Maximum-Likelihood Joint Image Reconstruction/ Motion Estimation in Attenuation-Corrected Respiratory Gated PET/CT Using a Single Attenuation Map

Alexandre Bousse, Ottavia Bertolli, David Atkinson, Simon Arridge, Sébastien Ourselin, Brian F. Hutton, *Senior Member, IEEE*, and Kris Thielemans, *Senior Member, IEEE*

***Abstract*—This work provides an insight into positron emission tomography (PET) joint image reconstruction/motion estimation (JRM) by maximization of the likelihood, where the probabilistic model accounts for warped attenuation. Our analysis shows that maximum-likelihood (ML) JRM returns the same reconstructed gates for any attenuation map (μ-map) that is a deformation of a given μ-map, regardless of its alignment with the PET gates. We derived a joint optimization algorithm accordingly, and applied it to simulated and patient gated PET data. We first evaluated the proposed algorithm on simulations of respiratory gated PET/CT data based on the XCAT phantom. Our results show that independently of which μ-map is used as input to JRM: (i) the warped μ-maps correspond to the gated μ-maps, (ii) JRM outperforms the traditional post-registration reconstruction and consolidation (PRRC) for hot lesion quantification and (iii) reconstructed gated PET images are similar to those obtained with gated μ-maps. This suggests that a breath-held μ-map can be used. We then applied JRM on patient data with a μ-map derived from a breath-held high resolution CT (HRCT), and compared the results with PRRC, where each reconstructed PET image was obtained with a corresponding cine-CT gated μ-map. Results show that JRM with breath-held HRCT achieves similar reconstruction to that using PRRC with cine-CT. This suggests a practical low-dose solution for implementation of motion-corrected respiratory gated PET/CT.**

***Index Terms*—Attenuation correction, attenuation mismatch, direct motion estimation, gated PET, image reconstruction, maximum-likelihood.**

I. Introduction

PATIENT respiratory motion during positron emission tomography (PET) acquisition not only degrades image resolution [1] but also results in reconstruction errors due to possible mismatch of PET data with the sequentially acquired computed tomography (CT) attenuation map (μ-map) [2].

One approach to overcome this issue is to reconstruct from a single motion-free PET gate that is aligned with CT data acquired at the same respiratory phase [3], [4]. This approach can readily avoid the effects of respiratory motion but the limitation to a single gate reduces the number of usable counts which results in noisy PET reconstruction.

4-D CT images can be obtained following a "cine" protocol, i.e., by performing repeated axial acquisitions during one (or sometimes more) respiratory cycle(s) [5], [6]. Acquired data are then regrouped according to their respiratory phases, either using the CT data themselves, or using a respiratory-motion tracking system (see [7] for a review). This approach, combined with gated PET, allows 4-D attenuation-corrected (AC) PET reconstruction, but is vulnerable to artifacts in the gated CT due to variation in the breathing cycle [8].

Another way to tackle this issue consists of deriving a motion model from non-AC reconstructed PET gates and applying this model to a single CT image [9], [10] in order to generate a 4-D μ-maps sequence. However, motion estimation from individually reconstructed gated non-AC PET images is subject to noise and low contrast, leading to motion estimation errors that can propagate to the deformed CT image and eventually to the final PET reconstruction.

Along similar lines, a pre-estimated motion model can be estimated from the PET data [11], [12] or from another imaging modality such as CT [13]–[15] or magnetic resonance imaging (MRI) [16]–[18], and then can be directly incorporated into the PET system matrix and the attenuation correction factors.

Nuyts *et al.* [19]–[21] have demonstrated the potential of joint estimation of the activity distribution and the attenuation map/ coefficient factors from emission data. This class of methods could allow one to perform AC reconstruction from PET gated data with the μ-maps reconstructed at each gate. However, it is a very ill-posed problem on non time-of-flight (TOF) PET data [20], and TOF data only allow to reconstruct the activity up to a constant factor [22].

A completely different type of method consists of performing a penalized maximum-likelihood (PML) joint image reconstruction/motion estimation (JRM), where the probabilistic model accounts for an unknown motion applied to both the

Manuscript received April 23, 2015; revised July 22, 2015; accepted July 28, 2015. Date of publication August 03, 2015; date of current version December 29, 2015. This work was supported by UK EPSRC (EP/K005278/1), the NIHR-funded UCH Biomedical Research Centre and EPSRC CASE Award 13220093 with GE Healthcare. *Asterisk indicates corresponding author.*

*A. Bousse is with the Institute of Nuclear Medicine, University College London, NW1 2BU London, U.K. (e-mail: a.bousse@ucl.ac.uk).

O. Bertolli, B. F. Hutton, and K. Thielemans are with the Institute of Nuclear Medicine, University College London, NW1 2BU London, U.K.

D. Atkinson is with the Centre for Medical Imaging, University College London, NW1 2PG London, U.K.

S. Arridge and S. Ourselin are with the Centre for Medical Image Computing, University College London, WC1E 6BT London, U.K.

B. F. Hutton is also with the Centre for Medical Radiation Physics, University of Wollongong, Wollongong NSW 2522, Australia.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2015.2464156

activity volume and μ-map. This approach has the advantage of allowing one to estimate the motion directly from the raw data (without having to pre-register the volumes) and to utilize only one μ-map, warped alongside the activity. JRM has already been used in the past [23], [24] but these papers ignored attenuation. This is mainly for two reasons: (i) the derivation of an optimization algorithm proved to be challenging; (ii) the μ-map used in the likelihood function normally needs to be aligned to the motion-free activity volume. As the activity distribution is unknown, it is unclear how the attenuation map should be chosen.

This paper is an extension of [23], [24], with incorporation of the attenuation in the likelihood term, and follows the work we initiated in [25]. Similarly to [24], the estimated activity volume is a "virtual" image that is not associated to a particular gate. The main result of this paper is that in such a setting, JRM reconstructs the same gated activity distribution regardless of the μ-map input, provided the input can be derived as deformations of a unique μ-map. This means that JRM can be potentially used with a breath-held CT-derived μ-map that is not synchronized with any of the PET gates.

The objective of this work was to assess the feasibility of using JRM with attenuation and its ability to deal with misaligned μ-maps. We implemented an algorithm for PML reconstruction that accounts for the necessary transformations of the μ-map, based on the discretization scheme proposed in [23].

We evaluated JRM using both simulated and patient data. Simulations included μ-map misalignment correction and reconstruction of gated PET data. We compared JRM with post-reconstruction registration and consolidation (PRRC), where each PET gate was reconstructed with the corresponding exactly aligned μ-map. We also reported an initial evaluation on patient data, where JRM was applied with a breath-held high resolution CT (HRCT) derived μ-map and the results were compared with PRRC, where each gate was reconstructed using cine-CT gated μ-maps.

II. Theory

A. Attenuated PET Measurement Without Motion

The activity distribution and attenuation map can be modeled as functions $f \in \mathcal{C}^+$ and $\mu \in \mathcal{C}^+$ respectively, where $\mathcal{C}^+$ denotes the set of non-negative continuous functions on $\mathbb{R}^3$. The activity and attenuation at position $\boldsymbol{r} \in \mathbb{R}^3$ are $f(\boldsymbol{r})$ and $\mu(\boldsymbol{r})$. Although f is unknown, we assume that an attenuation map μ is obtained from a separate measurement such as an X-ray CT acquisition. The PET measured counts are modeled as a Poisson random vector $\boldsymbol{g} \in \mathbb{N}^{n_\mathrm{b}}$ with independent entries $[\boldsymbol{g}]_i = g_i$, $i = 1, \ldots, n_\mathrm{b}$, n_b being the number of detector bins:

$$g_i \sim \mathrm{Poisson}(\bar{g}_i(f,\mu)), \quad \bar{g}_i(f,\mu) = \tau a_i(\mu)\mathcal{H}_i f + s_i$$

where τ is the scan duration, $\mathcal{H}_i f$ is the expected number of counts per unit of time in absence of attenuation and background events,

$$\mathcal{H}_i f = \int_\Omega f(\boldsymbol{r}) h_i(\boldsymbol{r}) \mathrm{d}\boldsymbol{r}, \tag{1}$$

$a_i(\mu)$ is the attenuation factor along the segment L_i connecting the detectors of bin i,

$$a_i(\mu) = \exp\left(-\int_{L_i} \mu(\boldsymbol{r})\mathrm{d}\boldsymbol{r}\right) \tag{2}$$

and s_i is the expected number of background events (scatter/random coincidence) at bin i. The function h_i in (1) is the PET system response at bin i and $\Omega \subset \mathbb{R}^3$ is a compact set representing the field of view.

B. Attenuated PET Measurement With Motion

In presence of patient motion, both activity and attenuation are deformed. The set of diffeomorphic functions on $\mathbb{R}^3$ is denoted $\mathcal{D}$. A deformation is modeled by a diffeomorphism $\varphi \in \mathcal{D}$, yielding a warping operator $\mathcal{W}_\varphi : \mathcal{C}^+ \to \mathcal{C}^+$ defined as

$$\mathcal{W}_\varphi f = f \circ \varphi, \quad \forall f \in \mathcal{C}^+.$$

$\mathcal{W}_\varphi$ is invertible on $\mathcal{C}^+$ with $[\mathcal{W}_\varphi]^{-1} = \mathcal{W}_{\varphi^{-1}}$ and $\mathcal{W}_\varphi \mathcal{W}_\psi = \mathcal{W}_{\psi\circ\varphi}$.

Under the assumption that patient respiratory motion is quasi-cyclic, acquired data are regrouped into n_g gates. We will ignore intra-gate motion and assume that on each gate $l = 1, \ldots, n_\mathrm{g}$ the patient is static. At gate l, the activity distribution and attenuation map are deformed versions of f and μ with a diffeomorphism $\varphi_l \in \mathcal{D}$. In the gated case, the measured data are a Poisson random vector $\boldsymbol{g}_l \in \mathbb{N}^{n_\mathrm{b}}$ with independent entries $[\boldsymbol{g}_l]_i = g_{i,l}$,

$$\begin{aligned} g_{i,l} &\sim \mathrm{Poisson}(\bar{g}_{i,l}(f,\varphi_l,\mu)), \\ \bar{g}_{i,l}(f,\varphi_l,\mu) &= \tau_l a_i(\mathcal{W}_{\varphi_l}\mu)\mathcal{H}_i\mathcal{W}_{\varphi_l} f + s_{i,l} \end{aligned} \tag{3}$$

where $\mathcal{H}_i$ and a_i were defined in (1) and (2), τ_l is the duration of gate l and $s_{i,l}$ is the expected number of background events at bin i, gate l.

C. Joint Maximum-Likelihood

Omitting terms independent on f and φ_l, the log-likelihood of the observables $(\boldsymbol{g}_l)_{l=1}^{n_\mathrm{g}} \in \mathbb{N}^{n_\mathrm{b}\times n_\mathrm{g}}$ is

$$L(f, \boldsymbol{\varphi}, \mu) = \sum_{l=1}^{n_\mathrm{g}} \Lambda_l(\bar{\boldsymbol{g}}_l(f,\varphi_l,\mu)) \tag{4}$$

with $\Lambda_l(\bar{\boldsymbol{g}}) \triangleq \sum_{i=1}^{n_\mathrm{b}} g_{i,l}\log\bar{g}_i - \bar{g}_i$, $\boldsymbol{\varphi} \triangleq (\varphi_l)_{l=1}^{n_\mathrm{g}} \in \boldsymbol{\mathcal{D}}$, $\boldsymbol{\mathcal{D}} \triangleq \mathcal{D}^{n_\mathrm{g}}$. Joint PET image reconstruction/motion estimation (JRM) by maximum-likelihood (ML) consists of estimating a pair $(\hat{f}, \hat{\boldsymbol{\varphi}})$ that explains the measurements $(\boldsymbol{g}_l)_{l=1}^{n_\mathrm{g}}$ by solving the following optimization problem:

$$\max_{f\in\mathcal{C}^+, \boldsymbol{\varphi}\in\boldsymbol{\mathcal{D}}} L(f, \boldsymbol{\varphi}, \mu). \tag{5}$$

At each gate $l = 1, \ldots, n_\mathrm{g}$, the resulting estimated activity image is $\hat{f}_l \triangleq \mathcal{W}_{\hat{\varphi}_l}\hat{f}$ and is obtained by warping the "virtual" activity image $\hat{f}$ estimated from the entire dataset $(\boldsymbol{g}_l)_{l=1}^{n_\mathrm{g}}$.

In fact, it can be seen that the log-likelihood only indirectly depends on f and μ, i.e., via $\mathcal{W}_{\varphi_l} f$ and $\mathcal{W}_{\varphi_l}\mu$. This observation leads us to consider the gated images as the final output of the estimation problem, with the virtual image f and the deformation only as intermediate objects. This shift in emphasis leads to the main result of this paper, as described in the next section.

D. On the Attenuation Map Dependency

Statistical model (3) suggests that μ and f should correspond to the same patient state. In fact, we can see that the reconstructed gated images obtained by solving (5) are identical for all μ-maps resulting from deformations of a common μ-map. We denote $\psi \circ \boldsymbol{\varphi} \triangleq (\psi \circ \varphi_1, \ldots, \psi \circ \varphi_{n_\mathrm{g}})$ for all $\boldsymbol{\varphi} \in \boldsymbol{\mathcal{D}}$, $\psi \in \mathcal{D}$.

Proposition 1: Let $\psi \in \mathcal{D}$, $\tilde{\mu} = \mu \circ \psi$ and $(\hat{f}, \hat{\boldsymbol{\varphi}}) \in \mathcal{C}^+ \times \boldsymbol{\mathcal{D}}$ be a maximizer of $(f, \boldsymbol{\varphi}) \mapsto L(f, \boldsymbol{\varphi}, \mu)$, then $(\tilde{f}, \tilde{\boldsymbol{\varphi}}) = (\mathcal{W}_\psi \hat{f}, \psi^{-1} \circ \hat{\boldsymbol{\varphi}})$ is a maximizer of $(f, \boldsymbol{\varphi}) \mapsto L(f, \boldsymbol{\varphi}, \tilde{\mu})$ and for all $l = 1, \ldots, n_\mathrm{g}$:

$$\mathcal{W}_{\tilde{\varphi}_l} \tilde{f} = \mathcal{W}_{\hat{\varphi}_l} \hat{f}, \quad \mathcal{W}_{\tilde{\varphi}_l} \tilde{\mu} = \mathcal{W}_{\hat{\varphi}_l} \mu. \tag{6}$$

Reciprocally, if $(\tilde{f}, \tilde{\boldsymbol{\varphi}}) \in \mathcal{C}^+ \times \boldsymbol{\mathcal{D}}$ is a maximizer of $(f, \boldsymbol{\varphi}) \mapsto L(f, \boldsymbol{\varphi}, \tilde{\mu})$, then $(\hat{f}, \hat{\boldsymbol{\varphi}}) = (\mathcal{W}_{\psi^{-1}} \tilde{f}, \psi \circ \tilde{\boldsymbol{\varphi}})$ is a maximizer of $(f, \boldsymbol{\varphi}) \mapsto L(f, \boldsymbol{\varphi}, \mu)$ and (6) is also verified.

Proof: We have the following identity for all $f \in \mathcal{C}^+$, $\mu \in \mathcal{C}^+$ and $\boldsymbol{\varphi} \in \boldsymbol{\mathcal{D}}$:

$$\begin{aligned} L(f, \boldsymbol{\varphi}, \mu) &= L(f \circ \psi, \psi^{-1} \circ \boldsymbol{\varphi}, \mu \circ \psi) \\ &= L(\mathcal{W}_\psi f, \psi^{-1} \circ \boldsymbol{\varphi}, \tilde{\mu}). \end{aligned} \tag{7}$$

Let $(\hat{f}, \hat{\boldsymbol{\varphi}})$ be a maximizer of $(f, \boldsymbol{\varphi}) \mapsto L(f, \boldsymbol{\varphi}, \mu)$:

$$L(\hat{f}, \hat{\boldsymbol{\varphi}}, \mu) = \max_{f \in \mathcal{C}^+, \boldsymbol{\varphi} \in \boldsymbol{\mathcal{D}}} L(f, \boldsymbol{\varphi}, \mu). \tag{8}$$

Combining (7) and (8) leads to

$$L(\mathcal{W}_\psi \hat{f}, \psi^{-1} \circ \hat{\boldsymbol{\varphi}}, \tilde{\mu}) = \max_{f \in \mathcal{C}^+, \boldsymbol{\varphi} \in \boldsymbol{\mathcal{D}}} L(f, \boldsymbol{\varphi}, \tilde{\mu})$$

and $(\tilde{f}, \tilde{\boldsymbol{\varphi}}) = (\mathcal{W}_\psi \hat{f}, \psi^{-1} \circ \hat{\boldsymbol{\varphi}})$ is a maximiser of $(f, \boldsymbol{\varphi}) \mapsto L(f, \boldsymbol{\varphi}, \tilde{\mu})$. We also have

$$\forall l = 1, \ldots, n_\mathrm{g}, \quad \mathcal{W}_{\tilde{\varphi}_l} \tilde{f} = \mathcal{W}_{\hat{\varphi}_l} [\mathcal{W}_\psi]^{-1} \mathcal{W}_\psi \hat{f} = \mathcal{W}_{\hat{\varphi}_l} \hat{f}$$

and similarly $\mathcal{W}_{\tilde{\varphi}_l} \tilde{\mu} = \mathcal{W}_{\hat{\varphi}_l} \mu$. The converse is demonstrated the same way. ∎

Proposition 1 does not assert the existence nor the uniqueness of a maximizer, but it can be interpreted as follows: if $(\hat{f}, \hat{\boldsymbol{\varphi}})$ is a likely candidate for (5), then $(\tilde{f}, \tilde{\boldsymbol{\varphi}}) = (\mathcal{W}_\psi \hat{f}, \psi^{-1} \circ \hat{\boldsymbol{\varphi}})$ is an equally likely candidate for (5) with $\tilde{\mu}$, and satisfies (6). In practice, this result means that *the JRM-reconstructed gated images are the same when using the attenuation map* μ or a deformed version $\tilde{\mu} = \mathcal{W}_\psi \mu$. For example, ψ can represent the mismatch between the attenuation acquired with a deep breath-hold X-ray CT scan and the attenuation map during the PET scan.

III. Joint Image Reconstruction/Motion Estimation Algorithm

A. Discretization

Instead of discretizing (3) directly, we adopted the "warp and project" discretization scheme from Jacobson & Fessler [23]. This approach allows one to perform deformation and projection separately, and thus, simplifies implementation.

1) Warping Matrix: We assume that an image function f and μ can be decomposed on basis functions centered on a voxel grid $\mathcal{G} = \{\boldsymbol{r}_k, k = 1, \ldots, n_\mathrm{v}\} \subset \Omega$ that coincides with the voxel centers:

$$f(\boldsymbol{r}) = \sum_{k=1}^{n_\mathrm{v}} f_k w(\boldsymbol{r} - \boldsymbol{r}_k), \quad \mu(\boldsymbol{r}) = \sum_{k=1}^{n_\mathrm{v}} \mu_k w(\boldsymbol{r} - \boldsymbol{r}_k) \tag{9}$$

where $w : \mathbb{R}^3 \to \mathbb{R}^+$ is an interpolating function of the form $w(\boldsymbol{r}) = e(x)e(y)e(z)$, $\boldsymbol{r} = (x, y, z)$, and $e : \mathbb{R} \to \mathbb{R}^+$ is positive, symmetric and continuously differentiable. The discretized activity image and attenuation map are represented by the non-negative coefficients $\boldsymbol{f} = (f_j)_{j=1}^{n_\mathrm{v}}$ and $\boldsymbol{\mu} = (\mu_j)_{j=1}^{n_\mathrm{v}}$. Deformation of $\boldsymbol{f}$ is performed with a deformation of the continuous images f followed by a re-sampling on $\mathcal{G}$: for all $j = 1, \ldots, n_\mathrm{v}$,

$$\begin{aligned} \mathcal{W}_\varphi f(\boldsymbol{r}_j) &= \sum_{k=1}^{n_\mathrm{v}} f_k w(\varphi(\boldsymbol{r}_j) - \boldsymbol{r}_k) \\ &= [\boldsymbol{W}_\varphi \boldsymbol{f}]_j \end{aligned} \tag{10}$$

where the warping square matrix $\boldsymbol{W}_\varphi \in \mathbb{R}^{n_\mathrm{v} \times n_\mathrm{v}}$ is defined by $[\boldsymbol{W}_\varphi]_{j,k} \triangleq w(\varphi(\boldsymbol{r}_j) - \boldsymbol{r}_k)$. The warped discrete activity and attenuation map are $\boldsymbol{W}_\varphi \boldsymbol{f}$ and $\boldsymbol{W}_\varphi \boldsymbol{\mu}$.

2) B-Spline Model for Deformation: Similarly to [23], we use a B-spline representation for φ. Let $\tilde{\mathcal{G}} = \{\tilde{\boldsymbol{r}}_n, n = 1, \ldots, n_\mathrm{c}\} \subset \mathcal{G}$ be a uniform sub-grid of $\mathcal{G}$ comprising n_c control points. The motion model is parametrized by $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y, \boldsymbol{\alpha}^Z) \in \mathbb{R}^{n_\mathrm{w}}$, where $n_\mathrm{w} = 3 \times n_\mathrm{c}$ is the number of warping parameters and $\boldsymbol{\alpha}^X = (\alpha_n^X)_{n=1}^{n_\mathrm{c}}$, $\boldsymbol{\alpha}^Y = (\alpha_n^Y)_{n=1}^{n_\mathrm{c}}$, $\boldsymbol{\alpha}^Z = (\alpha_n^Z)_{n=1}^{n_\mathrm{c}}$ are the motion B-spline coefficients along each axis X, Y and Z:

$$\varphi_{\boldsymbol{\alpha}}(\boldsymbol{r}) \triangleq \boldsymbol{r} + \begin{bmatrix} \sum_{n=1}^{n_\mathrm{c}} \alpha_n^X \mathcal{B}\left(\frac{\boldsymbol{r} - \tilde{\boldsymbol{r}}_n}{\sigma}\right) \\ \sum_{n=1}^{n_\mathrm{c}} \alpha_n^Y \mathcal{B}\left(\frac{\boldsymbol{r} - \tilde{\boldsymbol{r}}_n}{\sigma}\right) \\ \sum_{n=1}^{n_\mathrm{c}} \alpha_n^Z \mathcal{B}\left(\frac{\boldsymbol{r} - \tilde{\boldsymbol{r}}_n}{\sigma}\right) \end{bmatrix} = \begin{bmatrix} \varphi_{\boldsymbol{\alpha}}^X(\boldsymbol{r}) \\ \varphi_{\boldsymbol{\alpha}}^Y(\boldsymbol{r}) \\ \varphi_{\boldsymbol{\alpha}}^Z(\boldsymbol{r}) \end{bmatrix} \tag{11}$$

with $\boldsymbol{r} = (x, y, z) \in \mathbb{R}^3$, $\mathcal{B}(\boldsymbol{r}) \triangleq b(x)b(y)b(z) \in \mathbb{R}^+$, $b : \mathbb{R} \to \mathbb{R}^+$ is the cubic B-spline function and σ is the distance between control points. For notation compactness, $\boldsymbol{W}_{\varphi_{\boldsymbol{\alpha}}}$ is denoted $\boldsymbol{W}_{\boldsymbol{\alpha}}$. At each gate l, the motion B-spline coefficients vector is denoted $\boldsymbol{\alpha}_l$. The deformed activity and attenuation map are $\boldsymbol{W}_{\boldsymbol{\alpha}_l} \boldsymbol{f}$ and $\boldsymbol{W}_{\boldsymbol{\alpha}_l} \boldsymbol{\mu}$.

3) PET and CT Projector: The unattenuated PET system is modeled by a matrix $\boldsymbol{H} \in \mathbb{R}_+^{n_\mathrm{b} \times n_\mathrm{v}}$ where $[\boldsymbol{H}]_{i,j}$ is the probability that an annihilation occurring at voxel j is detected in detector pair i. Similarly, the X-ray line integral operator used in (2) is a matrix $\boldsymbol{L} \in \mathbb{R}_+^{n_\mathrm{b} \times n_\mathrm{v}}$ where $[\boldsymbol{L}]_{i,j} = \ell_{i,j}$ is the length of the intersection of L_i with voxel j.

B. Discrete Penalized Log-Likelihood

The expected number of detected events at bin i, gate l, introduced in (3) is redefined as a function of $\boldsymbol{f}$, $\boldsymbol{\alpha}_l$ and $\boldsymbol{\mu}$:

$$\bar{g}_{i,l}(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu}) = \tau_l a_i(\boldsymbol{W}_{\boldsymbol{\alpha}_l} \boldsymbol{\mu}) [\boldsymbol{H} \boldsymbol{W}_{\boldsymbol{\alpha}_l} \boldsymbol{f}]_i + s_{i,l} \tag{12}$$

with

$$a_i(\boldsymbol{\mu}) \triangleq \exp(-[\boldsymbol{L}\boldsymbol{\mu}]_i).$$

The vectorial writing of (12) is

$$\bar{\boldsymbol{g}}_l(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu}) = \tau_l \boldsymbol{A}(\boldsymbol{W}_{\boldsymbol{\alpha}_l} \boldsymbol{\mu}) \boldsymbol{H} \boldsymbol{W}_{\boldsymbol{\alpha}_l} \boldsymbol{f} + \boldsymbol{s}_l \tag{13}$$

where $\boldsymbol{A}(\boldsymbol{\mu}) \in \mathbb{R}^{n_\mathrm{b} \times n_\mathrm{b}}$ is a diagonal matrix defined by $[\boldsymbol{A}(\boldsymbol{\mu})]_{i,i} = a_i(\boldsymbol{\mu})$ for all $i = 1, \ldots, n_\mathrm{b}$, and $\boldsymbol{s}_l$ is the vector of expected background events at gate l. We also define the attenuation-corrected (AC) PET system matrix

$$\boldsymbol{H}_\mathrm{a}(\boldsymbol{\mu}) \triangleq \boldsymbol{A}(\boldsymbol{\mu})\boldsymbol{H} \tag{14}$$

and the AC motion-compensated (AC-MC) PET system matrix:

$$\boldsymbol{H}_\mathrm{a}^\mathrm{m}(\boldsymbol{\alpha}_l, \boldsymbol{\mu}) \triangleq \boldsymbol{H}_\mathrm{a}(\boldsymbol{W}_{\boldsymbol{\alpha}_l}\boldsymbol{\mu})\boldsymbol{W}_{\boldsymbol{\alpha}_l}. \tag{15}$$

An alternative formulation of (13) is

$$\bar{\boldsymbol{g}}_l(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu}) = \tau_l \boldsymbol{H}_\mathrm{a}^\mathrm{m}(\boldsymbol{\alpha}_l, \boldsymbol{\mu})\boldsymbol{f} + \boldsymbol{s}_l.$$

The entire collection of motion parameters is denoted $\boldsymbol{\theta} = (\boldsymbol{\alpha}_l)_{l=1}^{n_\mathrm{g}} \in \mathbb{R}^{n_\mathrm{w} \times n_\mathrm{g}}$. The log-likelihood (4) is redefined as a function of $\boldsymbol{f}$, $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$:

$$L(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu}) = \sum_{l=1}^{n_\mathrm{g}} \Lambda_l(\bar{\boldsymbol{g}}_l(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu}))$$

where Λ_l was defined in (4).

In order to control noise, two quadratic penalty terms $U(\boldsymbol{f})$ and $V(\boldsymbol{\theta})$ are introduced:

$$U(\boldsymbol{f}) = -\frac{1}{2}\sum_{j=1}^{n_\mathrm{v}} \sum_{k \in \mathcal{N}_j} \omega_{j,k}(f_j - f_k)^2, V(\boldsymbol{\theta}) = -\sum_{l=1}^{n_\mathrm{g}} \tau_l v(\boldsymbol{\alpha}_l)$$

with

$$\forall \boldsymbol{\alpha}, \quad v(\boldsymbol{\alpha}) = \frac{1}{2}\sum_{n=1}^{n_\mathrm{c}} \sum_{m \in \tilde{\mathcal{N}}_n} \tilde{\omega}_{n,m} \sum_{C \in \{X,Y,Z\}} (\alpha_n^C - \alpha_m^C)^2$$

where $\mathcal{N}_j$ and $\tilde{\mathcal{N}}_n$ are the neighborhood of j in $\mathcal{G}$ and of n in $\tilde{\mathcal{G}}$ respectively, $\omega_{j,k}$ is the inverse distance between voxel j and k, $\tilde{\omega}_{n,m}$ is the inverse distance between control points n and m. Note that U and V can be replaced by edge-preserving penalties [26]. The penalized log-likelihood is

$$\Phi(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu}) = L(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu}) + \beta U(\boldsymbol{f}) + \gamma V(\boldsymbol{\theta}). \tag{16}$$

JRM by penalized-ML (PML) is performed by solving the optimization problem

$$\max_{\boldsymbol{f} \geq \boldsymbol{0}, \boldsymbol{\theta}} \Phi(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu}).$$

Although it is not optimized, $\boldsymbol{\mu}$ is present as a hyper-parameter in order to assess the effect of a deformed attenuation map.

The maximization of Φ is performed in alternation between $\boldsymbol{f}$ and $\boldsymbol{\theta}$.

C. Motion Update

The motion estimation part consists of maximizing the penalized log-likelihood Φ (16) with respect to $\boldsymbol{\theta}$. We used a quasi-Newton (QN) approach with an approximated Hessian matrix. Most efforts are focused on the derivation of the Jacobian of the expected projection, which is detailed in Appendix A. The gradient is similar to [24] with inclusion of the warped attenuation map $\boldsymbol{W}_{\boldsymbol{\alpha}}\boldsymbol{\mu}$.

Given a fixed $\boldsymbol{f}$, $\Phi(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu})$ is maximized with respect to $\boldsymbol{\theta} = (\boldsymbol{\alpha}_l)_{l=1}^{n_\mathrm{g}}$. For notation compactness, $\Phi(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu})$, $L(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu})$ and $\bar{\boldsymbol{g}}_l(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu})$ are temporarily rewritten $\Phi(\boldsymbol{\theta})$, $L(\boldsymbol{\theta})$ and $\bar{\boldsymbol{g}}_l(\boldsymbol{\alpha}_l)$ respectively, and $U(\boldsymbol{f})$ is omitted. We rewrite Φ as function of $\boldsymbol{\theta}$ only:

$$\begin{aligned}\Phi(\boldsymbol{\theta}) &= L(\boldsymbol{\theta}) + \gamma V(\boldsymbol{\theta}) \\ &= \sum_{l=1}^{n_\mathrm{g}} \Lambda_l(\bar{\boldsymbol{g}}_l(\boldsymbol{\alpha}_l)) + \gamma V(\boldsymbol{\theta}).\end{aligned}$$

Let $\boldsymbol{\theta}^{(q)} = (\boldsymbol{\alpha}_l^{(q)})_{l=1}^{n_\mathrm{g}}$ be a current estimate of $\boldsymbol{\theta}$ at iteration q. The next estimate $\boldsymbol{\theta}^{(q+1)}$ is obtained by performing a maximization along a gradient ascent search direction $\boldsymbol{t}^{(q)}$:

$$\boldsymbol{\theta}^{(q+1)} = \boldsymbol{\theta}^{(q)} + \delta^{(q)}\boldsymbol{t}^{(q)}$$

with $\delta^{(q)}$ solution of

$$\max_{\delta \geq 0} \Phi(\boldsymbol{\theta}^{(q)} + \delta \boldsymbol{t}^{(q)}). \tag{17}$$

The search direction is of the form $\boldsymbol{t}^{(q)} = -\boldsymbol{B}^{(q)}\nabla\Phi(\boldsymbol{\theta}^{(q)})$, where $\boldsymbol{B}^{(q)}$ is a negative-definite matrix approximating the inverse of the Hessian $\nabla^2\Phi(\boldsymbol{\theta}^{(q)})$. It can be obtained at each iteration using a Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (see Chapter 6 in [27]). Because of the large dimension of $\boldsymbol{\theta}$, we computed $\boldsymbol{t}^{(q)}$ using a "limited-memory" BFGS (L-BFGS) algorithm (see Chapter 7 in [27]), which computes the matrix/vector product $\boldsymbol{t}^{(q)} = -\boldsymbol{B}^{(q)}\nabla\Phi(\boldsymbol{\theta}^{(q)})$ without storing $\boldsymbol{B}^{(q)}$. At this stage, any line-search method can be used to solve (17). We followed the approach described in [28] and its Fortran implementation [29] to compute $\boldsymbol{t}^{(q)}$ and $\delta^{(q)}$. The step length $\delta^{(q)}$ is computed to satisfy the Wolfe Conditions [30], which guarantee sufficient increase of Φ and discard unacceptably short steps.

The gradient $\nabla\Phi$ equals to $\nabla L + \gamma\nabla V$, and ∇L is a concatenation of its sub-gradients $\nabla_{\boldsymbol{\alpha}_l} L$ which are obtained by the chain rule,

$$\nabla_{\boldsymbol{\alpha}_l} L(\boldsymbol{\theta}) = \boldsymbol{J}(\bar{\boldsymbol{g}}_l(\boldsymbol{\alpha}_l))^\top \nabla\Lambda_l(\bar{\boldsymbol{g}}_l(\boldsymbol{\alpha}_l)),$$

where $\boldsymbol{J}(\cdot)$ denotes the Jacobian matrix of a vector with respect to $\boldsymbol{\alpha}$ and $\nabla\Lambda_l(\bar{\boldsymbol{g}}) = \boldsymbol{g}_l/\bar{\boldsymbol{g}} - \boldsymbol{1}$. The derivation of $\boldsymbol{J}(\bar{\boldsymbol{g}}_l(\boldsymbol{\alpha}_l))$ is given in Appendix A, (21).

D. Image Update

In absence of a penalty term $U(\boldsymbol{f})$ (i.e., $\beta = 0$), maximizing Φ with respect to $\boldsymbol{f}$ is performed with an ML expectation-maximization (ML-EM) algorithm [31], where the observable is the collection of the n_g gated PET data $(\boldsymbol{g}_l)_{l=1}^{n_\mathrm{g}}$, associated to their background event vectors $(\boldsymbol{s}_l)_{l=1}^{n_\mathrm{g}}$, and the system matrix is formed with the concatenation of all AC-MC PET system matrices:

$$\boldsymbol{H}_\mathrm{a}^{\mathrm{m,gated}}(\boldsymbol{\theta}, \boldsymbol{\mu}) \triangleq \begin{bmatrix} \tau_1 \boldsymbol{H}_\mathrm{a}^\mathrm{m}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}) \\ \vdots \\ \tau_{n_\mathrm{g}} \boldsymbol{H}_\mathrm{a}^\mathrm{m}(\boldsymbol{\alpha}_{n_\mathrm{g}}, \boldsymbol{\mu}) \end{bmatrix}.$$

The activity image $\boldsymbol{f}^{(p+1)}$ at iteration $p+1$ is obtained from $\boldsymbol{f}^{(p)}$ as follows:

$$\begin{aligned}\boldsymbol{f}^{(p+1)} &= \frac{\boldsymbol{f}^{(p)}}{\boldsymbol{\pi}(\boldsymbol{\theta}, \boldsymbol{\mu})} \sum_{l=1}^{n_\mathrm{g}} \tau_l \boldsymbol{H}_\mathrm{a}^\mathrm{m}(\boldsymbol{\alpha}_l, \boldsymbol{\mu})^\top \frac{\boldsymbol{g}_l}{\bar{\boldsymbol{g}}_l(\boldsymbol{f}^{(p)}, \boldsymbol{\alpha}_l, \boldsymbol{\mu})} \\ &\triangleq \boldsymbol{F}^\mathrm{em}(\boldsymbol{f}^{(p)})\end{aligned} \tag{18}$$

where we denoted $\boldsymbol{\pi}(\boldsymbol{\theta},\boldsymbol{\mu}) = \sum_{l=1}^{n_\mathrm{g}} \tau_l \boldsymbol{H}_\mathrm{a}^\mathrm{m}(\boldsymbol{\alpha}_l,\boldsymbol{\mu})^\top \mathbf{1}$. For a fixed $\boldsymbol{\theta}$, a sequence $(\boldsymbol{f}^{(p)})_{p\in\mathbb{N}}$ generated with (18) converges to a maximizer of Φ [32], [33].

Modified ML-EM algorithms (M-MLEM) to maximize the penalized log-likelihood have been proposed [34]–[37]. All cited work directly or indirectly require computation of the EM update $\boldsymbol{F}^\mathrm{em}(\boldsymbol{f}^{(p)},\boldsymbol{\theta},\boldsymbol{\mu})$ (18). We followed De Pierro approach [35] which consists of maximizing a surrogate function Q instead of Φ. The surrogate Q depends on the previous iterate $\boldsymbol{f}^{(p)}$ and separates the voxels, so that the image can be updated at each voxel independently. When U is a quadratic prior, maximizing Q reduces to finding the unique positive root of a second order polynomial, for each voxel j. For non-quadratic priors, separable paraboloid surrogates can be employed [36]. A complete description of our implementation can be found in Appendix B.

Algorithm Summary

We summarized the method in Algorithm 1. Some of the notations, such as the differentiated warping operator $\boldsymbol{W}_{\boldsymbol{\alpha}}^{\partial C}$ and the regularized image $\boldsymbol{F}^\mathrm{reg}$, are introduced in the appendices. The motion parameter $\boldsymbol{\theta}$ is initialized by $\mathbf{0}$. The image coefficient vector $\boldsymbol{f}$ is initialized by M-MLEM reconstruction from the first gate $\boldsymbol{g}_1$, using the system matrix $\tau_1\boldsymbol{H}_\mathrm{a}(\boldsymbol{\mu})$ and the background $\boldsymbol{s}_1$. The attenuation map $\boldsymbol{\mu}$ does not need to be aligned with $\boldsymbol{g}_1$ (cf. the discussion in Section II-D). In order to avoid artifacts on the image due to incomplete motion estimation, $\boldsymbol{f}$ is reinitialized to a blank image every r_reinit iterations. We used $r_\mathrm{reinit} = 5$, except for Evaluation 2 (Section IV-B) for which we used $r_\mathrm{reinit} = 1$. The L-BFGS computation of the line search direction $\boldsymbol{t}$ requires the gradient $\nabla\Phi$ and the estimates of $\boldsymbol{\theta}$ at several previous iterations, but they were omitted for compactness.

Proposition 1 requires the warping operators $\mathcal{W}_\varphi$ to be invertible. In practice, $\boldsymbol{W}_{\boldsymbol{\alpha}}$ is not invertible, not only because the B-spline mapping $\varphi_{\boldsymbol{\alpha}} : \mathbb{R}^3 \to \mathbb{R}^3$ is not invertible, but also because of the interpolating function w in (10). Thus the equality (6) does not hold in the discrete case. The presence of the penalty terms U and V also affects this result as Proposition 1 applies to the likelihood without penalty. Therefore, one of the objective of the evaluation is to verify that $(\hat{\boldsymbol{f}},\hat{\boldsymbol{\theta}}) = \arg\max_{\boldsymbol{f}\geq\mathbf{0},\boldsymbol{\theta}} \Phi(\boldsymbol{f},\boldsymbol{\theta},\boldsymbol{\mu})$ and $(\tilde{\boldsymbol{f}},\tilde{\boldsymbol{\theta}}) = \arg\max_{\boldsymbol{f}\geq\mathbf{0},\boldsymbol{\theta}} \Phi(\boldsymbol{f},\boldsymbol{\theta},\tilde{\boldsymbol{\mu}})$ satisfy

$$\forall l, \quad \boldsymbol{W}_{\hat{\boldsymbol{\alpha}}_l}\hat{\boldsymbol{f}} \approx \boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}_l}\tilde{\boldsymbol{f}} \quad \text{and} \quad \boldsymbol{W}_{\hat{\boldsymbol{\alpha}}_l}\boldsymbol{\mu} \approx \boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}_l}\tilde{\boldsymbol{\mu}} \tag{19}$$

for small values of β and γ.

Algorithm 1: Joint PET penalized maximum-likelihood motion estimation/image reconstruction

Input: Gated PET data $(\boldsymbol{g}_l)_{l=1}^{n_\mathrm{g}}$, attenuation map $\boldsymbol{\mu}$, image and motion smoothing priors β and γ
Output: PET image coefficients $\boldsymbol{f}$, B-spline motion parameter $\boldsymbol{\theta}$
initialization ;
$\boldsymbol{\theta} \leftarrow \mathbf{0}$;
$\boldsymbol{f} \leftarrow$ M-MLEM$(\boldsymbol{g}_1,\boldsymbol{\mu},\beta)$;
for $r = 1,\ldots,\mathrm{MaxIter}$ **do**
  motion estimation ;
  for $q = 1,\ldots,q_\mathrm{max}$ **do**
    for $l = 1,\ldots,n_\mathrm{g}$ **do**
      for $C \in \{X,Y,Z\}$, $\boldsymbol{h} \in \{\boldsymbol{f},\boldsymbol{\mu}\}$ **do**
        $\boldsymbol{J}_{\boldsymbol{h}}^C \leftarrow \mathrm{diag}\left\{\boldsymbol{W}_{\boldsymbol{\alpha}_l}^{\partial C}\boldsymbol{h}\right\}\boldsymbol{\mathcal{B}}$;
      end
      $\boldsymbol{J}_{\boldsymbol{f}} \leftarrow \left[\boldsymbol{J}_{\boldsymbol{f}}^X,\boldsymbol{J}_{\boldsymbol{f}}^Y,\boldsymbol{J}_{\boldsymbol{f}}^Z\right]$;
      $\boldsymbol{J}_{\boldsymbol{\mu}} \leftarrow \left[\boldsymbol{J}_{\boldsymbol{\mu}}^X,\boldsymbol{J}_{\boldsymbol{\mu}}^Y,\boldsymbol{J}_{\boldsymbol{\mu}}^Z\right]$;
      $\nabla\Lambda_l \leftarrow \boldsymbol{g}_l/\overline{\boldsymbol{g}}_l(\boldsymbol{f},\boldsymbol{\alpha}_l,\boldsymbol{\mu}) - \mathbf{1}$;
      $\nabla_{\boldsymbol{\alpha}_l} L \leftarrow -\tau_l \boldsymbol{J}_{\boldsymbol{\mu}}^\top \boldsymbol{L}^\top \mathrm{diag}\{\boldsymbol{H}_\mathrm{a}^\mathrm{m}(\boldsymbol{\alpha}_l,\boldsymbol{\mu})\boldsymbol{f}\}\nabla\Lambda_l +$
      $\tau_l \boldsymbol{J}_{\boldsymbol{f}}^\top \boldsymbol{H}_\mathrm{a}(\boldsymbol{W}_{\boldsymbol{\alpha}_l}\boldsymbol{\mu})^\top \nabla\Lambda_l$;
    end
    $\nabla\Phi \leftarrow [\nabla_{\boldsymbol{\alpha}_1} L^\top,\ldots,\nabla_{\boldsymbol{\alpha}_{n_\mathrm{g}}} L^\top]^\top + \gamma\nabla V(\boldsymbol{\theta})$;
    $\boldsymbol{t} \leftarrow \mathrm{LBFGS}(\nabla\Phi,\boldsymbol{\theta})$;
    $\delta^\star \leftarrow \arg\max_{\delta\geq 0} \Phi(\boldsymbol{\theta}+\delta\boldsymbol{t})$;
    $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \delta^\star\boldsymbol{t}$;
  end
  image reconstruction ;
  if $\mathrm{mod}(r, r_\mathrm{reinit}) = 0$ **then**
    $\boldsymbol{f} \leftarrow \mathbf{1}$;
  end
  $\boldsymbol{\pi} \leftarrow \sum_{l=1}^{n_\mathrm{g}} \boldsymbol{H}_\mathrm{a}^\mathrm{m}(\boldsymbol{\alpha}_l,\boldsymbol{\mu})^\top\mathbf{1}$;
  for $p = 1,\ldots,p_\mathrm{max}$ **do**
    $\boldsymbol{F}^\mathrm{em} \leftarrow \frac{\boldsymbol{f}}{\boldsymbol{\pi}} \sum_{l=1}^{n_\mathrm{g}} \boldsymbol{H}_\mathrm{a}^\mathrm{m}(\boldsymbol{\alpha}_l,\boldsymbol{\mu})^\top \frac{\boldsymbol{g}_l}{\overline{\boldsymbol{g}}_l(\boldsymbol{f},\boldsymbol{\alpha}_l,\boldsymbol{\mu})}$;
    if $\beta > 0$ **then**
      $\boldsymbol{F}^\mathrm{reg} \leftarrow \boldsymbol{F}^\mathrm{reg}(\boldsymbol{f})$;
      $\boldsymbol{f} \leftarrow \arg\max_{\boldsymbol{f}\geq\mathbf{0}} Q^L(\boldsymbol{f};\boldsymbol{F}^\mathrm{em}) + Q^U(\boldsymbol{f};\boldsymbol{F}^\mathrm{reg})$;
    else
      $\boldsymbol{f} \leftarrow \boldsymbol{F}^\mathrm{em}$;
    end
  end
end

IV. Evaluation

Some images in this section contain a horizontal line, in order to help motion/mismatches visualization. These lines are not related to reconstruction profile plots.

A. Reconstruction Methods

In this section we briefly summarize the method we utilized for evaluation. The methods used differ for each experiment. The interpolating function $w(\boldsymbol{r}) = e(x)e(y)e(z)$ used for the finite-dimensional representation (9) were the same B-splines functions used for $\mathcal{B}(\boldsymbol{r}) = b(x)b(y)b(z)$ in (11), although the former is defined on the voxel grid $\mathcal{G}$ and the latter is defined on the sub-grid $\tilde{\mathcal{G}}$.

1) JRM: We performed JRM by maximizing the penalized log-likelihood Φ with two different μ-maps: $\boldsymbol{\mu} = \boldsymbol{\mu}_1$, consistent with the first gate $\boldsymbol{g}_1$ (JRM1) and with $\tilde{\boldsymbol{\mu}}$, misaligned with each gate (JRM2):

$$(\hat{\boldsymbol{f}},\hat{\boldsymbol{\theta}}) \triangleq \underset{\boldsymbol{f}\geq\mathbf{0},\boldsymbol{\theta}}{\arg\max}\, \Phi(\boldsymbol{f},\boldsymbol{\theta},\boldsymbol{\mu}_1), (\tilde{\boldsymbol{f}},\tilde{\boldsymbol{\theta}}) \triangleq \underset{\boldsymbol{f}\geq\mathbf{0},\boldsymbol{\theta}}{\arg\max}\, \Phi(\boldsymbol{f},\boldsymbol{\theta},\tilde{\boldsymbol{\mu}}).$$

The reconstructed activity volumes at each gate $l = 1,\ldots,n_\mathrm{g}$ are denoted $\hat{\boldsymbol{f}}_l$ and $\tilde{\boldsymbol{f}}_l$,

$$\hat{\boldsymbol{f}}_l \triangleq \boldsymbol{W}_{\hat{\boldsymbol{\alpha}}_l}\hat{\boldsymbol{f}}, \quad \tilde{\boldsymbol{f}}_l \triangleq \boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}_l}\tilde{\boldsymbol{f}}.$$

2) PML Reconstruction of Each Gate Independently: This approach assumes that a collection of μ-maps $(\boldsymbol{\mu})_{l=1}^{n_\mathrm{g}}$ corresponding to the gated PET data $(\boldsymbol{g}_l)_{l=1}^{n_\mathrm{g}}$ is available. Each gate l is independently obtained by PML reconstruction (M-MLEM algorithm, [35]) from $\boldsymbol{g}_l$, using the system matrix $\tau_l\boldsymbol{H}_\mathrm{a}(\boldsymbol{\mu}_l)$ (14) and the background vector $\boldsymbol{s}_l$:

$$\hat{\boldsymbol{f}}_l^\mathrm{PML} = \underset{\boldsymbol{f}\geq\mathbf{0}}{\arg\max}\, \Lambda_l(\tau_l\boldsymbol{H}_\mathrm{a}(\boldsymbol{\mu}_l)\boldsymbol{f} + \boldsymbol{s}_l) + \tau_l\beta U(\boldsymbol{f}). \tag{20}$$

The reconstructed images depend on the parameter β.

(a) (b) (c)

Fig. 1. Evaluation 1: phantoms for μ-map realignment experiment: (a) $\boldsymbol{f}$; (b) $\boldsymbol{\mu}$; (c) $\tilde{\boldsymbol{\mu}}$.

3) Post-Reconstruction Registration and Consolidation (PRRC): PRRC also requires $(\boldsymbol{\mu})_{l=1}^{n_g}$ and consists of 4 steps:

(i) PML reconstruction following (20) to obtain $(\hat{\boldsymbol{f}}_l^{\mathrm{PML}})_{l=1}^{n_g}$

(ii) Selection of a reference gate l_0 and registration of the reconstructed volumes $(\hat{\boldsymbol{f}}_l^{\mathrm{PML}})_{l\neq l_0}$ to $\hat{\boldsymbol{f}}_{l_0}^{\mathrm{PML}}$. The registered volumes are denoted $\hat{\boldsymbol{f}}_l^{\mathrm{reg}}$, $l \neq l_0$.

(iii) Consolidation by averaging a single volume $\hat{\boldsymbol{f}}_{l_0}^{\mathrm{PRRC}}$:

$$\hat{\boldsymbol{f}}_{l_0}^{\mathrm{PRRC}} = \frac{1}{\sum_{l=1}^{n_g} \tau_l} \left(\tau_{l_0} \hat{\boldsymbol{f}}_{l_0}^{\mathrm{PML}} + \sum_{l \neq l_0} \tau_l \hat{\boldsymbol{f}}_l^{\mathrm{reg}} \right).$$

(iv) Re-warping of $\hat{\boldsymbol{f}}_{l_0}^{\mathrm{PRRC}}$ at each gate $l \neq l_0$ using the inverse motion. The output images are denoted $(\hat{\boldsymbol{f}}_l^{\mathrm{PRRC}})_{l=1}^{n_g}$.

PRRC utilizes all μ-maps, whereas JRM needs only one ($\boldsymbol{\mu}_1$ or $\tilde{\boldsymbol{\mu}}$). Therefore, PRRC should have a potential advantage over JRM.

4) JRM With Fixed μ-map (JRM-FM, Simulations Only): This approach consists of maximizing a modified version of Φ, denoted Φ^{fixMu}, obtained by replacing $\boldsymbol{W}_{\boldsymbol{\alpha}_l}\boldsymbol{\mu}$ with $\boldsymbol{\mu}$ in the expression of $\bar{\boldsymbol{g}}_l$ (12):

$$(\hat{\boldsymbol{f}}^{\mathrm{FM}}, \hat{\boldsymbol{\theta}}^{\mathrm{FM}}) \triangleq \underset{\boldsymbol{f}\geq \boldsymbol{0}, \boldsymbol{\theta}}{\arg\max}\, \Phi^{\mathrm{fixMu}}(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu}_1).$$

The gradient used for motion update is obtained using the Jacobian (21) without the first term. The reconstructed gates are denoted $\hat{\boldsymbol{f}}_l^{\mathrm{FM}} \triangleq \boldsymbol{W}_{\hat{\boldsymbol{\alpha}}_l^{\mathrm{FM}}} \hat{\boldsymbol{f}}^{\mathrm{FM}}$.

5) No Motion Compensation: This approach consists of a PML reconstruction of a single volume from the entire dataset $(\boldsymbol{g}_l)_{l=1}^{n_g}$, ignoring the motion and using a single attenuation map. The reconstructed volume is denoted $\hat{\boldsymbol{f}}_{\mathrm{noMC}}$.

6) Motion-Free Data (Simulations Only): For this approach, data are simulated without motion with the same number of counts. The reconstructed image, denoted $\hat{\boldsymbol{f}}_{\mathrm{static}}$, is obtained using PML with a consistent μ-map.

B. Evaluation 1: JRM μ-map Realignment On Noise-Free Simulated Data

To verify the analysis of Section II-D, we applied JRM in a single gate noise-free experiment. The activity $\boldsymbol{f}$ (Fig. 1(a)) and the corresponding aligned μ-map (Fig. 1(b)) are volumes generated from the XCAT phantom, cropped to a $113 \times 113 \times 42$ box containing the chest (3.125 mm edge cubic voxels). A deep inspiration μ-map, denoted $\tilde{\boldsymbol{\mu}}$ (Fig. 1(c)), was also generated. The PET projector $\boldsymbol{H}$ models a 5 mm FWHM point spread function for resolution. The same projector was used for $\boldsymbol{L}$ (projection of the μ-map). A noise-free sinogram $\boldsymbol{g}$ was generated by projection of $\boldsymbol{f}$ with a system response attenuated by $\boldsymbol{\mu}$ and a uniform background $\boldsymbol{s}$:

$$\boldsymbol{g} = \boldsymbol{H}_{\mathrm{a}}(\boldsymbol{\mu})\boldsymbol{f} + \boldsymbol{s}.$$

This model is a sub-case of (13) with $\boldsymbol{W} = \boldsymbol{I}$ (identity matrix). The motion parameter $\boldsymbol{\theta}$ is a single B-spline coefficient vector $\boldsymbol{\alpha}$. We estimated $(\tilde{\boldsymbol{f}}, \tilde{\boldsymbol{\alpha}})$ from $\boldsymbol{g}$ and $\tilde{\boldsymbol{\mu}}$ with $\beta = 0$ and γ set to a small value. We used a $n_{\mathrm{c}} = 40 \times 40 \times 16$ control points grid for the B-spline motion parametrization, and $\boldsymbol{f}$ was re-initialized after each iteration (i.e., $r_{\mathrm{reinit}} = 1$). Our aim is to verify whether the approximation (19) holds. In this particular case, it reduces to

$$\boldsymbol{f} \approx \boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}}\tilde{\boldsymbol{f}} \quad \text{and} \quad \boldsymbol{\mu} \approx \boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}}\tilde{\boldsymbol{\mu}}.$$

Figs. 2 and 3 show the warped estimated activities $\boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}}\tilde{\boldsymbol{f}}$ and the relative differences $\boldsymbol{\mu} - \boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}}\tilde{\boldsymbol{\mu}}$ at iterations 0 (reconstruction with $\tilde{\boldsymbol{\alpha}} = \boldsymbol{0}$), 1, 5, 10, 30 and 100. Each iteration corresponds to one realization of the outer loop $r = 1, \ldots, \mathrm{maxIter}$ in Algorithm 1. Results show that after 100 iterations the μ-map misalignment artifacts present in the initialization (near the liver and the myocardium) have disappeared. Also the warped attenuation $\boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}}\tilde{\boldsymbol{\mu}}$ is similar to $\boldsymbol{\mu}$, with some discrepancies on organ boundaries which can be attributed to partial volume effects, and also on the ribs due to the inability of the B-spline motion model to describe discontinuities in the motion field. These results show that JRM is a viable method for misaligned μ-map correction.

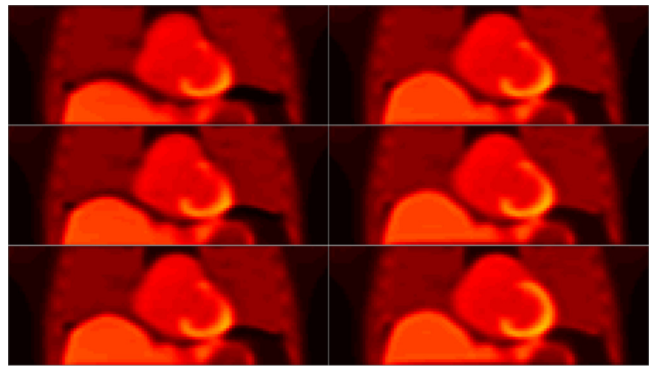


Fig. 2. Evaluation 1: from top left to bottom right: warped estimated activity $\boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}}\tilde{\boldsymbol{f}}$ at iterations 0 ($\tilde{\boldsymbol{\alpha}} = \boldsymbol{0}$), 1, 5, 10, 30 and 100.

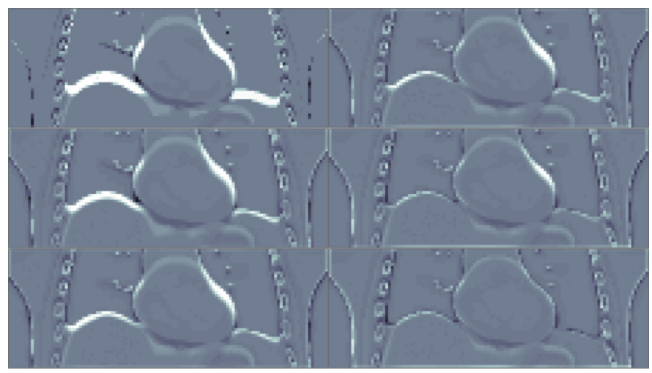
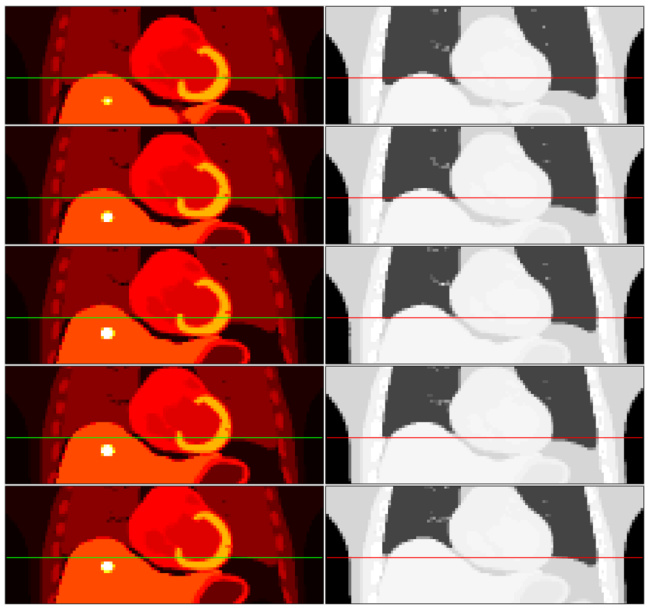
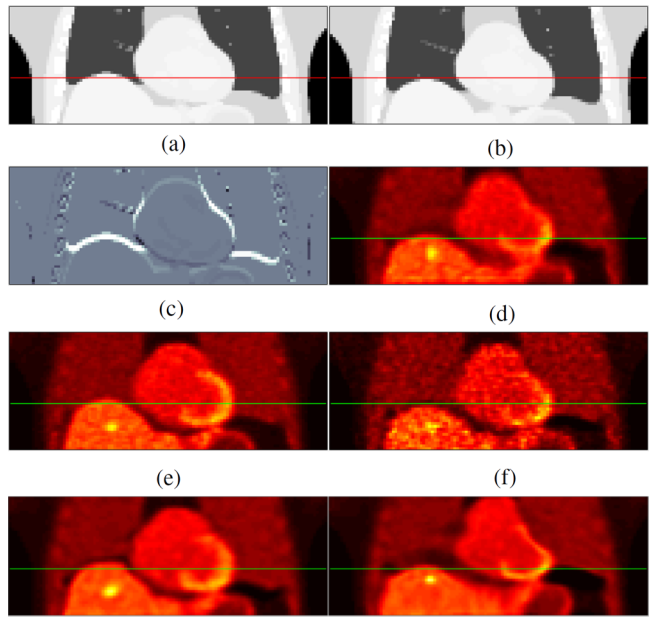


Fig. 3. Evaluation 1: from top left to bottom right: μ-maps relative difference $\boldsymbol{\mu} - \boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}}\tilde{\boldsymbol{\mu}}$ at iterations 0 ($\tilde{\boldsymbol{\alpha}} = \boldsymbol{0}$), 1, 5, 10, 30 and 100.



(a) (b)

Fig. 4. Evaluation 2: (a) top to bottom: XCAT phantom activity gates $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_5$; (b) top to bottom: corresponding XCAT phantom μ-maps $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_5$.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 5. Evaluation 2: (a) attenuation map $\boldsymbol{\mu}_1$; (b) misaligned attenuation map $\tilde{\boldsymbol{\mu}}$; (c) relative difference $\boldsymbol{\mu}_1 - \tilde{\boldsymbol{\mu}}$; (d) $\hat{\boldsymbol{f}}_{\text{noMC}}$: reconstruction using all the gates without motion compensation; (e) $\hat{\boldsymbol{f}}_{\text{static}}$: reconstruction from motion-free PET data; (f) $\tilde{\boldsymbol{f}}_1$: JRM reconstruction with $\tilde{\boldsymbol{\mu}}$ (gate 1) at iteration 0 (initialization); (g) $\hat{\boldsymbol{f}}_1^{\text{FM}}$: JRM-FM gate 1; (h) $\hat{\boldsymbol{f}}_5^{\text{FM}}$: JRM-FM gate 5.

C. Evaluation 2: Gated PET/CT Simulated Data

We generated $n_{\text{g}} = 5$ XCAT activity volumes $(\boldsymbol{f}_l)_{l=1}^5$ and their corresponding μ-maps $(\boldsymbol{\mu}_l)_{l=1}^5$ (Fig. 4). Each volume corresponds to one gate of the respiratory cycle ($l = 1$: inspiration, $l = 5$: expiration). Cardiac contraction was not simulated. In addition, a deep inspiration μ-map, $\tilde{\boldsymbol{\mu}}$ (Fig. 5(b)), was generated. This μ-map differs from $\boldsymbol{\mu}_1$ (see Fig. 5(c)) and from all other $(\boldsymbol{\mu}_l)_{l=2}^5$. A hot lesion was added to the activity volumes. We used same projectors and B-spline motion model as in Section IV-B.

$K = 50$ independent realizations of the PET gated projections $(\boldsymbol{g}_l)_{l=1}^5$ were simulated with $\tau_l = \tau$ for all l as

$$\boldsymbol{g}_l \sim \text{Poisson}(\tau \boldsymbol{H}_{\text{a}}(\boldsymbol{\mu}_l)\boldsymbol{f}_l + \boldsymbol{s}_l),$$

the background $\boldsymbol{s}_l$ being set to a uniform vector. A single realization of $(\boldsymbol{g})_{l=1}^{n_{\text{g}}}$ totals 10^8 counts (including background events). The activity volume was first reconstructed without motion correction ($\hat{\boldsymbol{f}}_{\text{noMC}}$, Fig. 5(d)) and from motion-free data ($\hat{\boldsymbol{f}}_{\text{static}}$, Fig. 5(e)). $\hat{\boldsymbol{f}}_{\text{static}}$ was obtained by PML reconstruction from the motion-free data from gate 1, generated as

$$\boldsymbol{g}_{\text{static}} \sim \text{Poisson}(5 \times (\tau \boldsymbol{H}_{\text{a}}(\boldsymbol{\mu}_1)\boldsymbol{f}_1 + \boldsymbol{s}_1)).$$

The former suffers from both motion and attenuation mismatch artifacts whereas the latter is artifact-free. JRM-FM first and fifth gates, $\hat{\boldsymbol{f}}_1^{\text{FM}} = \boldsymbol{W}_{\hat{\boldsymbol{\alpha}}_1^{\text{FM}}}\hat{\boldsymbol{f}}^{\text{FM}}$ and $\hat{\boldsymbol{f}}_5^{\text{FM}} = \boldsymbol{W}_{\hat{\boldsymbol{\alpha}}_5^{\text{FM}}}\hat{\boldsymbol{f}}^{\text{FM}}$, are shown in Figs. 5(g) and 5(h) respectively. $\hat{\boldsymbol{f}}_1^{\text{FM}}$ and more particularly $\hat{\boldsymbol{f}}_5^{\text{FM}}$ suffer from severe attenuation mismatch artifacts. This demonstrate that the μ-map needs to be warped alongside the activity.

We processed the $K = 50$ realizations of $(\boldsymbol{g}_l)_{l=1}^5$ with JRM1 ($\boldsymbol{\mu} = \boldsymbol{\mu}_1$) and JRM2 ($\boldsymbol{\mu} = \tilde{\boldsymbol{\mu}}$), as well PRRC, with $\beta = 0, 0.01$, 0.02, 0.03 and 0.1. The number of iterations was $\text{MaxIter} = 10$ for JRM1 and 30 for JRM2 (more iterations are necessary for JRM2 to correct for attenuation mismatches). The estimated activity volumes and motion are denoted $(\hat{\boldsymbol{f}}, \hat{\boldsymbol{\theta}})$ and $(\tilde{\boldsymbol{f}}, \tilde{\boldsymbol{\theta}})$. Because JRM2 was initialized with a PML reconstruction from the first gate dataset $\boldsymbol{g}_1$ with the AC system matrix $\tau \boldsymbol{H}_{\text{a}}(\tilde{\boldsymbol{\mu}})$ (Fig. 5(f)), the initial reconstruction suffers from misalignment artifacts on the liver and the myocardium. The PRRC reconstructed gates, $(\hat{\boldsymbol{f}}_l^{\text{PRRC}})_{l=1}^{n_{\text{g}}}$ were obtained using the same B-spline deformation model as JRM for post-reconstruction registration.

JRM and PRRC volumes (reconstructed with $\beta = 0.01$ and $\beta = 0.03$ respectively to match the variance, see Fig. 8) at each gate l, $\hat{\boldsymbol{f}}_l = \boldsymbol{W}_{\hat{\boldsymbol{\alpha}}_l}\hat{\boldsymbol{f}}$, $\tilde{\boldsymbol{f}}_l = \boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}_l}\tilde{\boldsymbol{f}}$ and $\hat{\boldsymbol{f}}_l^{\text{PRRC}}$ are shown in Figs. 6(a)–6(c). The 3 volume sequences appear similar, and the misalignment artifacts present in the initialization of $\tilde{\boldsymbol{f}}$ (Fig. 5(f)) have vanished. Reconstruction profiles were plotted along a section of the first gate within the coronal plane intersecting the lesion (Fig. 7). Results show that both JRM volumes $\hat{\boldsymbol{f}}_1$ and $\tilde{\boldsymbol{f}}_1$ are similar. It can be noticed on Fig. 7 that $\hat{\boldsymbol{f}}_1^{\text{PRRC}}$ has somewhat better defined organ edge boundaries compared to $\hat{\boldsymbol{f}}_1$ and $\tilde{\boldsymbol{f}}_1$. This is because $\hat{\boldsymbol{f}}_1$ and $\tilde{\boldsymbol{f}}_1$ used only one μ-map warped alongside the activity whereas $\hat{\boldsymbol{f}}_l^{\text{PRRC}}$ uses the entire exact sequence. As observed in Fig. 5(g), JRM-FM shows a sudden drop of activity near the liver.

We assessed the performance of JRM1, JRM2 and PRRC on the hot lesion on the first gate by plotting the mean square error

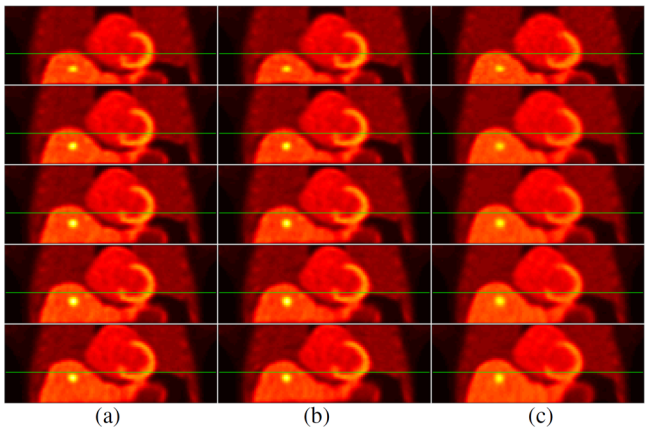


Fig. 6. Evaluation 2: reconstructed volumes at each gate $l = 1, \ldots, 5$ (from top to bottom). (a) $\hat{\boldsymbol{f}}_l = \boldsymbol{W}_{\hat{\boldsymbol{\alpha}}_l}\hat{\boldsymbol{f}}$ (using an aligned μ-map); (b) $\tilde{\boldsymbol{f}}_l = \boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}_l}\tilde{\boldsymbol{f}}$ (using a misaligned μ-map); (c) PRRC reconstructed gates $\hat{\boldsymbol{f}}_l^{\mathrm{PRRC}}$.

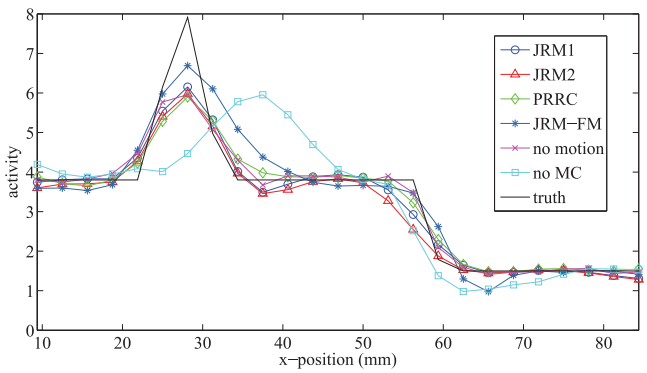


Fig. 7. Evaluation 2: reconstruction profiles along a section within the coronal plane (foot to head) intersecting the lesion (gate 1).

(MSE), in a small region R containing the lesion, against the total image variance (based on the K realizations of $(\boldsymbol{g}_l)_{l=1}^{n_{\mathrm{g}}}$), for each value of β:

$$\mathrm{MSE} = \frac{1}{|R|}\frac{1}{K}\sum_{j\in R}\sum_{\kappa=1}^{K}([\boldsymbol{h}^{[\kappa]}]_j - [\boldsymbol{f}_1]_j)^2,$$

$$\mathrm{Var} = \frac{1}{n_{\mathrm{v}}}\frac{1}{K-1}\sum_{j=1}^{n_{\mathrm{v}}}\sum_{\kappa=1}^{K}([\boldsymbol{h}^{[\kappa]}]_j - m_j(\boldsymbol{h}))^2,$$

with

$$m_j(\boldsymbol{h}) = \frac{1}{K}\sum_{\kappa=1}^{K}[\boldsymbol{h}^{[\kappa]}]_j$$

where $\boldsymbol{h}$ is either $\hat{\boldsymbol{f}}_1$, $\tilde{\boldsymbol{f}}_1$ or $\hat{\boldsymbol{f}}_1^{\mathrm{PRRC}}$ and the superscript $^{[\kappa]}$ denotes the noise realization. The 3 curves in Fig. 8 show that JRM1 achieves lower MSE for any variance level. JRM2 performs somewhere in between JRM1 and PRRC.

The performance of JRM can also be assessed by investigating the warped μ-map with the estimated motion field. Fig. 9 shows the warped μ-maps $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}_l}\boldsymbol{\mu}_1$ and $\boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}_l}\tilde{\boldsymbol{\mu}}$. Both volume sequences appear similar to the ground truth $\boldsymbol{\mu}_l$ (right column), which is consistent with the analysis of Section II-D.

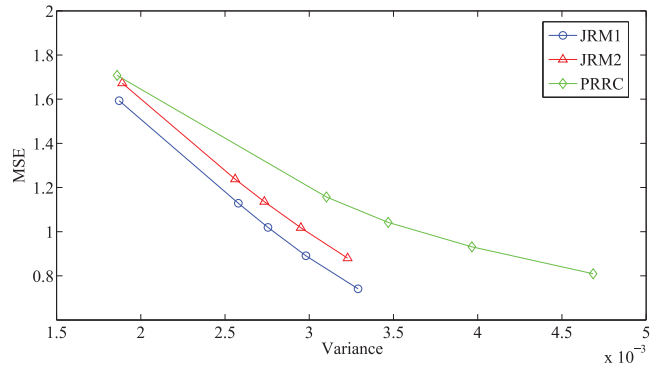


Fig. 8. Evaluation 2: MSE vs variance in the lesion for $\beta = 0, 0.01, 0.02, 0.03$ and 0.1.

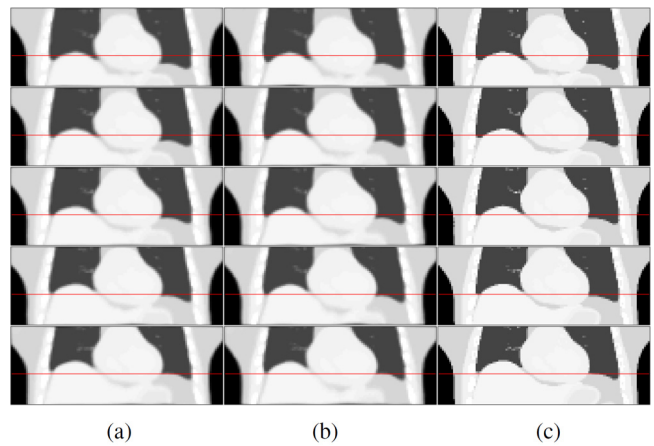


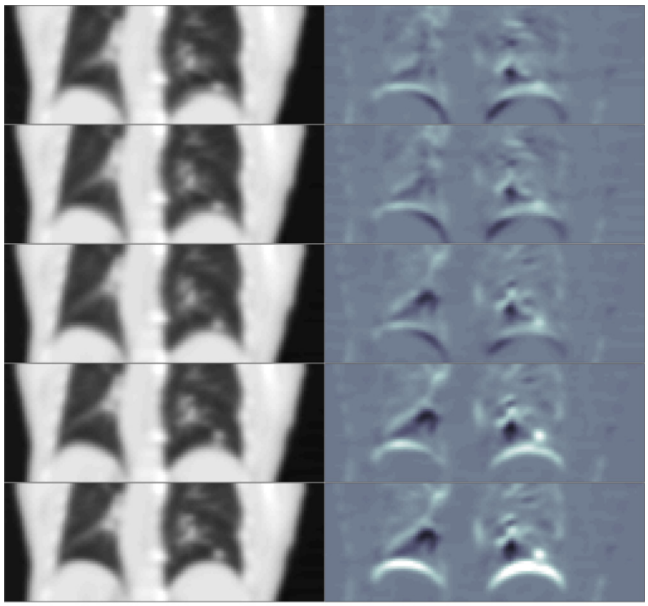
Fig. 9. Evaluation 2: warped μ-maps at each gate $l = 1, \ldots, 5$ (top to bottom). (a) $\boldsymbol{W}_{\hat{\boldsymbol{\alpha}}_l}\boldsymbol{\mu}_1$; (b) $\boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}_l}\tilde{\boldsymbol{\mu}}$; (c) ground truth $\boldsymbol{\mu}_l$ (for comparison).

D. Evaluation 3: Patient Data

Patient data for a clinical FDG PET/CT study were acquired on a GE Discovery STE [38]. 315 MBq of ^{18}F-FDG was injected 1 hour before the scan start according to normal clinical protocol. A cine-CT scan was performed to cover the lung PET bed position (140 kVp, 60 mA, 4 s duration, 0.5 s rotation period, 0.45 s time between reconstructed images, 9 bed positions, 8 axial slices (thickness 2.5 mm) per bed position), followed by a PET scan with list mode enabled. The patient was monitored with the Varian RPM system.

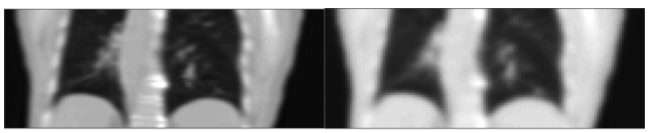
PET data were binned into 5 gates according to the value of the RPM signal at the detection time of the event. Gated CT images were obtained by using a weighted average of the cine-CT images where the weight was determined from the amount of overlap between the time interval during which the cine-CT slice was acquired and the time period that the RPM signal was in the range for each gate. Matching ranges were used for the gating of the cine-CT and PET data in an attempt to obtain spatially matched gated PET and CT data. The gate (fractional) durations were $\tau_1 = 0.0583$, $\tau_2 = 0.1276$, $\tau_3 = 0.2446$, $\tau_4 = 0.3405$ and $\tau_5 = 0.2290$.

The cine-CT derived μ-maps, $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_5$ are shown in Fig. 10(a). The μ-maps were down-sampled to the PET volume sizes ($192 \times 192 \times 47$). A μ-map, denoted $\tilde{\boldsymbol{\mu}}$ (Fig. 11(b)), was derived from the High Resolution CT (HRCT), acquired at full



(a) (b)

Fig. 10. Evaluation 3: (a) gated cine-CT μ-maps, $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_5$; (b) relative difference $\boldsymbol{\mu}_l - \tilde{\boldsymbol{\mu}}$ between gated cine-CT and HRCT-derived $\tilde{\boldsymbol{\mu}}$ (Fig. 11(b)).

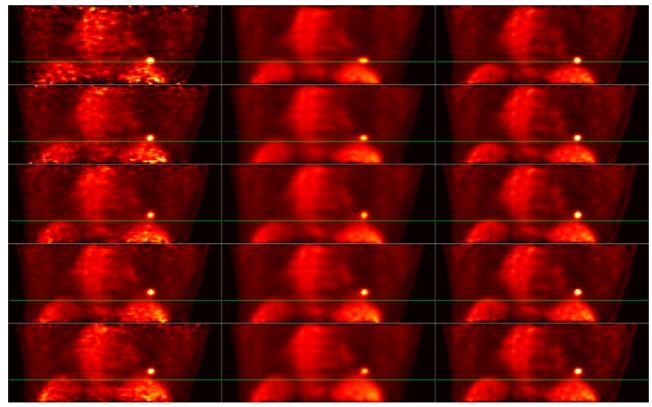


(a) (b)

Fig. 11. Evaluation 3: (a) HRCT reconstruction of the chest; (b) $\tilde{\boldsymbol{\mu}}$: HRCT-derived μ-map.

inspiration in helical mode (140 kVp, 148 mA, 0.8 s rotation period, pitch 1.375, slice thickness 3.27 mm) (Fig. 11(a)). This μ-map differs from the cine-CT μ-maps at each gate (see Fig. 10(b)).

We used the GE proprietary software for offline data processing (projection/backprojection, scatter/random estimation). Scatter was estimated using the cine-CT gated attenuation maps for PRRC and PML, and using the HRCT only for JRM. We proceeded with PRRC and individual PML reconstruction of each gate. The reconstructed volumes $\hat{\boldsymbol{f}}_l^{\mathrm{PML}}$ and $\hat{\boldsymbol{f}}_l^{\mathrm{PRRC}}$ are shown in Figs. 12(a) and 12(b). The gate with largest duration $l_0 = 4$ was chosen as reference for PRRC. We used the same $65 \times 65 \times 16$ control point grid for PRRC and JRM. JRM was achieved by maximization of Φ with $\boldsymbol{\mu} = \tilde{\boldsymbol{\mu}}$. The reconstructed gates, i.e., $\tilde{\boldsymbol{f}}_l = \boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}_l} \tilde{\boldsymbol{f}}$ are shown in Fig. 12(c). PML volumes suffer from noise as the gates were reconstructed individually. Both PML and PRRC also suffer from "partial volume effects" on the organ edges (liver and stomach) that can be attributed to the poor quality of the cine-CT μ-maps. In contrast, organs in JRM volumes (reconstructed using the HRCT-derived μ-map) appear sharper. A reconstruction profile along a section within the coronal plane and intersecting the hot lesion is shown in Fig. 13. It can be observed that the uptake on the hot lesion is higher with JRM. A non motion-corrected profile was plotted for comparison.



(a) (b) (c)

Fig. 12. Evaluation 3: reconstructed activity volumes at each gate $l = 1, \ldots, 5$ (top to bottom): (a) individually PML reconstructed gates $\hat{\boldsymbol{f}}_l^{\mathrm{PML}}$ (using cine-CT sequence, see Fig. 10(a)); (b) PRRC reconstructed gates $\hat{\boldsymbol{f}}_l^{\mathrm{PRRC}}$ (using cine-CT sequence, see Fig. 10(a)); (c) JRM reconstructed gates $\tilde{\boldsymbol{f}}_l$ (using $\tilde{\boldsymbol{\mu}}$, see Fig. 11(b)).

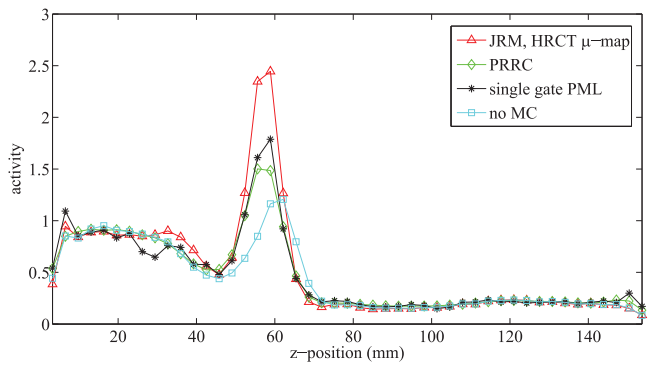


Fig. 13. Evaluation 3: reconstructed volume profiles (gate $l = 3$) along a section within the coronal plane (foot to head), intersecting the hot lesion.

For these patient data, there is no real approach to assess the JRM ability to warp $\tilde{\boldsymbol{\mu}}$ to the correct position because the true attenuation at each gate is unknown. We compared the warped μ-maps obtained with JRM with the gated cine-CT μ-maps (Fig. 14), but these are not completely reliable as mentioned in the introduction. Nevertheless, it can be noted that none of the JRM reconstructed gates seem to suffer from μ-map misalignment artifacts, suggesting that $\boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}_l}$ warps $\tilde{\boldsymbol{\mu}}$ appropriately.

V. Discussion

The major information is that the actual object of interest is not $\boldsymbol{f}$ but $\boldsymbol{W}_{\boldsymbol{\alpha}}\boldsymbol{f}$. It was mentioned in [39], Section 7.2.1, that incorporating the attenuation map is challenging because the coefficients vector $\boldsymbol{f}$ is "virtual" (only the warped version $\boldsymbol{W}_{\boldsymbol{\alpha}}\boldsymbol{f}$ matters) and therefore there is no corresponding μ-map. In fact, JRM estimates a "virtual" activity image and a deformation operator that accounts for both misalignments and patient motion. When the reconstruction is performed using a misaligned μ-map $\tilde{\boldsymbol{\mu}}$, the "virtual" activity image $\tilde{\boldsymbol{f}}$ is reconstructed in the "$\tilde{\boldsymbol{\mu}}$-space". Applying the estimated warping operator $\boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}_l}$ realigns $\tilde{\boldsymbol{f}}$ to the unobserved $\boldsymbol{\mu}$ and performs the motion correction to gate l. This is a direct consequence of Proposition 1, (6) and (7).

In this paper, JRM with misaligned μ-map requires a large number of iterations, indicating that for non-TOF-PET data the problem is ill-posed. However, since JRM uses a known μ-map, it is does not suffer from the same cross-talk issues as for joint reconstruction of activity and attenuation in non-TOF-PET [20]–[22]. TOF-PET data would likely accelerate convergence of JRM but we leave this for future work.

In our results based on the simulated data, the reconstructed gates with JRM appear free from motion/attenuation misalignment artifacts. Our quantitative analysis of MSE and variance shows that JRM outperforms PRRC in the quantification of a hot lesion. The fact that the deformed μ-map follows the deformed activity shows that gated CT is not necessary. With a misaligned μ-map, the reconstructed gates are the same, confirming the analysis of Section II-D, although reconstructing with a μ-map which is not aligned with any of the PET gates requires additional iterations. This result is important as in many PET/CT studies, a diagnostic CT is acquired during breath-hold. The result using a single PET gate suggests that JRM can be used for attenuation/PET mismatch correction in situations beyond respiratory gating.

The assessment of JRM on real data is more problematic because of the absence of ground-truth. The only references are the cine-CT gated μ-maps but they are prone to gating errors due to inter-cycle variation [8]. However, JRM with an HRCT-derived μ-map leads to artifact-free images, thus showing that $\boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}_l}$ warps both the activity volume and μ-map appropriately. As it can be seen in Fig. 11(a), the lesion appears smaller in the HRCT than in the cine-CT, probably because of motion during the acquisition. However, this seems to have had a negligible effect on the PET reconstruction.

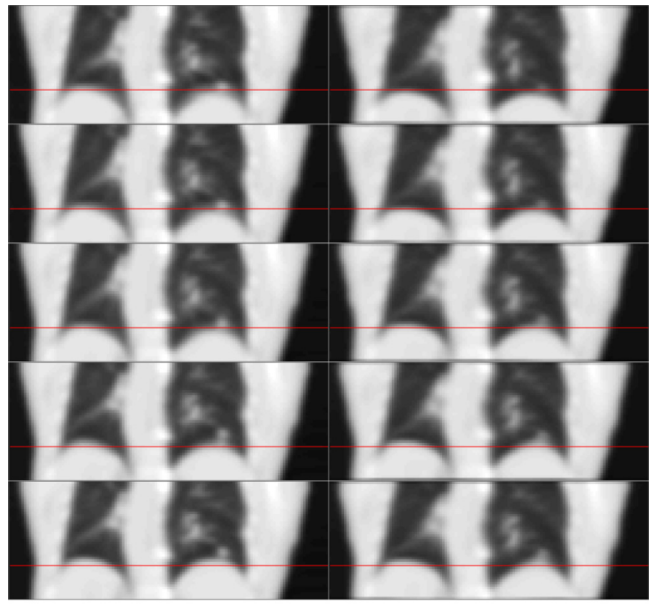
It should be noted that for real data, an accurate approximation of the background term $\boldsymbol{s}$ (scatter and coincidences) is of major importance. Ignoring this term leads to mismatches between $\boldsymbol{g}_l$ (observed counts) and $\bar{\boldsymbol{g}}_l$ (expected counts) that JRM will try to compensate with $\varphi_{\boldsymbol{\alpha}}$. For example, if $\bar{\boldsymbol{g}}_l$ is underestimated, JRM tries to compensate for the missing counts by enlarging the lungs (low attenuation) to artificially increase the number of expected counts. Accurate gate durations τ_l is also important for the same reasons.

We have not investigated the motion smoothness parameter γ. In preliminary experiments (results not shown), JRM with too small γ values led to an irregular motion field $\boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}}$ and "broken" images $\tilde{\boldsymbol{f}}$. Nevertheless, the product $\boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}}\tilde{\boldsymbol{f}}$ was largely unaffected, as well as the warped attenuation. This is another illustration that the warped images are the final result of the image reconstruction, not the "virtual" image.

The proposed algorithm is monotonic by definition, in the sense that each iteration r increases the penalized likelihood Φ. However it is not possible to demonstrate strict convergence (existence and uniqueness of a limit $(\boldsymbol{f}^\star, \boldsymbol{\theta}^\star)$) because of the non-concavity of $(\boldsymbol{f}, \boldsymbol{\theta}) \mapsto \Phi(\boldsymbol{f}, \boldsymbol{\theta})$.

VI. Conclusion

We demonstrated that it is possible to extend existing JRM methods [23], [24] to a more complete model that includes an attenuation map affected by the same motion. We demonstrated that the reconstructed gates are independent of the choice of the input μ-map, provided it results from the deformation of a common μ-map. We proposed an algorithm to monotonically maximize the penalized log-likelihood of the complete model. Results on XCAT simulated data showed that it is possible to use JRM to correct for misaligned attenuation and to reconstruct the activity from the entire dataset. Our algorithm applied on patient data successfully warped the breath-held HRCT μ-map and achieved similar reconstruction to PRRC reconstructions using cine-CT gated μ-maps. These results suggest that JRM can remove the need for cine-CT data for PET attenuation correction, with a corresponding reduction of patient dose.



(a) (b)

Fig. 14. Evaluation 3: gated cine-CT μ-maps and JRM-warped $\tilde{\boldsymbol{\mu}}$ at each gate $l = 1, \ldots, 5$ (top to bottom): (a) gated cine-CT μ-maps $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_5$; (b) warped $\tilde{\boldsymbol{\mu}}$, i.e., $\boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}_1}\tilde{\boldsymbol{\mu}}, \ldots, \boldsymbol{W}_{\tilde{\boldsymbol{\alpha}}_5}\tilde{\boldsymbol{\mu}}$.

Appendix A
Projection and Image Jacobian

The Jacobian of $\bar{\boldsymbol{g}}_l(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu})$ with respect to $\boldsymbol{\alpha}_l$ is

$$\boldsymbol{J}(\bar{\boldsymbol{g}}_l(\boldsymbol{f}, \boldsymbol{\alpha}_l, \boldsymbol{\mu})) = -\tau_l \mathrm{diag}\{\boldsymbol{H}_{\mathrm{a}}^{\mathrm{m}}(\boldsymbol{\alpha}_l, \boldsymbol{\mu})\boldsymbol{f}\}\boldsymbol{L}\boldsymbol{J}(\boldsymbol{W}_{\boldsymbol{\alpha}_l}\boldsymbol{\mu}) + \tau_l \boldsymbol{H}_{\mathrm{a}}(\boldsymbol{W}_{\boldsymbol{\alpha}_l}\boldsymbol{\mu})\boldsymbol{J}(\boldsymbol{W}_{\boldsymbol{\alpha}_l}\boldsymbol{f}) \quad (21)$$

where $\mathrm{diag}\{\boldsymbol{x}\}$ denotes the diagonal matrix generated from vector $\boldsymbol{x}$, $\boldsymbol{H}_{\mathrm{a}}$ and $\boldsymbol{H}_{\mathrm{a}}^{\mathrm{m}}$ are the AC and AC-MC system matrices defined in (14) and (15) respectively, $\boldsymbol{J}(\boldsymbol{W}_{\boldsymbol{\alpha}_l}\boldsymbol{f}) \in \mathbb{R}^{n_{\mathrm{v}} \times n_{\mathrm{w}}}$ and $\boldsymbol{J}(\boldsymbol{W}_{\boldsymbol{\alpha}_l}\boldsymbol{\mu}) \in \mathbb{R}^{n_{\mathrm{v}} \times n_{\mathrm{w}}}$ are the Jacobian matrices of the warped image coefficients $\boldsymbol{f}$ and $\boldsymbol{\mu}$ respectively. When $\boldsymbol{\mu} = \boldsymbol{0}$, (21) is equivalent to (28) and (29) in [24][1].

The Jacobian matrix $\boldsymbol{J}(\boldsymbol{W}_{\boldsymbol{\alpha}}\boldsymbol{f})$ (and similarly $\boldsymbol{J}(\boldsymbol{W}_{\boldsymbol{\alpha}_l}\boldsymbol{\mu})$) was already derived in [40], Chapter 5. We re-derived it with our notations. Recall the definition of $\boldsymbol{W}_{\boldsymbol{\alpha}}$ (10):

$$\forall \boldsymbol{f} \in \mathbb{R}^3, \quad [\boldsymbol{W}_{\boldsymbol{\alpha}}\boldsymbol{f}]_j = \sum_{k=1}^{n_{\mathrm{v}}} f_k w(\varphi_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - \boldsymbol{r}_k),$$

[1]In [24], $H(a, \cdot)$ is the system response at bin a and should not be mistaken with $\boldsymbol{H}_{\mathrm{a}}$, the attenuation corrected PET system matrix.

with $\varphi_{\boldsymbol{\alpha}} = (\varphi^X_{\boldsymbol{\alpha}}, \varphi^Y_{\boldsymbol{\alpha}}, \varphi^Z_{\boldsymbol{\alpha}}) : \mathbb{R}^3 \to \mathbb{R}^3$ defined in (11), and recall the definition of w:

$$\forall \boldsymbol{r} = (x, y, z) \in \mathbb{R}^3, \quad w(\boldsymbol{r}) = e(x)e(y)e(z).$$

Note that $\varphi^X_{\boldsymbol{\alpha}}$, $\varphi^Y_{\boldsymbol{\alpha}}$ and $\varphi^Z_{\boldsymbol{\alpha}}$ only depends on $\boldsymbol{\alpha}^X$, $\boldsymbol{\alpha}^Y$ and $\boldsymbol{\alpha}^Z$ respectively. By the chain rule, we have

$$\frac{\partial w(\varphi_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - \boldsymbol{r}_k)}{\partial \alpha^X_n} = \mathcal{B}\Big(\frac{\boldsymbol{r}_j - \tilde{\boldsymbol{r}}_n}{\sigma}\Big)\dot{e}(\varphi^X_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - x_k) \times e(\varphi^Y_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - y_k) \times e(\varphi^Z_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - z_k),$$

where $\dot{e}$ denotes the first-order derivative of e. The partial derivative of $[\boldsymbol{W}_{\boldsymbol{\alpha}}\boldsymbol{f}]_j$ w.r.t. α^X_n is

$$\frac{[\boldsymbol{W}_{\boldsymbol{\alpha}}\boldsymbol{f}]_j}{\partial \alpha^X_n} = \mathcal{B}\Big(\frac{\boldsymbol{r}_j - \tilde{\boldsymbol{r}}_n}{\sigma}\Big)\sum_{k=1}^{n_{\mathrm{v}}} f_k \dot{e}(\varphi^X_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - x_k) \times e(\varphi^Y_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - y_k) \times e(\varphi^Z_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - z_k)$$

(partial derivatives with respect to α^Y_n and α^Z_n are similarly obtained). Introducing the matrix $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{n_{\mathrm{v}} \times n_{\mathrm{c}}}$ defined as

$$[\boldsymbol{\mathcal{B}}]_{j,n} = \mathcal{B}\Big(\frac{\boldsymbol{r}_j - \tilde{\boldsymbol{r}}_n}{\sigma}\Big)$$

and $\boldsymbol{W}^{\partial X}_{\boldsymbol{\alpha}}$, $\boldsymbol{W}^{\partial Y}_{\boldsymbol{\alpha}}$ and $\boldsymbol{W}^{\partial Z}_{\boldsymbol{\alpha}} \in \mathbb{R}^{n_{\mathrm{v}} \times n_{\mathrm{v}}}$ defined as

$$\begin{aligned}
[\boldsymbol{W}^{\partial X}_{\boldsymbol{\alpha}}]_{j,k} &= \dot{e}(\varphi^X_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - x_k)e(\varphi^Y_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - y_k)e(\varphi^Z_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - z_k)\\
[\boldsymbol{W}^{\partial Y}_{\boldsymbol{\alpha}}]_{j,k} &= e(\varphi^X_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - x_k)\dot{e}(\varphi^Y_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - y_k)e(\varphi^Z_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - z_k)\\
[\boldsymbol{W}^{\partial Z}_{\boldsymbol{\alpha}}]_{j,k} &= e(\varphi^X_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - x_k)e(\varphi^Y_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - y_k)\dot{e}(\varphi^Z_{\boldsymbol{\alpha}}(\boldsymbol{r}_j) - z_k)
\end{aligned}$$

we can derive a matrix formulation:

$$\boldsymbol{J}(\boldsymbol{W}_{\boldsymbol{\alpha}}\boldsymbol{f}) = \Big[\mathrm{diag}\{\boldsymbol{W}^{\partial X}_{\boldsymbol{\alpha}}\boldsymbol{f}\}\boldsymbol{\mathcal{B}}, \mathrm{diag}\{\boldsymbol{W}^{\partial Y}_{\boldsymbol{\alpha}}\boldsymbol{f}\}\boldsymbol{\mathcal{B}}, \mathrm{diag}\{\boldsymbol{W}^{\partial Z}_{\boldsymbol{\alpha}}\boldsymbol{f}\}\boldsymbol{\mathcal{B}}\Big].$$

Appendix B
Separable Surrogates for PML Image Reconstruction

We adopted the methodology from [41]. As in Section III-C, $\Phi(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu})$ is rewritten $\Phi(\boldsymbol{f})$ and $V(\boldsymbol{\theta})$ is omitted:

$$\Phi(\boldsymbol{f}) = L(\boldsymbol{f}) + \beta U(\boldsymbol{f})$$

Φ is strictly concave if $\beta > 0$ and $\boldsymbol{g}_l^\top \boldsymbol{H}^{\mathrm{m}}_{\mathrm{a}}(\boldsymbol{\alpha}_l, \boldsymbol{\mu})\mathbf{1} \neq 0$ for at least one l (see [37], Section II-A1), so it has a unique maximizer in $\mathbb{R}^{n_{\mathrm{v}}}_+$.

A function $Q : \mathbb{R}^{n_{\mathrm{v}}}_+ \times \mathbb{R}^{n_{\mathrm{v}}}_+ \to \mathbb{R}$ is a surrogate for Φ if it verifies the following axioms:

1) $Q(\boldsymbol{f}; \boldsymbol{h}) - Q(\boldsymbol{h}; \boldsymbol{h}) \leq \Phi(\boldsymbol{f}) - \Phi(\boldsymbol{h}), \forall \boldsymbol{f}, \boldsymbol{h} \in \mathbb{R}^{n_{\mathrm{v}}}_+$
2) $\nabla_1 Q(\boldsymbol{h}; \boldsymbol{h}) = \nabla \Phi(\boldsymbol{h}), \boldsymbol{h} \in \mathbb{R}^{n_{\mathrm{v}}}_+$

where ∇_1 is the gradient with respect to the vector on the left of the semicolon. If for all $\boldsymbol{h} \in \mathbb{R}^{n_{\mathrm{v}}}_+$ the mapping $\boldsymbol{f} \mapsto Q(\boldsymbol{f}; \boldsymbol{h})$ has a (possibly non-unique) maximizer in $\mathbb{R}^{n_{\mathrm{v}}}_+$, then condition 1) guarantees that a sequence $(\boldsymbol{f}^{(p)})_{p \in \mathbb{N}}$ satisfying

$$Q(\boldsymbol{f}^{(p+1)}; \boldsymbol{f}^{(p)}) = \max_{\boldsymbol{f} \in \mathbb{R}^{n_{\mathrm{v}}}_+} Q(\boldsymbol{f}; \boldsymbol{f}^{(p)})$$

automatically verifies

$$\Phi(\boldsymbol{f}^{(p+1)}) \geq \Phi(\boldsymbol{f}^{(p)}).$$

Condition 2) serves to demonstrate convergence to a limit $\boldsymbol{f}^*$ satisfying the Karush-Kuhn-Tucker condition for Φ. We now summarize the framework proposed in [35] to derive separable surrogates for L and U.

Using the convexity inequality on log (see [35], (25)), it can be shown that Q^L defined as

$$Q^L(\boldsymbol{f}; \boldsymbol{F}^{\mathrm{em}}(\boldsymbol{f}^{(p)})) = \sum_{j=1}^{n_{\mathrm{v}}} \pi_j(\boldsymbol{\theta}, \boldsymbol{\mu}) \left(F^{\mathrm{em}}_j(\boldsymbol{f}^{(p)}) \log f_j - f_j\right),$$

where $F^{\mathrm{em}}_j(\boldsymbol{f}^{(p)}) = [\boldsymbol{F}^{\mathrm{em}}(\boldsymbol{f}^{(p)})]_j$ and $\pi_j(\boldsymbol{\theta}, \boldsymbol{\mu}) = [\boldsymbol{\pi}(\boldsymbol{\theta}, \boldsymbol{\mu})]_j$ are defined in Section III-D, is a surrogate for $\boldsymbol{f} \mapsto L(\boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{\mu})$. Using another convexity inequality on the regularizing term U (see [35], (31)), it can be shown that Q^U defined as

$$Q^U(\boldsymbol{f}; \boldsymbol{F}^{\mathrm{reg}}(\boldsymbol{f}^{(p)})) = -2\sum_{j=1}^{n_{\mathrm{v}}} \omega^{\mathrm{sum}}_j (f_j - F^{\mathrm{reg}}_j(\boldsymbol{f}^{(p)}))^2$$

where

$$F^{\mathrm{reg}}_j(\boldsymbol{f}^{(p)}) = \frac{1}{2\omega^{\mathrm{sum}}_j} \sum_{k \in \mathcal{N}_j} \omega_{j,k} \left(f^{(p)}_j + f^{(p)}_k\right)$$

and $\omega^{\mathrm{sum}}_j = \sum_{k \in \mathcal{N}_j} \omega_{j,k}$, is a surrogate for U. It has been noted in [41] that $\boldsymbol{F}^{\mathrm{reg}}(\boldsymbol{f}^{(p)}) = (F^{\mathrm{reg}}_j(\boldsymbol{f}^{(p)}))^{n_{\mathrm{v}}}_{j=1}$ can be seen as a regularized version of $\boldsymbol{f}^{(p)}$. Thus,

$$Q(\boldsymbol{f}; \boldsymbol{f}^{(p)}) \triangleq Q^L(\boldsymbol{f}; \boldsymbol{F}^{\mathrm{em}}(\boldsymbol{f}^{(p)})) + \beta Q^U(\boldsymbol{f}; \boldsymbol{F}^{\mathrm{reg}}(\boldsymbol{f}^{(p)}))$$

is a surrogate for Φ. More importantly, Q is a sum of terms depending on f_j only, so maximizing each of these terms with respect to f_j provides a maximizer for $\boldsymbol{f} \mapsto Q(\boldsymbol{f}; \boldsymbol{f}^{(p)})$. If $F^{\mathrm{em}}_j(\boldsymbol{f}^{(p)}) > 0$, $f^{(p+1)}_j$ is the unique strictly positive *root* of a second order polynomial in f_j. If $F^{\mathrm{em}}_j(\boldsymbol{f}^{(p)}, \boldsymbol{\theta}, \boldsymbol{\mu}) = 0$, $f^{(p+1)}_j$ is the unique positive *maximizer* of a (different) second order polynomial in f_j. It is demonstrated in Section III of [35] that such a sequence $(\boldsymbol{f}^{(p)})_{p \in \mathbb{N}}$ converges to the unique maximizer of Φ.

Acknowledgment

V. Bettinardi, Scientific Institute Ospedale San Raffaele, for providing the patient data, and B. F. Holman, Institute of Nuclear Medicine, University College London, for processing HRCT data.

References

[1] Q. Xu, K. Yuan, and D. Ye, "Respiratory motion blur identification and reduction in ungated thoracic PET imaging," *Phys. Med. Biol.*, vol. 56, pp. 4481–4498, 2011.
[2] M. J. Nyflot *et al.*, "Impact of CT attenuation correction method on quantitative respiratory-correlated (4D) PET/CT imaging," *Med. Phys.*, vol. 42, no. 1, pp. 110–120, 2015.
[3] S. A. Nehmeh *et al.*, "Four-dimensional (4D) PET/CT imaging of the thorax," *Med. Phys.*, vol. 31, no. 12, pp. 3179–3186, 2004.
[4] C. Liu *et al.*, "Quiescent period respiratory gating for PET/CT," *Med. Phys.*, vol. 37, no. 9, pp. 5037–5043, 2010.
[5] T. Pan, T. Y. Lee, E. Rietzel, and G. T. Chen, "4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT," *Med. Phys.*, vol. 31, no. 2, pp. 333–340, 2004.
[6] B.-K. Teo *et al.*, "The effect of breathing irregularities on quantitative accuracy of respiratory gated PET/CT," *Med. Phys.*, vol. 39, no. 12, pp. 7390–7397, 2012.
[7] S. A. Nehmeh and Y. E. Erdi, "Respiratory motion in positron emission tomography/computed tomography: A review," *Semin. Nucl. Med.*, vol. 38, no. 3, pp. 167–176, 2008.

[8] D. Han, J. Bayouth, and S. Bhatia, "Characterization and identification of spatial artifacts during 4D-CT imaging," *Med. Phys.*, vol. 38, no. 4, pp. 2074–2087, 2011.
[9] M. Dawood *et al.*, "Motion correction in respiratory gated cardiac PET/CT using multi-scale optical flow," *Med. Image Comput. Comput. Assist. Interv.*, vol. 11, no. 2, pp. 155–162, 2008.
[10] H. J. Fayad, F. Lamare, C. C. Le Rest, V. Bettinardi, and D. Visvikis, "Generation of 4-dimensional CT images based on 4-dimensional PET-derived motion fields," *J. Nucl. Med.*, vol. 54, no. 4, pp. 631–638, 2013.
[11] I. Polycarpou, C. Tsoumpas, and P. K. Marsden, "Analysis and comparison of two methods for motion correction in PET imaging," *Med. Phys.*, vol. 39, no. 10, pp. 6474–6483, 2012.
[12] D. Visvikis, F. Lamare, P. Bruyant, N. Boussion, and C. C. Le Rest, "Respiratory motion in positron emission tomography for oncology applications: Problems and solutions," *Nucl. Instrum. Meth. A*, vol. 569, no. 2, pp. 453–457, 2006.
[13] R. Manjeshwar, T. Xiaodong, E. Asma, and K. Thielemans, "Motion compensated image reconstruction of respiratory gated PET/CT," *IEEE Int. Symp. Biomed. Imag.*, pp. 674–677, 2006.
[14] F. Qiao, T. Pan, J. W. Clark, and O. R. Mawlawi, "A motion-incorporated reconstruction method for gated PET studies," *Phys. Med. Biol.*, vol. 51, pp. 3769–3783, 2006.
[15] F. Lamare *et al.*, "List-mode-based reconstruction for respiratory motion correction in PET using non-rigid body transformations," *Phys. Med. Biol.*, vol. 52, pp. 5187–5204, 2007.
[16] I. Polycarpou, C. Tsoumpas, A. P. King, and P. K. Marsden, "Impact of respiratory motion correction and spatial resolution on lesion detection in PET: A simulation study based on real MR dynamic data," *Phys. Med. Biol.*, vol. 59, pp. 697–713, 2014.
[17] S. Y. Chun *et al.*, "MRI-based nonrigid motion correction in simultaneous PET/MRI," *J. Nucl. Med.*, vol. 53, no. 8, pp. 1284–1291, 2012.
[18] Y. Petibon *et al.*, "Towards coronary plaque imaging using simultaneous PET-MR: A simulation study," *Phys. Med. Biol.*, vol. 59, pp. 1203–1222, 2014.
[19] J. Nuyts *et al.*, "Simultaneous maximum *A posteriori* reconstruction of attenuation and activity distributions from emission sinograms," *IEEE Trans. Med. Imag.*, vol. 18, no. 5, pp. 393–403, May 1999.
[20] A. Rezaei *et al.*, "Simultaneous reconstruction of activity and attenuation in time-of-flight PET," *IEEE Trans. Med. Imag.*, vol. 31, no. 12, pp. 2224–2233, Dec. 2012.
[21] A. Rezaei, M. Defrise, and J. Nuyts, "ML-reconstruction for TOF-PET with simultaneous estimation of the attenuation factors," *IEEE Trans. Med. Imag.*, vol. 33, no. 7, pp. 1563–1572, Jul. 2014.
[22] M. Defrise, A. Rezaei, and J. Nuyts, "Time-of-flight PET data determine the attenuation sinogram up to a constant," *Phys. Med. Biol.*, vol. 57, pp. 885–899, 2012.
[23] M. W. Jacobson and J. A. Fessler, "Joint estimation of image and deformation parameters in motion-corrected PET," in *IEEE Nucl. Sci. Symp. Conf. Rec.*, 2003, pp. 3290–3294.
[24] M. Blume, A. Martinez-Müller, A. Keil, N. Navab, and M. Rafecas, "Joint reconstruction of image and motion in gated positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 29, no. 11, pp. 1892–1906, Nov. 2010.
[25] A. Bousse and O. Bertolli *et al.*, "Direct joint motion estimation/image reconstruction in attenuation-corrected gated PET/CT using a single CT," *Fully Three-Dimensional Image Reconstruct. Radiol. Nucl. Med.*, 2015.
[26] D. F. Yu and J. A. Fessler, "Edge-preserving tomographic reconstruction with nonlocal regularization," *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 159–173, Feb. 2002.
[27] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York: Springer, 2006.
[28] R. H. Byrd, P. Lu, and J. Nocedal, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Stat. Comp.*, vol. 16, no. 5, pp. 1190–1208, 1995.
[29] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Trans. Math. Software*, vol. 23, no. 4, pp. 550–560, 1997.
[30] J. J. Moré and D. J. Thuente, "Line search algorithms with guaranteed sufficient decrease," *ACM Trans. Math. Software*, vol. 20, no. 3, pp. 286–307, 1994.
[31] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, no. 2, pp. 113–122, 1982.
[32] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomogr.*, vol. 8, no. 2, pp. 306–316, 1984.
[33] Y. Vardi, L. A. Shepp, and L. Kauffman, "A statistical model for positron emission tomography," *J. Am. Stat. Assoc.*, vol. 80, no. 389, pp. 8–20, 1985.
[34] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, no. 1, pp. 84–93, Mar. 1990.
[35] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imag.*, vol. 14, no. 1, pp. 132–137, Mar. 1995.
[36] H. Erdogan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Trans. Med. Imag.*, vol. 18, no. 9, pp. 801–14, Sep. 1999.
[37] S. Ahn and J. A. Fessler, "Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms," *IEEE Trans. Med. Imag.*, vol. 22, no. 5, pp. 613–626, May 2003.
[38] M. Teräs, T. Tolvanen, J. J. Johansson, J. J. Williams, and J. Knuuti, "Performance of the new generation of whole-body PET/CT scanners: Discovery STE and discovery VCT," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 34, no. 10, pp. 1683–1692, 2007.
[39] M. Blume, "Joint image and motion reconstruction for positron emission tomography," Ph.D. dissertation, Fakultät für Informatik Technische Universität München, Munich, Germany, 2011.
[40] M. W. Jacobson, "Approaches to Motion-corrected PET image reconstruction from respiratory gated projection data," Ph.D. dissertation, Univ. Michigan, Ann Arbor, 2006.
[41] G. Wang and J. Qi, "An optimization transfer algorithm for nonlinear parametric image reconstruction from dynamic PET data," *IEEE Trans. Med. Imag.*, vol. 31, no. 10, pp. 1977–1988, Oct. 2012.