Maximum-likelihood joint image reconstruction and motion estimation with misaligned attenuation in TOF-PET/CT

Fast Track Communication

Maximum-likelihood joint image reconstruction and motion estimation with misaligned attenuation in TOF-PET/CT

Alexandre Bousse[1], Ottavia Bertolli[1], David Atkinson[2], Simon Arridge[3], Sébastien Ourselin[4], Brian F Hutton[1,5] and Kris Thielemans[1]

[1] Institute of Nuclear Medicine, University College London, London NW1 2BU, UK
[2] Centre for Medical Imaging, University College London, London NW1 2PG, UK
[3] Department of Computer Science, University College London, London WC1E 6BT, UK
[4] Centre for Medical Image Computing, University College London, London WC1E 6BT, UK
[5] Centre for Medical Radiation Physics, University of Wollongong, Wollongong NSW 2522, Australia

E-mail: a.bousse@ucl.ac.uk

CrossMark

Abstract

This work is an extension of our recent work on joint activity reconstruction/motion estimation (JRM) from positron emission tomography (PET) data. We performed JRM by maximization of the penalized log-likelihood in which the probabilistic model assumes that the same motion field affects both the activity distribution and the attenuation map. Our previous results showed that JRM can successfully reconstruct the activity distribution when the attenuation map is misaligned with the PET data, but converges slowly due to the significant cross-talk in the likelihood. In this paper, we utilize time-of-flight PET for JRM and demonstrate that the convergence speed is significantly improved compared to JRM with conventional PET data.

Keywords: attenuation correction, time-of-flight PET, motion estimation, attenuation mismatch, maximum-likelihood

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

1. Introduction

Patient motion in positron emission tomography (PET) is a source of error due to possible mismatches between the PET data and the computed tomography (CT) attenuation map (μ-map) (Nyflot *et al* 2015). In Bousse *et al* (2016) we developed a motion compensated reconstruction scheme for gated PET data, namely joint reconstruction/motion estimation (JRM), to jointly estimate both the activity distribution and the motion field, by penalized likelihood maximization. Unlike previous works in this field (Jacobson and Fessler 2003, Blume *et al* 2010), our model assumed that both the activity distribution and the μ-map are affected by motion. This model led to the following result: the JRM-reconstructed PET gates are the same for any input μ-maps derived from deformations of a common μ-map. This is because the estimated motion automatically accounts for PET/μ-map misalignment. However, in case of large mismatches, our results in Bousse *et al* (2016) showed that JRM needs a high number of iterations due to the significant cross-talk in the joint-likelihood.

Recent work Rezaei *et al* (2012), (2014) demonstrated the potential of jointly reconstructing the activity distribution and the μ-map from time-of-flight (TOF) PET data. Since JRM with warped μ-map presents some similarities with joint activity/attenuation reconstruction from emission data, in the sense that the warped μ-map must match the PET projections at each gate, TOF PET can increase the JRM convergence rate—especially in situations where the input μ-map is misaligned with the PET data.

In this work we investigate the ability of JRM with TOF PET to deal with a misaligned μ-map. Section 2 presents the JRM optimization problem for TOF PET and non-TOF PET, and summarizes the algorithm. In section 3 we compare the convergence rate of JRM with TOF PET and non-TOF PET on a simulated end-expiration PET gate and a simulated end-inhalation μ-map. Results are discussed in section 4.

2. Method

2.1. TOF maximum-likelihood for joint image reconstruction and motion estimation

In this section we use notations similar to Bousse *et al* (2016, sections II and III).

2.1.1. Motion-free model.
The activity distribution and the attenuation map respectively take the form of 2 functions $f \in \mathcal{C}^+$ and $\mu \in \mathcal{C}^+$, where $\mathcal{C}^+$ denotes the set of non-negative continuous functions defined on $\mathbb{R}^3$. The attenuation map μ is reconstructed independently from separate measurements such as x-ray CT or segmented MRI. TOF measured counts are represented by a collection $\boldsymbol{g} = (g_{i,t})_{i,t=1}^{n_b,n_t} \in \mathbb{N}^{n_b \times n_t}$. The subscripts $i \in \{1, \ldots, n_b\}$ and $t \in \{1, \ldots, n_t\}$ are respectively the detector and the time bin indices. In absence of motion, $g_{i,t}$ follows a Poisson distribution of expectation $\bar{g}_{i,t}(f, \mu)$:

$$\bar{g}_{i,t}(f, \mu) = \tau a_i(\mu) \mathcal{H}_{i,t} f + s_{i,t},$$

with

$$\mathcal{H}_{i,t} f \triangleq \int_{\Omega} f(\boldsymbol{r}) h_{i,t}(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} \quad \text{and} \quad a_i(\mu) \triangleq \exp\left(-\int_{L_i} \mu(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r}\right) \tag{1}$$

where $h_{i,t} : \mathbb{R}^3 \to \mathbb{R}^+$ is the TOF PET system response function at detector/time bin (i, t), τ is the scanning time, L_i is the segment connecting the detectors of bin i, $s_{i,t}$ is the expected background events (random/scatter) at bin (i, t) and $\Omega \subset \mathbb{R}^3$ is a compact set representing the field of view.

2.1.2. Model with motion.   In practice, f and μ are affected by patient motion. For quasi-cyclic motion (respiratory, cardiac), acquired emission data are regrouped into n_g gates, each of which corresponds to a patient state. Each TOF-PET data vector $\boldsymbol{g}_l = (g_{i,t,l})_{i,t=1}^{n_b, n_t} \in \mathbb{N}^{n_b \times n_t}$, $l = 1, \ldots, n_g$, corresponds to a single gate in which the patient is assumed to be static. The motion at gate l is represented by a diffeomorphism $\varphi_l : \mathbb{R}^3 \to \mathbb{R}^3$ that deforms both the activity distribution f and the attenuation map μ. The deformed activities and μ-maps are $\mathcal{W}_{\varphi_l} f$ and $\mathcal{W}_{\varphi_l} \mu$, where $\mathcal{W}_\varphi : \mathcal{C}^+ \to \mathcal{C}^+$ is the warping operator defined as $\mathcal{W}_\varphi h = h \circ \varphi$ for all $h \in \mathcal{C}^+$. The number of detected counts $g_{i,t,l}$ at bin (i, t), gate l is a Poisson random variable of expectation

$$\bar{g}_{i,t,l}(f, \varphi_l, \mu) \triangleq \eta a_i(\mathcal{W}_{\varphi_l}\mu)\mathcal{H}_{i,t}\mathcal{W}_{\varphi_l}f + s_{i,t,l}, \tag{2}$$

where $s_{i,t,l}$ is the expected number of background events at bin (i, t), gate l, η is the duration of gate l and a_i and $\mathcal{H}_{i,t}$ are defined in (1).

2.1.3. Optimization problem.   The log-likelihood of the TOF-PET data is

$$L^{\mathrm{TOF}}(f, \boldsymbol{\varphi}, \mu) = \sum_{l=1}^{n_g} \sum_{i=1}^{n_b} \sum_{t=1}^{n_t} g_{i,t,l} \log \bar{g}_{i,t,l}(f, \varphi_l, \mu) - \bar{g}_{i,t,l}(f, \varphi_l, \mu), \tag{3}$$

where $\boldsymbol{\varphi} \triangleq (\varphi_l)_{l=1}^{n_g}$. Similarly, the non-TOF PET log-likelihood is

$$L^{\mathrm{PET}}(f, \boldsymbol{\varphi}, \mu) = \sum_{l=1}^{n_g} \sum_{i=1}^{n_b} g_{i,l} \log \bar{g}_{i,l}(f, \varphi_l, \mu) - \bar{g}_{i,l}(f, \varphi_l, \mu),$$

with $g_{i,l} = \sum_{t=1}^{n_t} g_{i,t,l}$ and $\bar{g}_{i,l}(f, \varphi_l, \mu) = \sum_{t=1}^{n_t} \bar{g}_{i,t,l}(f, \varphi_l, \mu)$.

Maximum-likelihood joint image reconstruction/motion estimation (JRM) in TOF PET and non-TOP PET consists of solving the following optimization problem

$$(\hat{f}, \hat{\boldsymbol{\varphi}}) = \underset{f \in \mathcal{C}^+, \boldsymbol{\varphi} \in \mathcal{D}}{\arg \max} L(f, \boldsymbol{\varphi}, \mu) \tag{4}$$

where L is either L^{TOF} or L^{PET}, $\mathcal{D} = \mathcal{D}^{n_g}$ and $\mathcal{D}$ denotes the set of diffeorphism on $\mathbb{R}^3$. Solving (4) returns a single image f reconstructed from all gates $\boldsymbol{g}_l$, $l = 1, \ldots, n_g$. The reconstructed activity image at gate l is $\mathcal{W}_{\hat{\varphi}_l}\hat{f} = \hat{f} \circ \hat{\varphi}_l$.

It should be noted that (2) and therefore (3) depend on f and μ only via $\mathcal{W}_{\varphi_l}f$ and $\mathcal{W}_{\varphi_l}\mu$. The activity distribution f corresponds to an unobserved state consistent with μ (but not necessarily with any $\boldsymbol{g}_l$). This led to a theoretical result for JRM in non-TOF PET in Bousse *et al* (2016, proposition 1), that can be naturally extended to TOF PET: if μ_1 and μ_2 are 2 different attenuation maps such that $\mu_2 = \mu_1 \circ \psi$ for some $\psi \in \mathcal{D}$, then there exists a bijection between the maximizers (when they exist) of $(f, \boldsymbol{\varphi}) \mapsto L(f, \boldsymbol{\varphi}, \mu_1)$ and the maximizers of $(f, \boldsymbol{\varphi}) \mapsto L(f, \boldsymbol{\varphi}, \mu_2)$ such that each pair of maximizers $(\hat{f}_1, \hat{\boldsymbol{\varphi}}^1)$–$(\hat{f}_2, \hat{\boldsymbol{\varphi}}^2)$, defined by this bijection, satisfies $\mathcal{W}_{\hat{\varphi}_l^1}\hat{f}_1 = \mathcal{W}_{\hat{\varphi}_l^2}\hat{f}_2$ and $\mathcal{W}_{\hat{\varphi}_l^1}\mu_1 = \mathcal{W}_{\hat{\varphi}_l^2}\mu_2$.

Proposition 1 in Bousse *et al* (2016) does not claim the existence and uniqueness of a maximizer for $(f, \boldsymbol{\varphi}) \mapsto L(f, \boldsymbol{\varphi}, \mu)$. However, if $(\hat{f}_1, \hat{\boldsymbol{\varphi}}^1)$ is a likely candidate to maximize $(f, \boldsymbol{\varphi}) \mapsto L(f, \boldsymbol{\varphi}, \mu_1)$, then *there exists* $(\hat{f}_2, \hat{\boldsymbol{\varphi}}^2) \in \mathcal{C}^+ \times \mathcal{D}$ that is an equally likely candidate to maximize $(f, \boldsymbol{\varphi}) \mapsto L(f, \boldsymbol{\varphi}, \mu_2)$ such that $\mathcal{W}_{\hat{\varphi}_l^1}\hat{f}_1 = \mathcal{W}_{\hat{\varphi}_l^2}\hat{f}_2$ and $\mathcal{W}_{\hat{\varphi}_l^1}\mu_1 = \mathcal{W}_{\hat{\varphi}_l^2}\mu_2$. Results from Bousse *et al* (2016) demonstrated that JRM with μ_1 and $\mu_2 = \mu_1 \circ \psi$ returns similar reconstructed gates, and more specifically, JRM can be performed with a μ-map misaligned with each gate, provided this misaligned results from some diffeomorphism. Moreover, the estimated motion realigns the μ-maps to the data at each gate l.

When μ is aligned with one gate $\boldsymbol{g}_{l_0}$, f can be initialized by maximizing the log-likelihood at gate l_0 with $\varphi_{l_0} = \mathrm{Id}$. If μ is misaligned with every gate, there is no consistent data to obtain a good initial estimate of f, thus rendering the optimization problem (4) more difficult. For example, in non-TOF PET (see (Bousse *et al* (2016), section IV.B)), it takes 50 to 100 iterations to solve (4) with a misaligned μ-map. The objective of this work is to see if TOF PET can facilitate the optimization.

2.2. Joint reconstruction scheme

2.2.1. Discretization. We used the same discretization scheme as in Bousse *et al* (2016), initially introduced in Jacobson and Fessler (2003) and Jacobson (2006). $M_{n,m}(\mathbb{R})$ denotes the space of real $n \times m$ matrices. The activity distribution and attenuation map are respectively represented by $\boldsymbol{f} \in \mathbb{R}_+^{n_{\mathrm{v}}}$ and $\boldsymbol{\mu} \in \mathbb{R}_+^{n_{\mathrm{v}}}$, where n_{v} denotes the total number of voxels. A deformation φ is parametrized by a B-spline coefficient vector $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y, \boldsymbol{\alpha}^Z) \in \mathbb{R}^{3 \times n_{\mathrm{c}}}$, where n_{c} denotes the number of B-spline control points. The warping operator $\mathcal{W}_{\varphi}$ becomes a square matrix $\boldsymbol{W}_{\boldsymbol{\alpha}} \in M_{n_{\mathrm{v}}, n_{\mathrm{v}}}(\mathbb{R})$, defined by the voxel interpolating functions, the B-spline coefficients $\boldsymbol{\alpha}$ and the B-spline basis. The B-spline coefficient corresponding to the motion φ_l at gate l is denoted $\boldsymbol{\alpha}_l$. The entire collection of B-spline coefficients is denoted $\boldsymbol{\theta} = (\boldsymbol{\alpha}_l)_{l=1}^{n_{\mathrm{g}}}$.

The TOF operator $\mathcal{H}_{i,t}$ takes the form of a system matrix $\boldsymbol{H} \in M_{n_{\mathrm{b}} \times n_{\mathrm{t}}, n_{\mathrm{v}}}(\mathbb{R})$ where $[\boldsymbol{H}]_{(i-1)n_{\mathrm{t}}+t, j} \triangleq p_{i,j,t}$ is the probability that an annihilation occurring at voxel j is detected in bin (i, t). The expected number of counts is redefined as

$$\bar{g}_{i,t,l}(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu}) \triangleq \tau a_i(\boldsymbol{W}_{\boldsymbol{\alpha}_l}\boldsymbol{\mu})[\boldsymbol{H}\boldsymbol{W}_{\boldsymbol{\alpha}_l}\boldsymbol{f}]_{(i-1)n_{\mathrm{t}}+t} + s_{i,t,l}, \tag{5}$$

where $a_i(\boldsymbol{\mu}) \triangleq \exp(-[\boldsymbol{L}\boldsymbol{\mu}]_i)$ and the line integral matrix $\boldsymbol{L} \in M_{n_{\mathrm{b}}, n_{\mathrm{v}}}(\mathbb{R})$ is defined by $[\boldsymbol{L}]_{i,j} = \ell_{i,j}$ where $\ell_{i,j}$ is the length of the intersection of L_i with voxel j.

2.2.2. Optimization method. The discrete TOF log-likelihood is derived from (3):

$$L^{\mathrm{TOF}}(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu}) = \sum_{l=1}^{n_{\mathrm{g}}} \sum_{i=1}^{n_{\mathrm{b}}} \sum_{t=1}^{n_{\mathrm{t}}} g_{i,t,l} \log \bar{g}_{i,t,l}(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu}) - \bar{g}_{i,t,l}(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu}). \tag{6}$$

In comparison, the non-TOF PET discrete log-likelihood is

$$L^{\mathrm{PET}}(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu}) \triangleq \sum_{l=1}^{n_{\mathrm{g}}} \sum_{i=1}^{n_{\mathrm{b}}} g_{i,l} \log \bar{g}_{i,l}(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu}) - \bar{g}_{i,l}(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu}) \tag{7}$$

with $g_{i,l} = \sum_{t=1}^{n_{\mathrm{t}}} g_{i,t,l}$ and $\bar{g}_{i,l}(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu}) = \sum_{t=1}^{n_{\mathrm{t}}} \bar{g}_{i,t,l}(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu})$.

To enforce image and motion smoothness, 2 penalty terms $U(\boldsymbol{f})$ and $V(\boldsymbol{\theta})$ are added to L^{TOF} and L^{PET} (quadratic penalties, see Bousse *et al* (2016)). The discrete JRM formulation is formulated as

$$\left(\hat{\boldsymbol{f}}^{\mathrm{TOF}}, \hat{\boldsymbol{\theta}}^{\mathrm{TOF}}\right) = \arg\max_{\boldsymbol{f} \geqslant \mathbf{0}, \boldsymbol{\theta}} \Phi^{\mathrm{TOF}}(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu}) \tag{8}$$

$$\left(\hat{\boldsymbol{f}}^{\mathrm{PET}}, \hat{\boldsymbol{\theta}}^{\mathrm{PET}}\right) = \arg\max_{\boldsymbol{f} \geqslant \mathbf{0}, \boldsymbol{\theta}} \Phi^{\mathrm{PET}}(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu}) \tag{9}$$

with $\Phi^{\mathrm{TOF}}(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu}) = L^{\mathrm{TOF}}(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu}) - \beta U(\boldsymbol{f}) - \gamma V(\boldsymbol{\theta})$, $\Phi^{\mathrm{PET}}(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu}) = L^{\mathrm{PET}}(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu}) - \beta U(\boldsymbol{f}) - \gamma V(\boldsymbol{\theta})$. The JRM-reconstructed activity images corresponding to gate l are $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}_l^{\mathrm{TOF}}}\hat{\boldsymbol{f}}^{\mathrm{TOF}}$ for TOF PET and $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}_l^{\mathrm{PET}}}\hat{\boldsymbol{f}}^{\mathrm{PET}}$ for non-TOF PET.

We solve (8) and (9) by alternating maximization in $\boldsymbol{f}$ and $\boldsymbol{\theta}$. We proceed as in Bousse *et al* (2016): a complete iteration consists of a maximization in $\boldsymbol{\theta}$—performed with a limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) quasi-Newton (QN) line-search algorithm (Nocedal and Wright 2006, chapter 7)—and a maximization in $\boldsymbol{f}$—performed with a motion compensated (MC) modified maximum-likelihood expectation-maximization (MMLEM) (De Pierro 1995) reconstruction from the gated data $(\boldsymbol{g}_l)_{l=1}^{n_g}$ using the current estimated motion parameter $\boldsymbol{\theta}$.

The gradient of L^{TOF} in $\boldsymbol{\alpha}_l$ is similar to the PET case described in Bousse *et al* (2016) with a summation over the time index t:

$$\nabla_{\boldsymbol{\alpha}_l} L(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu}) = \sum_{t=1}^{n_t} \boldsymbol{J}_{\boldsymbol{\alpha}_l}(\bar{\boldsymbol{g}}_{:,t,l}(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu}))^{\top} \left(\frac{\boldsymbol{g}_{:,t,l}}{\bar{\boldsymbol{g}}_{:,t,l}(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu})} - \boldsymbol{1} \right)$$

with $\boldsymbol{A}(\boldsymbol{\alpha}_l, \boldsymbol{\mu}) = \mathrm{diag}\{(a_i(\boldsymbol{W}_{\boldsymbol{\alpha}_l}\boldsymbol{\mu}))_{i=1}^{n_b}\}$, $\boldsymbol{g}_{:,t,l} = (g_{i,t,l})_{i=1}^{n_b}$, $\bar{\boldsymbol{g}}_{:,t,l}(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu}) = (\bar{g}_{i,t,l}(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu}))_{i=1}^{n_b}$, $\boldsymbol{H}_t \in M_{n_b,n_v}(\mathbb{R})$ defined by $[\boldsymbol{H}_t]_{i,j} = p_{i,j,t}$, and

$$\boldsymbol{J}_{\boldsymbol{\alpha}_l}(\bar{\boldsymbol{g}}_{:,t,l}(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu})) = \tau \boldsymbol{A}(\boldsymbol{\alpha}_l, \boldsymbol{\mu})(-\mathrm{diag}\{\boldsymbol{H}_t \boldsymbol{W}_{\boldsymbol{\alpha}_l} \boldsymbol{f}\} \boldsymbol{L} \boldsymbol{J}_{\boldsymbol{\alpha}_l}(\boldsymbol{W}_{\boldsymbol{\alpha}_l}\boldsymbol{\mu}) + \boldsymbol{H}_t \boldsymbol{J}_{\boldsymbol{\alpha}_l}(\boldsymbol{W}_{\boldsymbol{\alpha}_l}\boldsymbol{f})). \tag{10}$$

The Jacobian matrices $\boldsymbol{J}_{\boldsymbol{\alpha}}(\boldsymbol{W}_{\boldsymbol{\alpha}}\boldsymbol{f})$ and $\boldsymbol{J}_{\boldsymbol{\alpha}}(\boldsymbol{W}_{\boldsymbol{\alpha}}\boldsymbol{\mu})$ are derived in Bousse *et al* (2016).

The overall scheme is briefly summarized in algorithm 1, for $\Phi = \Phi^{\mathrm{TOF}}$ or $\Phi = \Phi^{\mathrm{PET}}$. Each sub-maximization algorithm (MC-MMLEM and L-LBFGS) is run until convergence. A detailed version of the scheme can be found in Bousse *et al* (2016).

Algorithm 1: Summary of JRM

Input: TOF PET gated data $(\boldsymbol{g}_l)_{l=1}^{n_g}$, attenuation map $\boldsymbol{\mu}$ (also image and motion smoothing priors weights β and γ).
Output: Activity image $\boldsymbol{f}$, B-spline motion parameter $\boldsymbol{\theta}$
$\boldsymbol{\theta} \leftarrow \boldsymbol{0}$;
$\boldsymbol{f} \leftarrow \mathrm{MMLEM}(\boldsymbol{g}_1)$ (reconstruction from gate 1 w/o motion compensation) ;
for $n = 1, \ldots, n_{\max}$ **do**
    $\boldsymbol{\theta} \leftarrow \arg\max_{\boldsymbol{\theta}} \Phi(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu})$ (with L-BFGS QN line-search) ;
    $\boldsymbol{f} \leftarrow \mathrm{MC\text{-}MMLEM}(\boldsymbol{g}, \boldsymbol{\theta})$;
end

2.2.3. Alternative approach for the single gate case. When the task is to reconstruct from a single gate l_0 only, the projection model (5) can be simplified by ignoring the motion on the activity and warp the μ-map only, i.e. using a Poisson model with expectation

$$\check{g}_{i,t,l_0}(\boldsymbol{f}, \boldsymbol{\alpha}_{l_0}, \boldsymbol{\mu}) \triangleq \tau a_i(\boldsymbol{W}_{\boldsymbol{\alpha}_{l_0}}\boldsymbol{\mu})[\boldsymbol{H}\boldsymbol{f}]_{(i-1)n_t+t} + s_{i,t,l_0}$$

so that the Jacobian (10) expression is simplified. The reconstructed activity $\hat{\boldsymbol{f}}$ matches the data $\boldsymbol{g}_{l_0}$ without the need of being warped. This problem was addressed in Rezaei and Nuyts (2013). However, the generalization of this model to the multi gate case requires to estimate one activity image $\hat{\boldsymbol{f}}$ per gate, as opposed to JRM which only reconstruct one image $\boldsymbol{f}$, warped to each gate.

3. Experiments on simulated data

For this experiment we used only one gate, i.e. $n_g = 1$. The sums over l in the log-likelihood (6) and (7) are dropped. We only reconstruct from a single TOF PET dataset $\boldsymbol{g} \in \mathbb{N}^{n_b \times n_t}$,
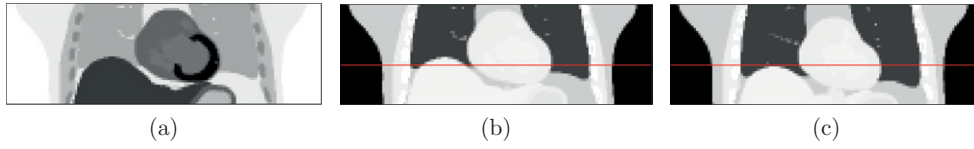
Figure 1. (a) $\boldsymbol{f}^{\star}$: true activity ; (b) $\boldsymbol{\mu}^{\star}$: true μ-map ; $\tilde{\boldsymbol{\mu}}$: deep breath-in misaligned μ-map.

$[\boldsymbol{g}]_{i,t} = g_{i,t}$. The motion parameter $\boldsymbol{\theta}$ reduces to a single $\boldsymbol{\alpha}$. The aim is to reconstruct a single activity image with a misaligned μ-map. The estimated deformation corresponds to the misalignment between the input μ-map and $\boldsymbol{g}$.

Each MC-MMLEM sub-optimimization was performed with 100 iterations which in our experiments was sufficient for convergence (no stopping rule was used), using the current $\boldsymbol{\alpha}$ estimate. Prior to each MC-MMLEM sub-optimimization, the image $\boldsymbol{f}$ was reinitialized to a blank image to avoid attenuation artifacts due to incomplete motion estimation in the previous iteration. We used the Fortran implementation proposed in Zhu *et al* (1997) for each L-BFGS sub-optimimization, with 80 iterations (sufficient for convergence in our experiments).

3.1. Simulation set-up

Our simulation set-up is similar to Bousse *et al* (2016, section IV.B) with Poisson noise. We used the XCAT phantom to generate ground truth end-expiration activity $\boldsymbol{f}^{\star}$ and attenuation map $\boldsymbol{\mu}^{\star}$ ($160 \times 160 \times 42$ volumes with 3.125 mm edge cubic voxels, corresponding to a 500 mm field of view, see figures 1(a) and (b)). We used a $n_{c} = 53 \times 53 \times 16$ B-spline grid to parametrize the motion. A misaligned end-inhalation μ-map, $\tilde{\boldsymbol{\mu}}$ (figure 1(c)), was also generated.

The spatial resolution of the PET was set to 6 mm FWHM and the temporal TOF resolution set to 500 ps FWHM. We used 10 TOF bins (332 ps width). The line integral operator $\boldsymbol{L}$ was adjusted to match the spatial resolution of the PET. We generated TOF and non-TOF data Poisson random vectors from $\boldsymbol{f}^{\star}$ and $\boldsymbol{\mu}^{\star}$ as:

$$g_{i,t} \sim \text{Poisson}(g_{i,t}^{\star}) \quad \text{and} \quad g_{i} = \sum_{t=1}^{n_{t}} g_{i,t},$$

with

$$g_{i,t}^{\star} = \tau a_{i}(\boldsymbol{\mu}^{\star})[\boldsymbol{H}\boldsymbol{f}^{\star}]_{(i-1)n_{t}+t} + s_{i,t}.$$

$\left(\hat{\boldsymbol{\alpha}}^{\text{TOF}}, \hat{\boldsymbol{f}}^{\text{TOF}}\right)$ and $\left(\hat{\boldsymbol{\alpha}}^{\text{PET}}, \hat{\boldsymbol{f}}^{\text{PET}}\right)$ were obtained with JRM using the misaligned attenuation map $\tilde{\boldsymbol{\mu}}$, i.e. by maximizing $(\boldsymbol{f}, \boldsymbol{\alpha}) \mapsto \Phi^{\text{TOF}}(\boldsymbol{f}, \boldsymbol{\alpha}, \tilde{\boldsymbol{\mu}})$ and $(\boldsymbol{f}, \boldsymbol{\alpha}) \mapsto \Phi^{\text{PET}}(\boldsymbol{f}, \boldsymbol{\alpha}, \tilde{\boldsymbol{\mu}})$ using algorithm 1.

JRM reconstructs the activity distribution $\boldsymbol{f}$ in the $\tilde{\boldsymbol{\mu}}$-space, which does not correspond to the observed PET gate, and $\boldsymbol{W}_{\boldsymbol{\alpha}}$ warps it to the observed gate. Therefore, we displayed the warped images $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}}\hat{\boldsymbol{f}}$, in order to be compared with the activity phantom $\boldsymbol{f}^{\star}$—used to generate the data.

We chose two different values of the image smoothness parameter β for Φ^{TOF} and Φ^{PET}. The image variance was estimated by reconstructing several noise replicates for different values of β, as in Bousse *et al* (2016), and the 2 values of β were chosen such that the variance in the output images $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}^{\text{PET}}}\hat{\boldsymbol{f}}^{\text{PET}}$ and $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}^{\text{TOF}}}\hat{\boldsymbol{f}}^{\text{TOF}}$ were the same.
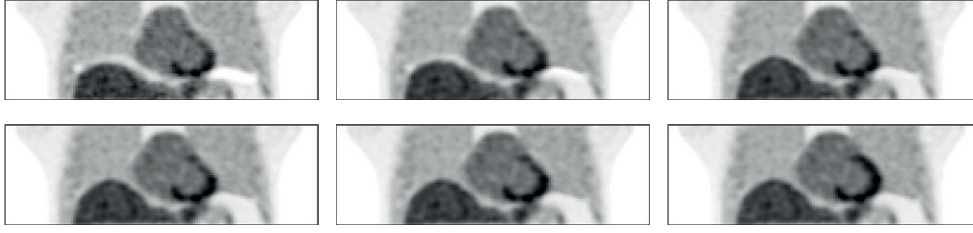
Figure 2. From top left to bottom right: $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}^{\mathrm{PET}}}\hat{\boldsymbol{f}}^{\mathrm{PET}}$ at iterations 0 (no motion compensation), 1, 10, 30, 50 and 100.
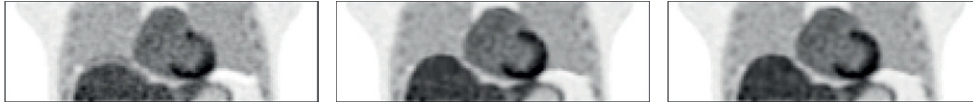
Figure 3. From left to right: $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}^{\mathrm{TOF}}}\hat{\boldsymbol{f}}^{\mathrm{TOF}}$ at iterations 0 (no motion compensation), 1 and 5.
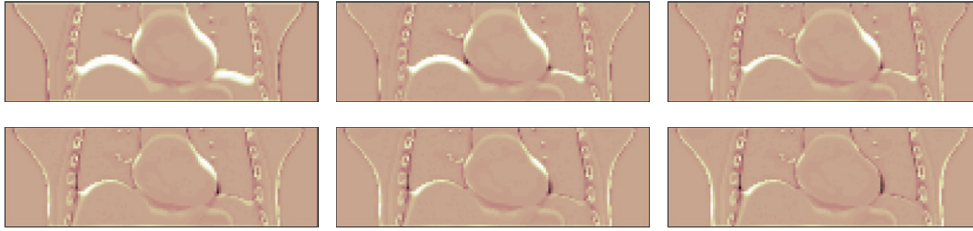
Figure 4. From top left to bottom right: relative difference $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}^{\mathrm{PET}}}\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}^{\star}$ at iterations 0 ($\hat{\boldsymbol{\alpha}} = \boldsymbol{0}$), 1, 10, 30, 50 and 100.

Figure 5. From left to right: relative difference $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}^{\mathrm{TOF}}}\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}^{\star}$ at iterations 0 ($\hat{\boldsymbol{\alpha}} = \boldsymbol{0}$), 1 and 5.

3.2. Results

Figures 2 and 3 show the JRM reconstructed activities $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}^{\mathrm{PET}}}\hat{\boldsymbol{f}}^{\mathrm{PET}}$ and $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}^{\mathrm{TOF}}}\hat{\boldsymbol{f}}^{\mathrm{TOF}}$ at several iterations (iteration 0 corresponds to the first MMLEM reconstruction with no motion compensation). It can be observed that PET JRM requires 50 to 100 iterations to reduce μ-map misalignment artifacts significantly whereas TOF JRM requires only 5. The relative differences $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}^{\mathrm{PET}}}\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}^{\star}$ and $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}^{\mathrm{TOF}}}\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}^{\star}$ are shown in figures 4 and 5[6]. Results show that the warped μ-map using TOF JRM estimated motion, i.e., $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}^{\mathrm{TOF}}}\tilde{\boldsymbol{\mu}}$, becomes similar to $\boldsymbol{\mu}^{\star}$ in 5 iterations, whereas it takes more than 50 iterations for the warped μ-map using non-TOF JRM estimated motion. Note that since TOF PET forward and back projections are computationally

---

[6] Note that figures 4(a) and 5(a) show the same image, i.e. $\boldsymbol{W}_{\boldsymbol{\alpha}}\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}^{\star}$ with $\boldsymbol{\alpha} = \boldsymbol{0}$.
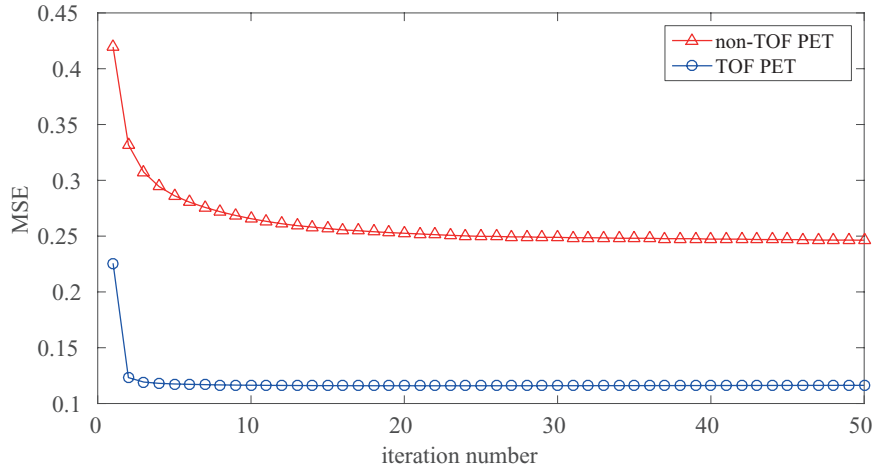
Figure 6. $\text{MSE}\left(W_{\hat{\alpha}^{\text{PET}}}\hat{f}^{\text{PET}}, f^{\star}\right)$ and $\text{MSE}\left(W_{\hat{\alpha}^{\text{TOF}}}\hat{f}^{\text{TOF}}, f^{\star}\right)$ VS iteration number.

more demanding, these observations should be put into perspective. Nevertheless, the reduction in number of iterations shows that using TOF PET significantly reduces the cross-talk in JRM as compared to non-TOF PET.

Quantitative assessment of the convergence rate of TOF JRM and non-TOF JRM is challenging because the objective functions to maximize are different, i.e., plotting the values of $\Phi^{\text{TOF}}\left(\hat{f}^{\text{TOF}}, \hat{\alpha}^{\text{TOF}}, \tilde{\mu}\right)$ and $\Phi^{\text{PET}}\left(\hat{f}^{\text{PET}}, \hat{\alpha}^{\text{PET}}, \tilde{\mu}\right)$ is not informative. We therefore relied on the mean square error (MSE), between the estimated activities, $W_{\hat{\alpha}^{\text{TOF}}}\hat{f}^{\text{TOF}}$ and $W_{\hat{\alpha}^{\text{PET}}}\hat{f}^{\text{PET}}$, and the ground truth $f^{\star}$. The MSE is defined for all $u, v \in \mathbb{R}^{n_{\text{v}}}$ as

$$\text{MSE}(u, v) = \frac{1}{n_{\text{v}}}\|u - v\|_2^2.$$

The plots of $\text{MSE}\left(W_{\hat{\alpha}^{\text{TOF}}}\hat{f}^{\text{TOF}}, f^{\star}\right)$ and $\text{MSE}\left(W_{\hat{\alpha}^{\text{PET}}}\hat{f}^{\text{PET}}, f^{\star}\right)$ are shown in figure 6. MSE results are consistent with the observations from figures 2 and 3: TOF JRM reaches a quasi-minimal MSE after 3 iterations whereas non-TOF JRM needs more than 30. The difference in MSE is due to the better conditioning of the TOF PET system matrix compared the non-TOF PET one.

4. Discussion and conclusion

This paper is an extension of our recent work Bousse *et al* (2016), and presents preliminary results on JRM with TOF PET data. Results from Rezaei *et al* (2012) and Rezaei *et al* (2014) demonstrated that TOF PET can significantly reduce the cross-talk in the joint-likelihood. Our experiments led to a similar observation: when the reconstructed attenuation is misaligned with the PET data, TOF JRM requires 10 to 20 times less iterations than non-TOF JRM. This demonstrates that TOF JRM significantly reduces the cross-talk compared to using non-TOF PET data, and extend its practicality to situations where the μ-map is severely misaligned with the PET.

Acknowledgments

References

Blume M, Martinez-Müller A, Keil A, Navab N and Rafecas M 2010 Joint reconstruction of image and motion in gated positron emission tomography *IEEE Trans. Med. Imaging* **29** 1892–906

Bousse A, Bertolli O, Atkinson D, Arridge S, Ourselin S, Hutton B F and Thielemans K 2016 Maximum-likelihood joint image reconstruction/motion estimation in attenuation-corrected respiratory gated PET/CT using a single attenuation map *IEEE Trans. Med. Imaging IEEE Trans. Med. Imag.* **35** 217–28

De Pierro A R 1995 A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography *IEEE. Trans. Med. Imaging* **14** 132–7

Jacobson M W and Fessler J A 2003 Joint estimation of image and deformation parameters in motion-corrected PET *IEEE Nuclear Science Symp. Conf. Record* pp 3290–4

Jacobson M W 2006 Approaches to motion-corrected PET image reconstruction from respiratory gated projection data *PhD Thesis* The University of Michigan, Ann Arbor, USA

Nocedal J and Wright S J 2006 *Numerical Optimization* 2nd edn (New York: Springer)

Nyflot M J, Lee T C, Alessio A M, Wollenweber S D, Stearns C W, Bowen S R and Kinahan P E 2015 Impact of CT attenuation correction method on quantitative respiratory-correlated (4D) PET/CT imaging *Med. Phys.* **42** 110–20

Rezaei A, Defrise M, Bal G, Michel C, Conti M, Watson C and Nuyts J 2012 Simultaneous reconstruction of activity and attenuation in time-of-flight PET *IEEE Trans. Med. Imaging* **31** 2224–33

Rezaei A, Defrise M and Nuyts J 2014 ML-reconstruction for TOF-PET with simultaneous estimation of the attenuation factors *IEEE Trans. Med. Imaging* **33** 1563–72

Rezaei A and Nuyts J 2013 Simultaneous reconstruction of the activity image and registration of the CT image in TOF-PET *IEEE Nuclear Science Symp. and Medical Imaging Conf. Record*

Zhu C, Byrd R H, Lu P and Nocedal J 1997 Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization *ACM Trans. Math. Softw.* **23** 550–560