

ML-EM ALGORITHM WITH KNOWN CONTINUOUS MOVEMENT MODEL

CAMILLE POUCHOL AND OLIVIER VERDIER

ABSTRACT. In Positron Emission Tomography, movement leads to blurry reconstructions when not accounted for. Whether known a priori or estimated jointly to reconstruction, motion models are increasingly defined in continuum rather than in discrete, for example by means of diffeomorphisms. The present work provides both a statistical and functional analytic framework suitable for handling such models. It is based on time-space Poisson point processes as well as regarding images as measures, and allows to compute the maximum likelihood problem for line-of-response data with a known movement model. Solving the resulting optimisation problem, we derive an Maximum Likelihood Expectation Maximisation (ML-EM) type algorithm which recovers the classical ML-EM algorithm as a particular case for a static phantom. The algorithm is proved to be monotone and convergent in the low-noise regime. Simulations confirm that it correctly removes the blur that would have occurred if movement were neglected.

1. INTRODUCTION

In Positron Emission Tomography (PET), line-of-response data consists of the times of simultaneous detections of two photons, in each of the m pairs of detectors, from which one aims at reconstructing the underlying image $\mu \in X$, for some suitable space of images X . When the phantom is static, the times are grouped into numbers of detections y_i per detector $i \in \{1, \dots, m\}$. A good statistical model for the problem is then $y = \mathcal{P}(A\mu)$, i.e., the data is obtained as m independent Poisson random variables y_i of mean $(A\mu)_i$ where $A: X \rightarrow \mathbb{R}^m$ is a known operator modelling the scanner geometry.

This inverse problem is in practice solved through variants of the iterative Maximum Likelihood Expectation Maximisation (ML-EM) algorithm

$$\mu_{k+1} = \frac{\mu_k}{A^T \mathbf{1}} A^T \left(\frac{y}{A\mu_k} \right),$$

which maximises the likelihood associated to the above statistical model, i.e., it minimises $\ell(\mu) := d(y||A\mu)$ over $\mu \in X$, $\mu \geq 0$, where d is the Kullback–Leibler divergence [21, 22, 17].

Reconstruction methods in medical imaging suffer from blurring effects if the phantom moves during acquisition time, unless movement taken into account in the reconstruction process. Cardiac or thoracic PET scans are typical instances of this problem.

Motion estimation refers to methods which take movement into account by estimating it. This can be done either prior to reconstruction, or jointly with it, see [8, 6, 18, 19] for a review. A first class of methods is based on a discrete description of movement, such as [10, 11]. However, models with a continuous description of movement, typically by means of diffeomorphisms, are gaining popularity in the context of PET [3, 9], and more broadly in imaging sciences [23].

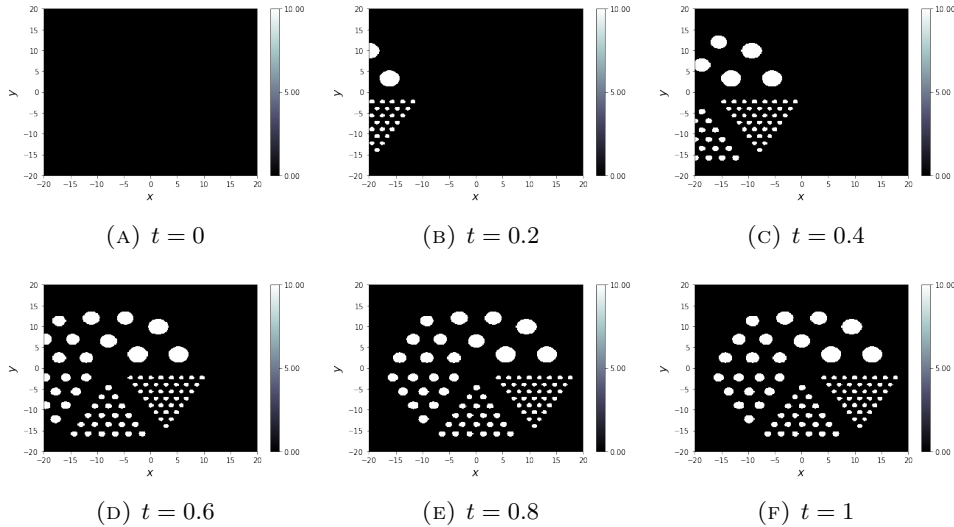


FIGURE 1. Template evolution through translation operators.

Considering X as a functional space of functions defined over a compact $K \subset \mathbb{R}^p$, these approaches assume that the activity is modified by operators $\mathcal{W}_t : X \rightarrow X$.

If we assume that \mathcal{W}_t is defined through a diffeomorphism for any $t \in [0, 1]$, for instance, as diffeomorphisms do not preserve grids, the latter approach does not easily lend itself to discretisation (in which case X becomes a finite-dimensional space).

The aim of the present paper is twofold.

- First, extending [17], we introduce a continuous mathematical framework which incorporates any movement model.
- Second, based on this construction, we propose and analyse an ML-EM type algorithm for the maximum likelihood problem associated to a known continuous movement model.

When the movement is in fact static, we recover the classical ML-EM algorithm as a particular case. Even in the static case, optimal solutions to the maximum likelihood problem can be singular measures (sums of point masses), and it is the standard outcome in the noisy case [14, 17]. We thus use measures to model images in this continuous context.

We also emphasise that in the case of *gated* data, i.e., when a physical device allows to group counts per phase in which the movement can be assumed to be stationary, an ML-EM algorithm has been derived informally in the literature [9, 13, 15]. We recover that algorithm as well when the \mathcal{W}_t are assumed to be piecewise constant in time.

The proposed approach applies to cases where the movement model is known. This can be for scans of phantoms where movement is controlled, or when the motion model is estimated prior to reconstruction. An example is provided in Figure 1 and Figure 2; see §4.2 for more details.

Our approach can also serve as a building block for joint motion estimation and reconstruction, when the movement model is unknown. A typical strategy to solve the resulting optimisation problem is to intertwine estimating the image and the transformation (typically a diffeomorphism) [9, 13, 15]. Thus, given a current

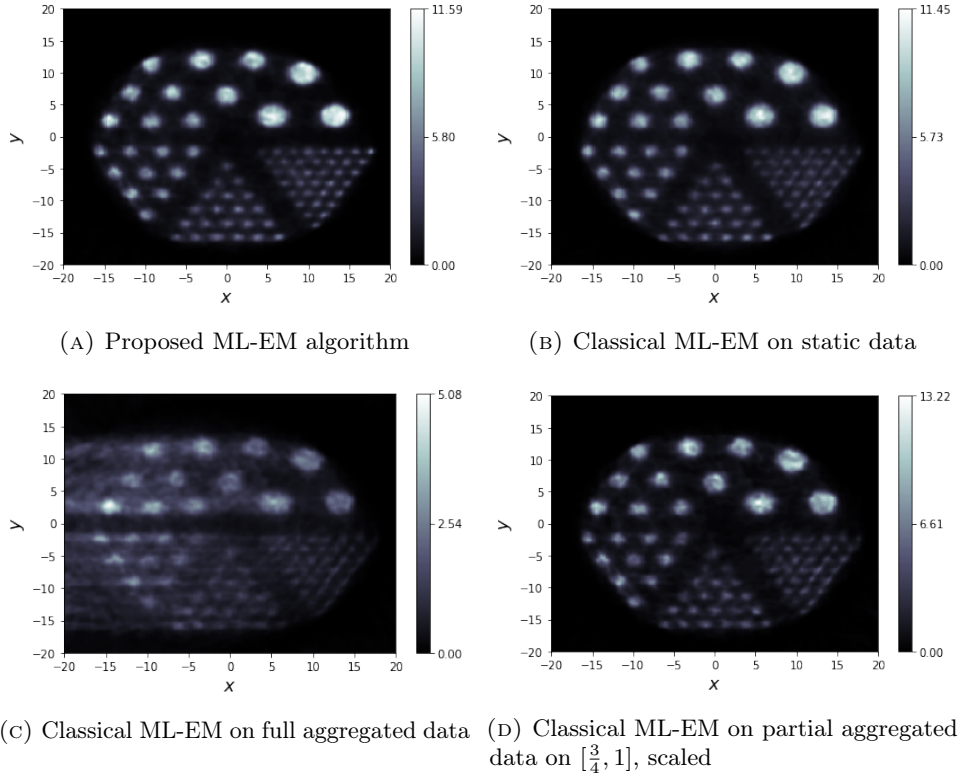


FIGURE 2. Tenth iterate of the proposed algorithm (3.1) compared to the classical ML-EM algorithm in various cases.

estimate of the transformations \mathcal{W}_t , the next iterate for the image is obtained by maximising the corresponding likelihood, which is exactly what our algorithm does.

Outline of the paper. The paper is organised as follows. In §2, we introduce the notations as well as the modelling through time-space Poisson point processes. We derive the likelihood associated to the times of detections for this statistical model. Finally, we provide the corresponding optimality conditions. The ML-EM algorithm is introduced in §3, where we prove its monotonicity and analyse its convergence properties. In §4, numerical simulations are presented, showing that our algorithm performs like the classical ML-EM, and avoids blurring effects when compared to neglecting motion.

2. MAXIMUM LIKELIHOOD FUNCTION

2.1. Notations. For a given compact $E \subset \mathbb{R}^d$, we denote $\mathcal{M}(E)$ the set of Radon measures defined on E , i.e., the topological dual of continuous functions $\mathcal{C}(E)$ over E . Endowing $\mathcal{M}(E)$ with its weak- $*$ topology, its dual is given by $\mathcal{C}(E)$. We denote by $\langle \mu, f \rangle$ the pairing of a measure $\mu \in \mathcal{M}_+(E)$ and a function $f \in \mathcal{C}(E)$.

The set of nonnegative measures will be denoted $\mathcal{M}_+(E)$. By the Riesz–Markov representation theorem, we may also regard a measure $\mu \in \mathcal{M}_+(E)$ as a Borel measure, and we will sometimes write $\mu(B)$ for the measure of a measurable set $B \subset E$.

For μ, ν two measures in $\mathcal{M}(E)$, the notation $\mu \ll \nu$ means that μ is absolutely continuous with respect to ν .

An operator $\mathcal{W}: \mathcal{M}(E) \rightarrow \mathcal{M}(E)$ i.e., a linear and continuous mapping in the weak-* topology, is defined through its adjoint $\mathcal{W}^*: \mathcal{C}(E) \rightarrow \mathcal{C}(E)$ by

$$\langle \mathcal{W}\mu, f \rangle := \langle \mu, \mathcal{W}^*f \rangle$$

for any $f \in \mathcal{C}(E)$.

Note that the adjoint of an operator of $\mathcal{M}(E)$ is well-defined as a mapping from $\mathcal{C}(E)$ onto itself, meaning that such an operator indeed may be defined through its adjoint, see [4, Proposition 3.14].

2.2. Modelling. We quickly recall the physics underlying Positron Emission Tomography. A radiotracer injected into the patient diffuses into tissues and then disintegrates by emitting positrons. A given positron, after a very short travel distance, interacts with an electron, yielding two photons emitted in (uniformly) random opposite directions. Such photons are then detected simultaneously by a pair of detectors.

We are given a compact $K \subset \mathbb{R}^p$ (having $p = 2, p = 3$ in mind for the application) on which the image is defined. We assume that the emission process is defined by a time-space Poisson point process. The intensity of the latter is a measure in $\mathcal{M}_+([0, 1] \times K)$, where, without loss of generality, we fix the final time to one.

We assume that the intensity at time t is given by a known linear transformation of the unknown measure $\mu \in \mathcal{M}_+(K)$. In other words, the activity at time t is the measure $\mathcal{W}_t\mu$, where the linear operators

$$\mathcal{W}_t: \mathcal{M}(K) \rightarrow \mathcal{M}(K), \quad t \in [0, 1],$$

are known.

Hence, we define the measure underlying the time-space Poisson process by

$$[t_1, t_2] \times B \mapsto \int_{t_1}^{t_2} (\mathcal{W}_t\mu)(B) dt, \quad 0 \leq t_1 < t_2 \leq 1, \quad B \subset K \text{ Borel set,}$$

which we denote $\mathcal{W}_t\mu$ with a slight abuse of notation. This measure is well-defined under the assumption (2.3), see the full set of assumptions in the next subsection.

Independently of the emission process associated, a point emitted at $x \in K$ and time $t \in [0, 1]$ then has a relative probability $a_i(x)$ to be detected in detector i , and we thus assume

$$(1) \quad a_i \in \mathcal{C}(K), \quad a_i \geq 0, \quad i = 1, \dots, m.$$

We assume that the detection also occurs at time t , which is an excellent approximation.

We now define the PET operator $A: \mathcal{M}(K) \rightarrow \mathbb{R}^m$ by

$$(A\mu)_i := \langle \mu, a_i \rangle = \int_K a_i d\mu, \quad i = 1, \dots, m.$$

Note that the adjoint $A^*: \mathbb{R}^m \rightarrow \mathcal{C}(K)$ of A is given by

$$A^*\lambda = \sum_{i=1}^m \lambda_i a_i, \quad \lambda \in \mathbb{R}^m.$$

2.3. Assumptions on the Transformations. We make the following assumptions on the transformations \mathcal{W}_t , for $t \in [0, 1]$.

Nonnegativity Preservation:

$$(2) \quad \mu \geq 0 \implies \mathcal{W}_t\mu \geq 0, \quad t \in [0, 1]$$

Integrability: denoting 1 for the constant function in $\mathcal{C}(K)$, we assume

$$(3) \quad \int_{[0,1] \times K} \mathcal{W}_t^*1 < +\infty.$$

Defining $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^m$, the condition (2.3) allows us to define the following integral:

$$(4) \quad f := \int_0^1 \mathcal{W}_t^* A^* \mathbf{1} dt$$

A consequence of (2.3) and (2.2) is that $f \geq 0$. We make the further blanket assumption that $f \neq 0$, that is (since K is compact)

$$(5) \quad \exists c > 0, \quad f(x) \geq c \quad x \in K.$$

The assumption (2.3) equivalently writes $g \geq 0 \implies \mathcal{W}_t^* g \geq 0$ for any $t \in [0, 1]$, $g \in \mathcal{C}(K)$, $g \geq 0$. It means that a transformation \mathcal{W}_t cannot create illegal (negative) activity. The assumption (2.3) essentially ensures that there is a finite activity over time. The blanket assumption (2.3) is natural: if $f = 0$, then $\mathcal{W}_t^* a_i = 0$ for all detectors i and times t , which means that nothing is ever measured. We simply exclude this trivial case.

2.4. Maximum likelihood. We assume that there are n_i detections in a given detector $i \in \{1, \dots, m\}$, detected at times t_j^i , $j = 1, \dots, n_i$. We do not require that the times be ordered. We also denote

$$n := \sum_{i=1}^m n_i,$$

the total number of detections.

Our aim is to derive the likelihood for the problem of estimating μ from the data given by the number of points and times of detections, namely

$$n_i, \quad t_j^i, \quad j = 1, \dots, n_i, \quad i = 1, \dots, m,$$

We introduce a couple of additional notations. First, we define for a time $t \in [0, 1]$ and the measure $\mu \in \mathcal{M}_+(K)$ the function

$$\beta_i(t) := (A\mathcal{W}_t\mu)_i, \quad i = 1, \dots, m,$$

or equivalently, $\beta_i(t) = \langle \mathcal{W}_t\mu, a_i \rangle$. We also define a collection Γ of n continuous functions by

$$\Gamma := \{ \mathcal{W}_{t_j^i}^* a_i \mid i = 1, \dots, m, j = 1, \dots, n_i \}.$$

We first identify the point process from which the data is drawn.

Proposition 2.1. *For $i \in \{1, \dots, m\}$, the number of points n_i and times t_j^i are drawn from independent inhomogeneous (in time) Poisson processes on $[0, 1]$, with respective intensities given by the functions β_i .*

Proof. Since the emission process is independent from that of the detection process, we use the thinning property ([12, Theorem 5.8]) to assert that for each $i \in \{1, \dots, m\}$ the point process defined by the points detected by detector i is also a time-space Poisson process, with underlying measure

$$a_i \mathcal{W}_t\mu \in \mathcal{M}_+([0, 1] \times K),$$

i.e., the measure

$$[t_1, t_2] \times B \longmapsto \int_{t_1}^{t_2} (a_i \mathcal{W}_t\mu)(B) dt, \quad 0 \leq t_1 < t_2 \leq 1, \quad B \subset K \text{ Borel set.}$$

Furthermore and still by the thinning property, all these processes are independent.

For a given $i \in \{1, \dots, m\}$, the number of points n_i and the times of detections t_j^i are that of a (time) Poisson process defined on $[0, 1]$, since it corresponds to marginalising with respect to $x \in K$ the associated time-space Poisson process. The

resulting object is thus an inhomogeneous (in time) Poisson process, with intensity obtained by integrating the measure $a_i \mathcal{W}_t \mu$ in space:

$$\forall t \in [0, 1], \quad \int_K d(a_i \mathcal{W}_t \mu) = \langle \mathcal{W}_t \mu, a_i \rangle = \beta_i(t).$$

□

We are now in a position to derive the maximum likelihood function, namely:

Corollary 2.2. *The maximum likelihood problem is given by*

$$\min_{\mu \in \mathcal{M}_+(K)} \ell(\mu),$$

where

$$\ell(\mu) := \langle \mu, f \rangle - \sum_{\gamma \in \Gamma} \log(\langle \mu, \gamma \rangle),$$

with the convention that $\ell(\mu) = +\infty$ for μ not in

$$\text{dom}(\ell) := \{ \mu \in \mathcal{M}_+(K) \mid \langle \mu, \gamma \rangle > 0, \quad \gamma \in \Gamma \}.$$

Proof. The negative log-likelihood for the number of points n and times t_j for an inhomogeneous Poisson point process of intensity $\beta \in L^1(0, 1)$ is known to be given up to constants by

$$\int_0^1 \beta(t) dt - \sum_{j=1}^n \log(\beta(t_j)).$$

The integrability condition (2.3) together with the preservation of nonnegativity (2.3) implies that $t \mapsto \mathcal{W}_t^* a_i \in L^1(0, 1; \mathcal{C}(K))$, and in particular the intensity functions β_i defined in the above proof all lie in $L^1(0, 1)$.

By independence, summing the above negative log-likelihood over $i \in \{1, \dots, m\}$ we find the full negative log-likelihood

$$\begin{aligned} \ell(\mu) &= \sum_{i=1}^m \int_0^1 \beta_i(t) dt - \sum_{i=1}^m \sum_{j=1}^{n_i} \log(\beta_i(t_j^i)), \\ &= \langle \mu, f \rangle - \sum_{\gamma \in \Gamma} \log(\langle \mu, \gamma \rangle). \end{aligned}$$

The last line just uses the fact that $\beta_i(t_j^i) = \langle \mu, W_{t_j^i}^* a_i \rangle$ whereas the exchange of order of integration in the first term can be performed by Fubini's theorem, owing to (2.3). □

Remark 2.3. In the stationary case, i.e., $\mathcal{W}_t = \text{Id}$ over $[0, 1]$, we have $f = A^* \mathbf{1}$ and for $\gamma \in \Gamma$ defined by $i \in \{1, \dots, m\}$, $j \in i \in \{1, \dots, n_i\}$, $\gamma = \mathcal{W}_{t_j^i}^* a_i = a_i$. Hence

$$\sum_{\gamma \in \Gamma} \log(\langle \mu, \gamma \rangle) = \sum_{i=1}^m \sum_{j=1}^{n_i} \log(\langle \mu, a_i \rangle) = \sum_{i=1}^m n_i \log((A\mu)_i),$$

so that the loss function writes

$$\ell(\mu) = \langle \mu, A^* \mathbf{1} \rangle - \sum_{i=1}^m n_i \log((A\mu)_i).$$

Up to constants, this is nothing but the usual negative log-likelihood $d(y||A\mu)$ with $y = (n_1, \dots, n_m)$ used in the static case [17], and d the Kullback-Leibler divergence defined for nonnegative vectors $u, v \in \mathbb{R}^m$ by

$$d(u||v) = \sum_{i=1}^m u_i \log\left(\frac{u_i}{v_i}\right) - u_i + v_i,$$

with value $+\infty$ if there exists $i \in \{1, \dots, m\}$ such that $u_i > 0, v_i = 0$.

2.5. Optimality conditions. We derive below the optimality conditions for the maximum likelihood problem.

To derive optimality conditions, we endow $\mathcal{M}(K)$ with its strong topology. The function ℓ takes finite values and is differentiable on its domain $\text{dom}(\ell)$, which is open. For $\mu \in \text{dom}(\ell)$, we readily compute

$$(6) \quad \nabla \ell(\mu) = f - \sum_{\gamma \in \Gamma} \frac{\gamma}{\langle \mu, \gamma \rangle},$$

an element of the dual of $\mathcal{M}(K)$ which we may identify with an element of $\mathcal{C}(K)$.

Proposition 2.4. *A measure $\mu^* \in \text{dom}(\ell)$ is optimal if and only if*

$$(7) \quad \begin{aligned} \nabla \ell(\mu^*) &\geq 0 \text{ on } K, \\ \nabla \ell(\mu^*) &= 0 \text{ on } \text{supp}(\mu^*). \end{aligned}$$

Proof. Since the function ℓ is convex, a measure $\mu^* \in \text{dom}(\ell)$ is optimal if and only if

$$\nabla \ell(\mu^*) \in -N_{\mathcal{M}_+(K)}(\mu^*).$$

where

$$N_{\mathcal{M}_+(K)}(\mu) := \{g \in \mathcal{C}(K) \mid \forall \nu \in \mathcal{M}_+(K), \langle \nu - \mu, g \rangle \leq 0\}$$

is the normal cone to $\mathcal{M}_+(K)$ at μ . From [17, Lemma 3.5], the normal cone can be characterised by

$$N_{\mathcal{M}_+(K)}(\mu) = \{g \in \mathcal{C}(K) \mid g \leq 0 \text{ on } K, g = 0 \text{ on } \text{supp}(\mu)\},$$

and the claim follows. \square

Note that this optimality criterion shows that if there exists a measure μ such that

$$\sum_{\gamma \in \Gamma} \frac{\gamma}{\langle \mu, \gamma \rangle} = f,$$

then μ is optimal.

Corollary 2.5. *The infimum of ℓ is a minimum.*

Proof. If an optimum exists, since $\nabla \ell(\mu^*) = 0$ vanishes on the support of μ^* , we obtain $\langle \mu^*, \nabla \ell(\mu^*) \rangle = 0$. A computation using (2.5) shows on the other hand that

$$\langle \mu^*, \nabla \ell(\mu^*) \rangle = \langle \mu^*, f \rangle - \#\Gamma = \langle \mu^*, f \rangle - n.$$

As a result, it suffices to minimise ℓ on the set $\{\mu \in \mathcal{M}_+(K) \mid \langle \mu, f \rangle = n\}$. Any measure μ in the previous set satisfies $\mu(K) \leq 1/c$ thanks to the lower bound on the function f . Thus, the set is bounded and hence weak-* compact by the Banach-Alaoglu theorem [20]. Since ℓ is clearly weak-* continuous, the claim follows. \square

3. ML-EM ALGORITHM

In this section, we define the ML-EM algorithm, and prove that it is monotone and convergent in the low noise regime. Since all measures and continuous functions will from now on all be defined on the compact K in this section, we drop the reference to K in the functional spaces, denoting them \mathcal{M} , \mathcal{M}_+ and \mathcal{C} respectively.

3.1. Definition and well-posedness. For $\mu_0 \in \text{dom}(\ell)$, we define the iterates

$$(8) \quad \mu_{k+1} = \frac{\mu_k}{f} \sum_{\gamma \in \Gamma} \frac{\gamma}{\langle \mu_k, \gamma \rangle}.$$

We first observe that the algorithm is well-defined. This is because

$$\mu_k \in \text{dom}(\ell) \implies \mu_{k+1} \in \text{dom}(\ell).$$

Indeed, using the lower bound (2.3) on f , we find for any $\gamma \in \Gamma$:

$$\langle \mu_{k+1}, \gamma \rangle \geq \frac{1}{c \langle \mu_k, \gamma \rangle} \langle \mu_k, \gamma^2 \rangle > 0,$$

since the Cauchy–Schwarz inequality $\langle \mu_k, \gamma^2 \rangle \mu_k(K) \geq \langle \mu_k, \gamma \rangle^2$ prevents $\langle \mu_k, \gamma^2 \rangle$ from vanishing. More precisely, after defining the compact

$$\tilde{K} := \bigcup_{\gamma \in \Gamma} \text{supp}(\gamma),$$

we find that $\text{supp}(\mu_k) = \tilde{K}$ for all $k \geq 1$, provided that $\text{supp}(\mu_0) = K$ and $\mu_0 \in \text{dom}(\ell)$. Note that the optimality conditions (2.4) prove that any optimal measure μ^* satisfies $\text{supp}(\mu^*) \subset \tilde{K}$.

Finally, we remark that

$$(9) \quad \forall k \geq 1, \quad \langle \mu_k, f \rangle = n.$$

Remark 3.1. Assume that for some times $t_0 < t_1, \dots < t_N$ we have

$$\forall t \in (t_{s-1}, t_s), \quad \mathcal{W}_t = \mathcal{W}_{t_s}, \quad s = 1, \dots, N,$$

i.e., the movement is piecewise constant on (t_{s-1}, t_s) for $s \in \{1, \dots, N\}$. We may then also group points by phase and detector by denoting n_i^s the number of points detected in detector i between t_{s-1} and t_s , and $n^s = (n_1^s, \dots, n_m^s)$. Using the notations $A_s := A\mathcal{W}_{t_s}$, $\Delta t_s = t_s - t_{s-1}$, the algorithm then rewrites

$$(10) \quad \mu_{k+1} = \frac{\mu_k}{\sum_{s=1}^N (\Delta t_s) A_s^* \mathbf{1}} \sum_{s=1}^N A_s^* \left(\frac{n^s}{A_s \mu_k} \right).$$

In other words, we recover the algorithm for *gated* data, proposed in [9] for the intensity-preserving action, and generalised in [13, 15]. The ensuing analysis is up to our knowledge the first rigorous justification for these informally-derived algorithms, under the assumption of piecewise-constant movement.

Note also that (3.1) can be rewritten as

$$\mu_{k+1} = \frac{\mu_k}{\sum_{s=1}^N \tilde{A}_s^* \mathbf{1}} \sum_{s=1}^N \tilde{A}_s^* \left(\frac{n^s}{\tilde{A}_s \mu_k} \right).$$

where $\tilde{A}_s := (\Delta t_s) A_s$. In this case, this is simply the standard ML-EM algorithm with the new operator $\tilde{A} : \mathcal{M} \rightarrow \mathbb{R}^{mN}$ defined by $\tilde{A} := [\tilde{A}_1, \dots, \tilde{A}_N]$.

In particular, we also recover the classical ML-EM algorithm, since if $\mathcal{W}_t = \text{Id}$ for all $t \in [0, 1]$, the above simplifies to

$$\mu_{k+1} = \frac{\mu_k}{A^* \mathbf{1}} A^* \left(\frac{y}{A \mu_k} \right),$$

with $y = (n_1, \dots, n_m)$.

3.2. Monotonicity. For a nonnegative function $\gamma \in \mathcal{C}$ and a nonnegative measure $\mu \in \mathcal{M}_+$, we define the measure

$$\nu_\gamma(\mu) := \frac{\gamma\mu}{\langle \mu, \gamma \rangle}.$$

Note that ν_γ is a probability measure over K , and that, following (3.1),

$$\mu_{k+1} = \frac{1}{f} \sum_{\gamma \in \Gamma} \nu_\gamma(\mu_k).$$

Define the set $X_k := \{ \mu \in \mathcal{M}_+ \mid \mu_{k+1} \ll \mu \ll \mu_k, \langle \mu, f \rangle = 1 \}$ for $k \in \mathbb{N}$. We now define the surrogate function $Q_k: X_k \rightarrow \mathbb{R}$ by

$$Q_k(\mu) := \ell(\mu) + \sum_{\gamma \in \Gamma} D(\nu_\gamma(\mu_k) \parallel \nu_\gamma(\mu)),$$

where D is the Kullback-Leibler divergence defined for any two probability measures $\mu \ll \nu$ over K by

$$D(\mu \parallel \nu) := \int_K \log \left(\frac{d\mu}{d\nu} \right) d\mu,$$

with $\frac{d\mu}{d\nu}$ standing for the Radon–Nikodym derivative of μ with respect to ν .

Lemma 3.2. *The following holds for all $k \geq 1$:*

- (i) $Q_k(\mu) \geq \ell(\mu), \quad \mu \in X_k$
- (ii) $Q_k(\mu_k) = \ell(\mu_k)$
- (iii) $Q_k(\mu) - Q_k(\mu_{k+1}) = \langle \mu - \mu_{k+1}, f \rangle + D(f\mu_{k+1} \parallel f\mu), \quad \mu \in X_k$
- (iv) $Q_k(\mu_k) - Q_k(\mu_{k+1}) = D(f\mu_{k+1} \parallel f\mu_k)$

Proof. The properties of divergences allows us to conclude about (i) and (ii).

One computes

$$Q_k(\mu) = \langle \mu, f \rangle - \sum_{\gamma \in \Gamma} \left\langle \nu_\gamma(\mu_k), \log \left(\langle \mu_k, \gamma \rangle \frac{d\mu}{d\mu_k} \right) \right\rangle.$$

This gives

$$\begin{aligned} Q_k(\mu) - Q_k(\mu_{k+1}) &= \\ &= \langle \mu - \mu_{k+1}, f \rangle + \sum_{\gamma \in \Gamma} \left\langle \nu_\gamma(\mu_k), \log \left(\langle \mu_k, \gamma \rangle \frac{d\mu_{k+1}}{d\mu_k} \right) - \log \left(\langle \mu_k, \gamma \rangle \frac{d\mu}{d\mu_k} \right) \right\rangle \\ &= \langle \mu - \mu_{k+1}, f \rangle + \underbrace{\left\langle \sum_{\gamma \in \Gamma} \nu_\gamma(\mu_k), \log \left(\frac{d\mu_{k+1}}{d\mu} \right) \right\rangle}_{f\mu_{k+1}} \\ &= \langle \mu - \mu_{k+1}, f \rangle + D(f\mu_{k+1} \parallel f\mu). \end{aligned}$$

which proves (iii). Finally, (iv) is a consequence of (iii) and (3.1). \square

These computations yield the monotony of the function ℓ along iterates.

Corollary 3.3. *For any $\mu_0 \in \text{dom}(\ell)$, we have*

$$0 \leq Q_k(\mu_k) - Q_k(\mu_{k+1}) \leq \ell(\mu_k) - \ell(\mu_{k+1}) \quad k \in \mathbb{N}.$$

Proof. It is a consequence of Lemma 3.2. First, using Lemma 3.2-(iv), we get $Q_k(\mu_k) - Q_k(\mu_{k+1}) \geq 0$. Now, conclude noticing that Lemma 3.2-(i) and Lemma 3.2-(ii) imply $Q_k(\mu_k) - Q_k(\mu_{k+1}) = \ell(\mu_k) - Q_k(\mu_{k+1}) \leq \ell(\mu_k) - \ell(\mu_{k+1})$. \square

3.3. Convergence. In this section, we highlight the main ideas of proofs, which largely follow [17].

Proposition 3.4. *For any $\mu_0 \in \text{dom}(\ell)$, the weak-* cluster points $\bar{\mu}$ of $\{\mu_k\}_{k \in \mathbb{N}}$ exist and are fixed points of the algorithm, namely*

$$\bar{\mu} = \frac{\bar{\mu}}{f} \sum_{\gamma \in \Gamma} \frac{\gamma}{\langle \bar{\mu}, \gamma \rangle}.$$

Proof. We first prove that the sequence is weak-* compact in \mathcal{M}_+ . Integrating the defining relation of ML-EM (3.1) and using the assumption (2.3) that $f \geq c > 0$, we indeed find

$$\mu_{k+1}(K) = \int_K d\mu_{k+1} \leq \frac{1}{c} \sum_{\gamma \in \Gamma} \frac{\langle \mu_k, \gamma \rangle}{\langle \mu_k, \gamma \rangle} = \frac{n}{c}.$$

From the Banach–Alaoglu theorem [20], we extract a weak-* converging subsequence to a given $\bar{\mu}$, and we denote the subsequence $\{\mu_{\varphi(k)}\}_{k \in \mathbb{N}}$.

We also note that ℓ is weak-* continuous. In particular, we have $\bar{\mu} \in \text{dom}(\ell)$ since otherwise $\{\ell(\mu_{\varphi(k)})\}_{k \in \mathbb{N}}$ would diverge to $+\infty$, in contradiction with its monotonicity.

Using Corollary 3.3 and Lemma 3.2 (iv), we can then follow the lines of [17, Proposition 4.3] to conclude the proof. \square

We refer to absolutely continuous measures for those that are absolutely continuous with respect to the Lebesgue measure on K , and proceed with a further assumption on the regularity of minimisers:

(11) there exists $\mu^* \in \arg \min_{\mu \in \mathcal{M}_+} \ell(\mu)$ absolutely continuous with $\text{supp}(\mu^*) = \tilde{K}$.

Remark 3.5. Using the results in [7], one typically expects the existence of such absolutely continuous measures in the low noise regime [17]. More precisely, in the static case and if $y = (n_1, \dots, n_m)$ is in the interior of the cone $\{A\mu \mid \mu \in \mathcal{M}_+\}$ in \mathbb{R}^m , such measures do exist [17].

Theorem 3.6. *Assume that assumption (3.3) holds. Then, for any $\mu_0 \in \text{dom}(\ell)$ absolutely continuous with a continuous and positive density over K , the algorithm is convergent in the sense that*

$$\ell(\mu_k) \xrightarrow{k \rightarrow +\infty} \min_{\mu \in \mathcal{M}_+} \ell(\mu).$$

Furthermore, any weak-* limit point $\bar{\mu}$ of the algorithm satisfies $\text{supp}(\bar{\mu}) = \tilde{K}$.

Proof. We argue in several steps, letting $\bar{\mu}$ be a cluster point of the ML-EM iterates $\{\mu_k\}_{k \in \mathbb{N}}$.

- (1) The assumptions on μ_0 ensure that μ_1 is absolutely continuous, with $\text{supp}(\mu_1) = \tilde{K}$, whence $D(f\mu^* \parallel f\mu_1) < +\infty$, where μ^* is defined by (3.3). We can then prove similarly to [17, Proposition 4.7] that for all $k \geq 1$,

$$D(f\mu^* \parallel f\mu_{k+1}) \leq D(f\mu^* \parallel f\mu_k).$$

Taking a subsequence along which $\{\mu_k\}_{k \in \mathbb{N}}$ converges weakly-* to $\bar{\mu}$ and using the weak-* lower semicontinuity of the Kullback–Leibler divergence [16], we find

$$D(f\mu^* \parallel f\bar{\mu}) < \infty.$$

This shows that $f\mu^* \ll f\bar{\mu}$, whence $\text{supp}(\bar{\mu}) = \tilde{K}$ thanks to the positivity of f .

- (2) We now make use of the fact that $\bar{\mu}$ must also be a fixed point of the algorithm, by virtue of Proposition 3.4. In other words, we have

$$\bar{\mu} = \frac{\bar{\mu}}{f} \sum_{\gamma \in \Gamma} \frac{\gamma}{\langle \bar{\mu}, \gamma \rangle} \iff \bar{\mu} \nabla \ell(\bar{\mu}) = 0.$$

This implies $\nabla \ell(\bar{\mu}) = 0$ on $\text{supp}(\bar{\mu})$. Outside of $\text{supp}(\bar{\mu}) = \tilde{K}$,

$$\nabla \ell(\bar{\mu}) = f - \sum_{\gamma \in \Gamma} \frac{\gamma}{\langle \bar{\mu}, \gamma \rangle} \geq 0,$$

since the right-hand side vanishes by definition of \tilde{K} , whereas f is positive. Hence, $\bar{\mu}$ satisfies both optimality conditions (2.4), which shows that $\bar{\mu}$ is optimal and hence the claim that any cluster point is optimal.

- (3) For the convergence of $\{\ell(\mu_k)\}_{k \in \mathbb{N}}$ towards the minimum, we just recall that the sequence is non-increasing from Corollary 3.3, hence its limit must coincide with $\ell(\bar{\mu})$ for any cluster point $\bar{\mu}$. The optimality of such cluster points concludes the proof. \square

The interest of emphasising the property $\text{supp}(\bar{\mu}) = \tilde{K}$ for cluster points is that this prevents them from being sparse measures.

4. NUMERICAL SIMULATIONS

All simulations are run in Python and use the Operator Discretization Library (`odl`) for manipulating operators [2], `neuron` for warping utilities [5], which itself uses `tensorflow` [1].

4.1. General approach. We work with the Derenzo phantom, displayed in the Introduction, see Figure 1.

The noise level is controlled by the dose (or time) by which we multiply the phantom. In all experiments we have run, we have multiplied the Derenzo phantom by ten. We denote it μ_r .

Given a time-evolution of the template through operators \mathcal{W}_t , we first generate the number of points and times per detector, using the result established in Proposition 2.1. The latter states that they are for each $i \in \{1, \dots, m\}$ independently drawn according to an inhomogeneous Poisson point process over $[0, 1]$, with intensity defined by

$$t \in [0, 1] \mapsto \beta_i(t) = \int_K a_i d(\mathcal{W}_t \mu_r) = \int_K (\mathcal{W}_t^* a_i) d\mu_r.$$

These processes are simulated using the rejection method, i.e., by finding a bound M_i such that $\beta_i \leq M_i$ on $[0, 1]$, simulating a homogeneous Poisson process of intensity M_i on $[0, 1]$, and then accepting a drawn time t_j^i with probability $\beta_i(t_j^i)/M_i$.

In each of the test cases presented below, we also compute the total number of points detected by detector n_i , $i \in \{1, \dots, m\}$, thus obtaining a sinogram, which we call *aggregated data*. Data can be aggregated on the whole interval of time $[0, 1]$ or on a small portion of it to curb the effect of movement. This is what is commonly done in modern scanners to alleviate blur coming from movement of organs, such as the lungs.

Finally, we will also compute a fictitious sinogram data, as obtained from the static phantom acquired for the same amount of time.

All in all, this gives us three benchmarks again which we may test our results for a given fixed number of iterates k_0 :

- k_0 iterates of the classical ML-EM algorithm on data acquired from the static phantom.

- k_0 iterates of the classical ML-EM algorithm on the aggregated data on the whole interval $[0, 1]$.
- k_0 iterates of the classical ML-EM algorithm on the aggregated data on a relevant subinterval of $[0, 1]$.

We expect our algorithm to perform as well as the classical ML-EM algorithm on a static phantom, thus avoiding both the blur observed when aggregating data on the whole time-interval because of movement, and the higher noise observed when only a portion of the aggregated data is kept.

4.2. Translation. We work with a 2D PET operator A with 90 angles (views) and 64 tangential positions, and the image space is a square with resolution 256×256 .

The image μ_t is given as the evolution of μ_r through operators \mathcal{W}_t defined by means of translations. We choose a mapping $c : [0, 1] \mapsto \mathbb{R}$ for a speed of translation, the operators are correspondingly defined through their adjoint for functions f by

$$x \mapsto \mathcal{W}_t^* f(x) = f(x + c(t)).$$

In the experiment the image is translated from left to right at speed t , up until it is at the center of the image, at $t = \frac{3}{4}$. The translation then stops. In other words,

$$c(t) = (c_1(t), 0), \quad c_1(t) = (at + b) \mathbb{1}_{[0, \frac{3}{4}]}(t), \quad t \in [0, 1]$$

for appropriately chosen constants a and b . The resulting evolving image is depicted in Figure 1.

Figure 2 shows the result of 10 iterates of our algorithm on the times t_i^j , compared to 10 iterates of the classical ML-EM algorithm as obtained either on a fictitious static case, or on partially or fully aggregated data. The partially aggregated data is taken from the time-interval $[\frac{3}{4}, 1]$, namely when the translation has stopped.

As expected, the result is almost undistinguishable from the static case, whereas the classical ML-EM algorithm on the full aggregated data leads to a poor image due to the movement. The same applied to partially aggregated data performs well since the image is static on the last portion of the time-window, but 3/4th of the data is unused, resulting in a noisier image. The small circles towards the center are indeed less easily distinguished in the case (d) as they can be in cases (a) and (b).

We observe a lower maximum for image (c) than for images (a) and (b). This is to be expected, as classical ML-EM iterates are such that $\langle \mu_k, A^* \mathbf{1} \rangle = n$, whereas it is $\langle \mu_k, f \rangle = n$ (see (3.1)) for our algorithm. Since the function f is less than $A^* \mathbf{1}$ in this translating situation, this explains why maxima must differ.

Finally, we stress that we have scaled the last image (d) by a multiplication of 4 to make the comparison with (a) and (b) easier, since there is only 1/4th of the data.

4.3. Mass-preserving action of diffeomorphisms. The forward operator is now a 2D PET operator A with 45 angles (views) and 64 tangential positions, and the image space is a square with resolution 128×128 .

The image μ_t is given as the evolution of μ_r through operators \mathcal{W}_t defined by means of diffeomorphisms φ_t and the mass-preserving action, i.e., for all $f \in \mathcal{C}$,

$$\mathcal{W}_t^* f(x) = f(\varphi_t(x)), \quad x \in K.$$

Now, if μ is absolutely continuous with density g with respect to the Lebesgue measure, we find

$$\langle \mathcal{W}\mu, f \rangle = \int_E f(\varphi(x))g(x) dx = \int_E f(x)|D\varphi^{-1}(x)|g(\varphi^{-1}(x)) dx,$$

which shows that $\mathcal{W}\mu$ is absolutely continuous with density $x \mapsto |D\varphi^{-1}(x)|g(\varphi^{-1}(x))$ with respect to the Lebesgue measure.

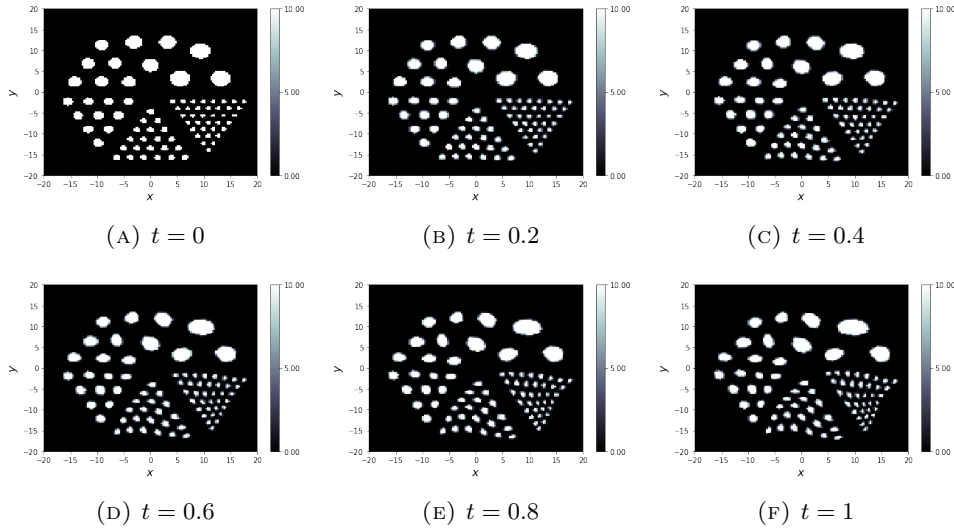


FIGURE 3. Template evolution through mass-preserving operators.

The diffeomorphisms are obtained by integration of a stationary vector field, as $\varphi_t = \exp(tv)$. In other words, they are obtained by integrating the following Cauchy problem over $[0, 1]$:

$$\begin{cases} \partial_t \varphi_t(x) = v(\varphi_t(x)), \\ \varphi_0 = \text{Id}. \end{cases}$$

The resulting evolving image is depicted in Figure 3.

Figure 4 presents the results after ten iterations of the proposed algorithm. Again, the results of the classical ML-EM algorithm on the static phantom after the same number of iterations look very similar.

However, the reconstruction obtained on the full aggregated data is unsurprisingly blurred, as all the positions are “averaged” in the reconstruction. The last reconstruction obtained from aggregating data on $[0, \frac{1}{4}]$ exhibit less blur but more noise.

REFERENCES

- [1] ABADI, M., AND ET AL. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [2] ADLER, J., KOHR, H., AND ÖKTEM, O. ODL-a Python framework for rapid prototyping in inverse problems. *Royal Institute of Technology* (2017).
- [3] BLUME, M., MARTINEZ-MOLLER, A., KEIL, A., NAVAB, N., AND RAFECAS, M. Joint reconstruction of image and motion in gated positron emission tomography. *IEEE Transactions on Medical Imaging* 29, 11 (2010), 1892–1906.
- [4] BREZIS, H. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.
- [5] DALCA, A. V., GUTTAG, J., AND SABUNCU, M. R. Anatomical priors in convolutional networks for unsupervised biomedical segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 9290–9299.
- [6] DAWOOD, M., JIANG, X., AND SCHÄFERS, K. P., Eds. *Correction Techniques in Emission Tomography*. Series in Medical Physics and Biomedical Engineering. CRC Press, 2008.
- [7] GEORGIU, T. T. Solution of the general moment problem via a one-parameter imbedding. *IEEE transactions on automatic control* 50, 6 (2005), 811–826.
- [8] GIGENGACK, F., JIANG, X., DAWOOD, M., AND SCHÄFERS, K. P. *Motion Correction in Thoracic Positron Emission Tomography*. Springer-Verlag, 2015.

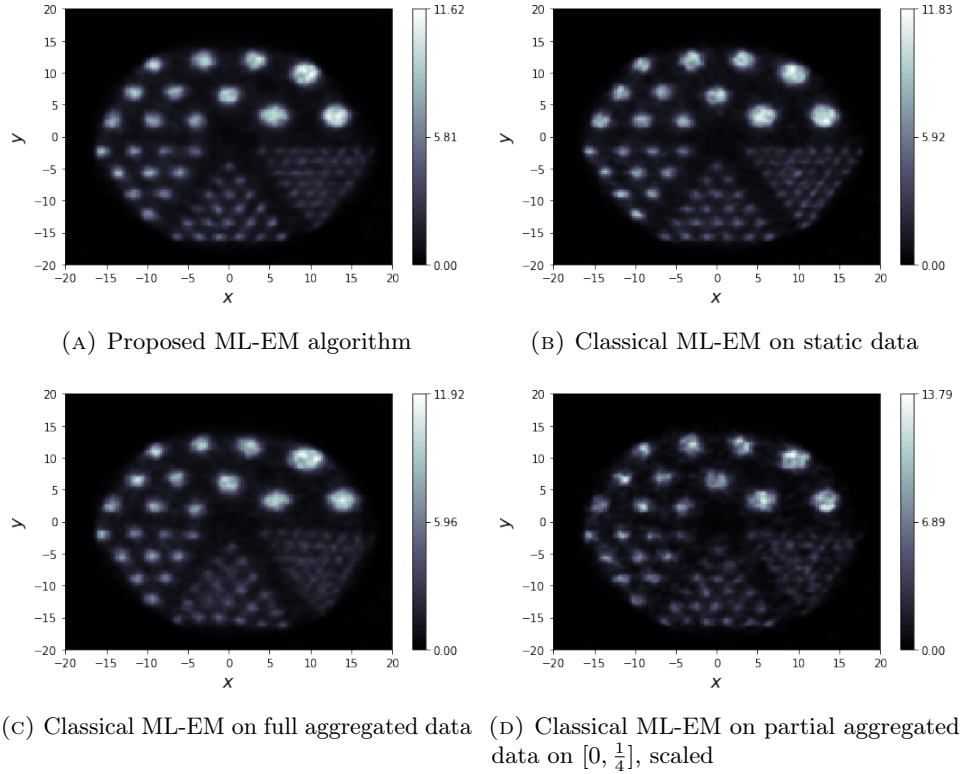


FIGURE 4. 10th iterate of our algorithm compared to the classical ML-EM algorithm in various cases.

- [9] HINKLE, J., SZEGEDI, M., WANG, B., SALTER, B., AND JOSHI, S. 4D CT image reconstruction with diffeomorphic motion model. *Medical image analysis* 16, 6 (2012), 1307–1316.
- [10] JACOBSON, M. W., AND FESSLER, J. A. Joint estimation of image and deformation parameters in motion-corrected PET. In *2003 IEEE Nuclear Science Symposium Conference Record* (2003), pp. 3290–3294.
- [11] JACOBSON, M. W., AND FESSLER, J. A. Joint estimation of respiratory motion and activity in 4D PET using CT side information. In *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro. Arlington, VA, April 6–9, 2006* (2006), pp. 275–278.
- [12] LAST, G., AND PENROSE, M. *Lectures on the Poisson process*, vol. 7. Cambridge University Press, 2017.
- [13] LI, T., ZHANG, M., QI, W., ASMA, E., AND QI, J. Motion correction of respiratory-gated pet image using deep learning based image registration framework. In *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine* (2019), vol. 11072, International Society for Optics and Photonics, p. 110720Q.
- [14] MAIR, B., RAO, M., AND ANDERSON, J. Positron emission tomography, Borel measures and weak convergence. *Inverse Problems* 12, 6 (1996), 965.
- [15] ÖKTEM, O., POUCHOL, C., AND VERDIER, O. Spatiotemporal pet reconstruction using ml-em with learned diffeomorphic deformation. In *International Workshop on Machine Learning for Medical Image Reconstruction* (2019), Springer, pp. 151–162.
- [16] POSNER, E. Random coding strategies for minimum entropy. *IEEE Transactions on Information Theory* 21, 4 (1975), 388–391.
- [17] POUCHOL, C., AND VERDIER, O. The ML-EM algorithm in continuum: sparse measure solutions. *Inverse Problems* 36, 3 (2020).
- [18] RAHMIM, A., TANG, J., AND ZAIDI, H. Four-dimensional image reconstruction strategies in cardiac-gated and respiratory-gated PET imaging. *PET Clinics* 8, 1 (2013), 51–67.
- [19] READER, A. J., AND VERHAEGHE, J. 4D image reconstruction for emission tomography. *Physics in Medicine and Biology* 59, 22 (2014), R371–R418.

- [20] RUDIN, W. *Functional analysis*, second ed. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, 1991.
- [21] SHEPP, L. A., AND VARDI, Y. Maximum likelihood reconstruction for emission tomography. *IEEE transactions on medical imaging* 1, 2 (1982), 113–122.
- [22] VARDI, Y., SHEPP, L., AND KAUFMAN, L. A statistical model for positron emission tomography. *Journal of the American statistical Association* 80, 389 (1985), 8–20.
- [23] YOUNES, L. *Shapes and Diffeomorphisms*, vol. 171 of *Applied Mathematical Sciences*. Springer-Verlag, 2010.

MAP5 LABORATORY, FP2M, CNRS FR 2036, UNIVERSITÉ DE PARIS, 75006 PARIS, FRANCE.

E-mail address: `camille.pouchol@parisdescartes.fr`

DEPARTMENT OF MATHEMATICS, KTH ROYAL INSTITUTE OF TECHNOLOGY, 100 44 STOCKHOLM, SWEDEN.

E-mail address: `olivier@kth.se`

DEPARTMENT OF COMPUTING, ELECTRICAL ENGINEERING AND MATHEMATICAL SCIENCES, WESTERN NORWAY UNIVERSITY OF APPLIED SCIENCES, BERGEN, NORWAY.

E-mail address: `olivier.verdier@hvl.no`