value for sparse C-PB-ADMM is set to 0.003.

Figure 11 shows the registered re-sampled T1-weighted MR image as well as reconstructed images of this data using MLEM, C-PB-MLEM and sparse C-PB-ADMM algorithms. MLEM
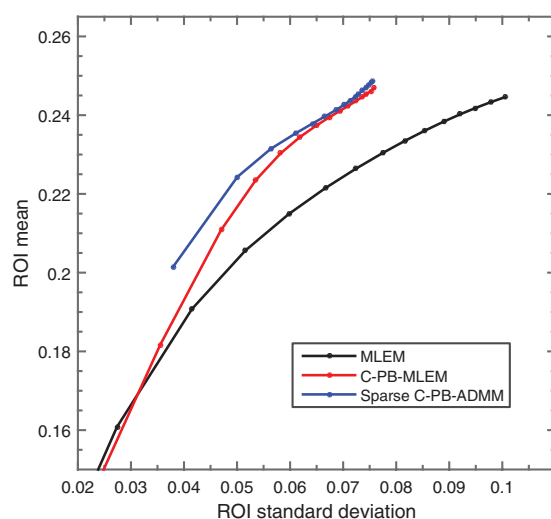
Figure 12. The mean activity versus the standard deviation in the right caudate in the real data reconstructed using MLEM and C-PB-MLEM and sparse C-PB-ADMM methods for increasing iterations.

is shown after 50 iterations. Post-smoothed MLEM is obtained by applying a Gaussian kernel of size 5 by 5 by 5 and σ equal to 1 voxel to the reconstructed image at iteration 50. The images reconstructed using C-PB-MLEM and saprse C-PB-ADMM are shown after 25 iterations. Figure 12 compares the mean activity in the right caudate excluding the edges versus standard deviation within this region with increasing iteration for MLEM, C-PB-MLEM and sparse C-PB-ADMM. One can note that both proposed methods can deliver a higher ROI mean within this region for every standard deviation.

6. Conclusion

In this work, a novel re-parameterization framework which uses patch-based basis functions learned from a registered prior image for PET image reconstruction is devised. In addition, a method for finding sparse coefficients for a re-parameterized Poisson log-likelihood objective function is proposed. The two approaches are then combined together to find sparse coefficients for patch-based basis functions. The proposed methods are validated by computer simulations. The results indicate that the proposed methods can achieve better quality images compared to the RD-MAP, Q-MAP and sparse PB-MAPEM approaches.

The core of this work is to use non-negative dictionaries extracted from prior images to parameterize the PET image reconstruction task. Although this work only looks into the use of T1-weighted MR images for basis vector extraction, once corrected for inverted intensities (if needed), any prior image with enough boundary information can be utilized in this framework. It is important to note that the proposed framework does permit flexibility, as demonstrated by the reconstruction of features not present in the MR (the lesions in the ground truth PET distribution) -so abnormal features in the radioactivity distribution can still be successfully reconstructed.

Note that there is a trade off between the amount of structural detail and the insensitivity of the method to misalignment. This trade off can be controlled by the number of clusters. In fact, in order to avoid sensitivity to mis-registration, one can decrease the number of clusters while increasing the number of atoms in each cluster. In the extreme case, an identical dictionary

learned from the entire image can be utilized to represent each and every patch in the image at the expense of increased variance. In the other extreme, we can partition the prior image patches into many clusters; leading to many dictionaries each with few atoms. This will result in low variance at the expense of increase in bias and becoming sensitive to slight mis-registrations.

Recently, advanced non-linear registration techniques have been used to construct 4D PET atlases (Bieth *et al* 2013). A future study investigating the use of a PET atlas for basis extraction and reconstructing PET images of the same radiotracer would be of great interest.

Acknowledgment

Appendix. Alternate direction method of multipliers

The alternate direction method of multipliers (ADMM) is a powerful optimization method introduced by Glowinski and Marroco (1975) that has recently gained attention in many image processing applications such as image denoising and restoration, e.g. Afonso *et al* (2010), Figueiredo and Bioucas-Dias (2010) and Matakos *et al* (2013). Through the variable splitting method, ADMM allows the optimization of the sum of two or more functionals to be done separately in an alternating manner. For example let us consider $f(\mathbf{x}) + g(\mathbf{Ax})$ as the objective function to be optimized. By introducing an auxiliary variable $\mathbf{z} = \mathbf{Ax}$ the problem can be written as a constrained minimization:

$$
\begin{aligned}
&\min \quad f(\mathbf{x}) + g(\mathbf{z}) \\
&\text{such that} \quad \mathbf{Ax} - \mathbf{z} = 0
\end{aligned}
\tag{A.1}
$$

ADMM uses the augmented Lagrangian method to relax the above problem to :

$$
L_p(\mathbf{x}, \mathbf{z}, \mathbf{u}) = f(\mathbf{x}) + g(\mathbf{z}) + \mu^T(\mathbf{Ax} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{Ax} - \mathbf{z}\|^2
\tag{A.2}
$$

where μ is the Lagrangian multiplier and $\rho > 0$ is the penalty parameter. ADMM then solves the above optimization problem in an iterative alternating manner as follows:

$$
\mathbf{x}^{k+1} = \arg \min_{\mathbf{x}} L_p(\mathbf{x}, \mathbf{z}^k, \mu^k)
\tag{A.3}
$$

$$
\mathbf{z}^{k+1} = \arg \min_{\mathbf{z}} L_p(\mathbf{x}^{k+1}, \mathbf{z}, \mu^k)
\tag{A.4}
$$

$$
\mu^{k+1} = \mu^k + \rho(\mathbf{Ax}^{k+1} - \mathbf{z}^{k+1})
\tag{A.5}
$$

For a thorough review of the ADMM method see Boyd *et al* (2011).

References

Afonso M V, Bioucas-Dias J M and Figueiredo M A 2010 Fast image recovery using variable splitting and constrained optimization *IEEE Trans. Image Process.* **19** 2345–56

Ahn S *et al* 2015 Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET *Phys. Med. Biol.* **60** 5733

Ardekani B A *et al* 1996 Minimum cross-entropy reconstruction of PET images using prior anatomical information *Phys. Med. Biol.* **41** 2497

Bieth M *et al* 2013 Atlas construction for dynamic (4D) PET using diffeomorphic transformations *Medical Image Computing and Computer-Assisted Intervention* (New York: Springer) pp 35–42

Bowsher J E *et al* 1996 Bayesian reconstruction and use of anatomical *a priori* information for emission tomography *IEEE Trans. Medical Imaging* **15** 673–86

Boyd S *et al* 2011 Distributed optimization and statistical learning via the alternating direction method of multipliers *Found. Trends Mach. Learn.* **3** 1–122

Buades A, Coll B and Morel J-M 2005 A review of image denoising algorithms, with a new one *Multiscale Model. Simul.* **4** 490–530

Chatterjee P and Milanfar P 2009 Clustering-based denoising with locally learned dictionaries *IEEE Trans. Image Process.* **18** 1438–51

Chatterjee P and Milanfar P 2012 Patch-based near-optimal image denoising *IEEE Trans. Image Process.* **21** 1635–49

Chen S *et al* 2015 Sparse representation and dictionary learning penalized image reconstruction for positron emission tomography *Phys. Med. Biol.* **60** 807

Cheng-Liao J and Qi J 2011 PET image reconstruction with anatomical edge guided level set prior *Phys. Med. Biol.* **56** 6899

Dabov K *et al* 2007 Image denoising by sparse 3D transform-domain collaborative filtering *IEEE Trans. Image Process.* **16** 2080–95

Daubechies I, Defrise M and De Mol C 2004 An iterative thresholding algorithm for linear inverse problems with a sparsity constraint *Commun. Pure Appl. Math.* **57** 1413–57

De Pierro A R 1995 A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography *IEEE Trans. Med. Imaging* **14** 132

Fessler J A 1997 Grouped coordinate descent algorithms for robust edge-preserving image restoration *Proc. SPIE* **3071** 184–94

Figueiredo M A T and Bioucas-Dias J M 2010 Restoration of poissonian images using alternating direction optimization *IEEE Trans. Image Process.* **19** 3133–45

Glowinski R and Marroco A 1975 Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires *ESAIM: Math. Modelling Numer. Anal.—Modélisation Math. Anal. Numér.* **9** 41–76

Green P J 1990 On use of the em for penalized likelihood estimation *J. R. Stat. Soc.* B **52** 443–52

Hoyer P O 2004 Non-negative matrix factorization with sparseness constraints *J. Mach. Learn. Res.* **5** 1457–69

Jacobs F, Matej S and Lewitt R 1998 Image reconstruction techniques for PET *Technical Report* Department of Radiology, University of Pennsylvania, Philadelphia MIPG245

Jiao J *et al* 2015 Detail-preserving pet reconstruction with sparse image representation and anatomical priors *Information Processing in Medical Imaging* (New York: Springer) pp 540–51

Kervrann C and Boulanger J 2006 Optimal spatial adaptation for patch-based image denoising *IEEE Trans. Image Process.* **15** 2866–78

Lewitt R M 1990 Multidimensional digital image representations using generalized kaiser–bessel window functions *J. Opt. Soc. Am.* A **7** 1834–46

Lewitt R M 1992 Alternatives to voxels for image representation in iterative reconstruction algorithms *Phys. Med. Biol.* **37** 705

Matakos A, Ramani S and Fessler J A 2013 Accelerated edge-preserving image restoration without boundary artifacts *IEEE Trans. Image Process.* **22** 2019

Novosad P and Reader A J 2016 MR-guided dynamic PET reconstruction with the kernel method and spectral temporal basis functions *Phys. Med. Biol.* **61** 4624

Nuyts J *et al* 2003 A penalized-likelihood image reconstruction method for emission tomography, compared to postsmoothed maximum-likelihood with matched spatial resolution *IEEE Trans. Med. Imaging* **22** 1042–52

Nuyts J, Bequé D, Dupont P and Mortelmans L 2002 A concave prior penalizing relative differences for maximum-*a posteriori* reconstruction in emission tomography *IEEE Trans. Nucl. Sci.* **49** 56–60

Peharz R and Pernkopf F 2012 Sparse nonnegative matrix factorization with l0-constraints *Neurocomputing* **80** 38–46

Rangarajan A, Hsiao T and Gindi G 2000 A Bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction *J. Math. Imaging Vis.* **12** 199–217

Reader A J and Verhaeghe J 2014 4D image reconstruction for emission tomography *Phys. Med. Biol.* **59** R371

Reilhac A *et al* 2006 Creation and application of a simulated database of dynamic [18F] MPPF PET acquisitions incorporating inter-individual anatomical and biological variability *IEEE Trans. Med. Imaging* **25** 1431–9

Somayajula S, Asma E and Leahy R M 2005 PET image reconstruction using anatomical information through mutual information based priors *IEEE Nuclear Science Symp. Conf. Record* vol 5 pp 2722–6

Tahaei M S and Reader A J 2014 Combining different variance reduction approaches for PET image reconstruction *IEEE Nuclear Science Symp. and Med. Imaging Conf.* pp 1–6

Tang J and Rahmim A 2009 Bayesian PET image reconstruction incorporating anato-functional joint entropy *Phys. Med. Biol.* **54** 7063

Tang J *et al* 2014 Sparsity-based pet image reconstruction using MRI learned dictionaries *IEEE 11th Int. Symp. on Biomedical Imaging* pp 1087–90

Tropp J *et al* 2007 Signal recovery from random measurements via orthogonal matching pursuit *IEEE Trans. Inf. Theory* **53** 4655–66

Wang G and Qi J 2012 Penalized likelihood pet image reconstruction using patch-based edge-preserving regularization *IEEE Trans Med. Imaging* **31** 2194–204

Wang G and Qi J 2015 PET image reconstruction using kernel method *IEEE Trans. Med. Imaging* **34** 61–71

Wells R G *et al* 2000 Comparing filtered backprojection and ordered-subsets expectation maximization for small-lesion detection and localization in Ga-67 SPECT *J. Nucl. Med.* **41** 1391–9