An EM Algorithm for Dynamic SPECT

Heinz H. Bauschke, Dominikus Noll,* Anna Celler, *Member, IEEE*, and Jonathan M. Borwein

Abstract— In this paper we present two variants of the EM algorithm for dynamic SPECT imaging. A version based on compartmental modeling which fits a sum of exponentials and a more general approach allowing for arbitrary decaying activities. The underlying probabilistic models are discussed and the incomplete and complete data spaces are shown to be physically meaningful.

We indicate that the second method, leading to a convex program in the M step, is easier to treat numerically and we present a possible numerical approach. Some preliminary numerical tests indicating the feasibility of the method are included.

I. INTRODUCTION

S INGLE photon emission computed tomography (SPECT) is a nuclear medicine diagnosis technique which measures the three dimensional (3-D) distribution of a radioactively labeled pharmaceutical injected in the body. As compared to standard imaging techniques such as computed tomography (CT), the significance of SPECT lies in the fact that it reveals the function of the body rather than its structure. For example, if a radio pharmaceutical is absorbed by an unhealthy tissue and rejected by healthy tissue, then SPECT will reveal the unhealthy tissue as a bright region. A related technique is positron emission tomography (PET), [34].

A SPECT camera works by rotating around the patient a scintillation detector (camera head) that records gamma rays emitted by the patient. A collimator placed in front of the detector rejects rays that are not perpendicular to the camera face (see Fig. 1). The images resulting in the camera are two-dimensional 2-D projections of the original 3-D activity distribution in the patient. Some devices use double or triple head cameras, or even ring SPECT instruments, to improve the number of detected counts and therefore the statistics.

Current clinical applications of SPECT are based on the hypothesis that the injected radio activity in the organ of interest is stationary over the acquisition period (usually about 20 min). However, physiological processes in the body are usually dynamic and some organs (kidney and heart) show a significant decay of activity due to washout. Being able to trace activity in space and time is therefore of importance and is expected to significantly enhance diagnostic possibilities. A
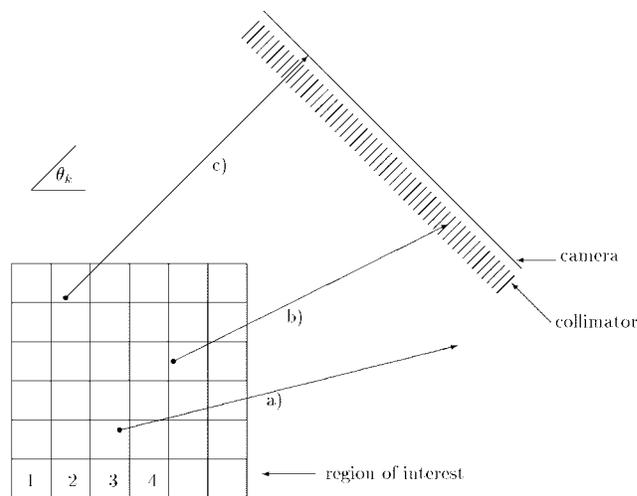
Fig. 1. Photons radiating from the region of interest. (a) Misses the camera. (b) Absorbed by the collimator. (c) Passes the collimator and hits the camera.

dynamic SPECT reconstruction, if numerically and algorithmically possible, might be presented as a movie rather than a static image.

With a significant decay of tracer activity over the acquisition period, the filtered back projection method (FBP) may no longer be reliably used, and new approaches must be developed. The purpose of the present paper is to introduce a probabilistic model for dynamic SPECT whose numerical solution is based on an instance of the EM algorithm. We discuss its numerical aspects along with those of other models as, for instance, presented in [20].

Let us mention that dynamic SPECT is currently already possible with a ring SPECT instrument, rarely available in hospitals, or with a triple head camera doing fast camera rotations. Here, the time-varying activity is handled by acquiring a number of fast sequential SPECT studies and fitting curves on a voxel-by-voxel basis to the resulting SPECT images (cf. [30]). This approach suffers from poor data statistics and is not compatible with the majority of the transmission scan-based attenuation-correction methods (cf. [32], [31], [9], [18], [19]). Our investigation aims at a dynamic method which works even for single-head cameras doing slow rotations adapted to the time scale of the tracer dynamics.

Clinical applications of dynamic SPECT which could be realized in the near future include renal studies using the dynamic tracers Tc-99mMAG3 and Tc-99mDTPA used in planar renal imaging, myocardial viability studies using I-123 labeled fatty acids, or brain blood-flow imaging using Xe-133, Xe127, or I-123, (cf. [15], [23]). The biological half lives of

any of these dynamic tracers is within the time scale of a typical scan of 20–25 min duration.

II. PROBABILISTIC MODEL

In order to model dynamics in SPECT we consider an instrument with a single camera head, available in most hospitals. The case of double- or triple- head cameras or even-ring SPECT devices could be treated similarly. The camera rotates through 180° with S stops, $60 \leq S \leq 120$, with each stop taking T/S minutes. T is the total acquisition time, usually some 20 min. The camera plane is of typical size 30 × 40 cm. Depending on its resolution, the camera is divided into M receptor bins (pixels), a typical choice being some 6 × 6 mm, in which case we have $M \approx 3000$ receptor elements. The camera rotates about a fixed axis with a radius of 20–30 cm.

The gamma rays originate from a region of interest of approximately 30 × 30 × 30 cm³. This region of interest is divided into N little cubes called voxels. In our example, assuming a spatial resolution of 0.5–1 cm leads to $N = 30^3$ up to $N = 60^3$ elements. During the kth stop the camera is at a fixed position with angle $\theta_k = (k-1) \cdot 180°/S$, $k = 1, \ldots, S$. What we are trying to reconstruct is the radio activity $x_{ik} := x_i(t_k)$ of the isotope in voxels $i = 1, \ldots, N$ at times t_k. The projected data y_{jk} measured during each stop k present the activities in the receptor bins $j = 1, \ldots, M$. Ideally, while the camera head is at position k only gammas traveling along a line with angle θ_k are allowed to pass the collimator and hit some photo cell (see Fig. 1).

We assume that activity is constant during the time interval of a single stop, but is allowed to vary in time over the whole acquisition period. Here, activity x_{ik} of voxel i during the kth stop is proportional to the number of photons that leave the voxel during this time interval, radiating in any possible direction.

Let Y_{jk} be the random variable which counts the number of events in the camera bin j during the stop k. The physics of radio activity make it reasonable to assume that the Y_{jk} are Poisson distributed. The data y_{jk} collected in the camera bins at different positions form a sample for the Y_{jk}.

Let c_{ijk} denote the probability that a photon leaves voxel i in direction θ_k and hits the detector bin j while the camera is actually at position k. The coefficients c_{ijk} are supposed to be known. They are usually referred to as the geometry of the model but, in addition to geometry, they may include probabilistic effects such as gamma ray scattering and, more importantly, attenuation and collimator blurring (see [1], [9], [36], [33]).

While including the collimator response function into the model parameters c_{ijk} is routinely possible, correcting for attenuation causes a major problem which, in practice, is handled in two ways. The first is to calculate the attenuation map beforehand, either by doing a transmission scan in parallel with the SPECT acquisition (see for instance [9]) or by estimating attenuation mathematically. For the latter approach see the method proposed in [24] and reported to be practical in [35]. A second way, at least mathematically possible,
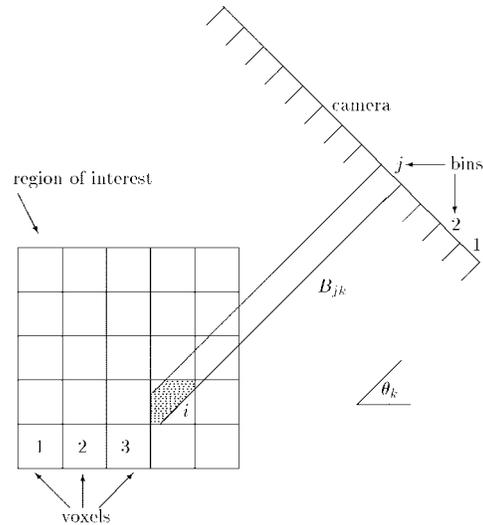

Fig. 2. The coefficient c_{ijk} is the relative volume of voxel i lying within the beam B_{jk} connecting i and the receptor bin j during the camera position θ_k.

consists in estimating the unknown attenuation and source simultaneously as proposed, for instance, in [6] and, more recently, in [13].

The way attenuation is estimated does not have any influence on our dynamic protocols. The coefficients c_{ijk} may, in principle, even comprise scattering, although the number of nonzero c_{ijk} then increases dramatically, leading to a numerically difficult problem.

In a simplified model where collimator blurring, attenuation, and scatter are ignored c_{ijk} is proportional to the volume of that part of voxel i which lies in the beam B_{jk} connecting i with the receptor bin j during the position k given by the angle θ_k (see Fig. 2). Based on the geometry, the expected values of the Y_{jk} are

$$E(Y_{jk}) = \sum_{i=1}^{N} c_{ijk} x_{ik} > 0. \tag{1}$$

For later use, let us fix some notation here. The coefficients c_{ijk} give rise to two linear operators. We define $C : \mathbb{R}^{NS} \to \mathbb{R}^{MS}$ by $(Cx)_{jk} = \sum_{i=1}^{N} c_{ijk} x_{ik}$ and $\Gamma : \mathbb{R}^{NS} \to \mathbb{R}^{NMS}$ by $(\Gamma x)_{ijk} = c_{ijk} x_{ik}$. We observe that, in practice, Γ is usually one to one, while C is typically not.

III. DYNAMIC TRACING

The critical part in modeling the dynamics of SPECT is the time behavior of activity $x_i(t)$. If radioactive decay were the only significant part, a standard model of the form $x_i(t) = A_i e^{-\lambda t}$ with known decay constant $\lambda > 0$ would apply. Dynamic SPECT is, however, based on the hypothesis that due to a flow in the organ, each voxel i may have its own individual decay profile. Therefore, more sophisticated models for decay with rates varying in time and space are needed. Based on experimental evidence for fatty acid myocardial viability [14], the following parametric model of decay:

$$x_i(t) = A_i e^{-\lambda_i t} + B_i e^{-\mu_i t} + C_i \tag{2}$$

with unknown parameters $A_i \geq 0, B_i \geq 0, C_i \geq 0$, and $\lambda_i > 0, \mu_i > 0$ was used in [20] and [5]. Discretizing (2) according to the S stops, $t_k = (k-1)T/S$, T the total acquisition time, leads to a parametric model with five degrees of freedom for each i.

Another approach suggested in the investigation of NMR-relaxation data (cf. [37]) uses a model of the form

$$x_i(t) = \int_0^\infty a_i(\lambda) e^{-\lambda t} \, d\lambda. \tag{3}$$

Discretizing the time axis at the S stops t_k now leads to a sum of exponentials with S degrees of freedom.

A third approach, which we propose here, is to not insist on any specific form of the decay curves. That is, in each position i we allow for an activity profile with S degrees of freedom x_{ik}, $k = 1, \ldots, S$, assuming only that activity decays in time

$$x_{i1} \geq x_{i2} \geq x_{i3} \geq \cdots \geq x_{iS} \quad \text{and} \quad x_{iS} \geq 0. \tag{4}$$

As compared to (2), this increases the flexibility of modeling decay, but increasing the degrees of freedom bares the risk of producing an under fitted model. Fortunately, experiments indicate (see Section VIII) that the model works well, and that its algorithmic and numeric advantages over (2) should be exploited. To introduce some standing notation we denote the set of $x = \{x_{ik}\}$ satisfying (4) by Ω.

IV. MAXIMUM LIKELIHOOD

Let us now derive the maximum-likelihood estimation for the unknown parameters $x \in \Omega$ based on the data y. Let $g(y; x)$ be the probability density of the measurements Y given the activity curves x. Then we consider

$$(\mathrm{ML}_1) \quad \begin{array}{ll} \text{maximize} & E(\log g(y; x)) \\ \text{subject to} & x \in \Omega \end{array}$$

i.e., we pick those parameters $x \in \Omega$ which were the ones most likely to have produced the data $y = \{y_{jk}\}$. As the Y_{jk} are Poisson variables with means $\sum_{i=1}^N c_{ijk} x_{ik}$, up to some constant the negative log-likelihood function $-\log g(y; x)$ equals

$$\phi(x) := \sum_{j=1}^M \sum_{k=1}^S \left(\sum_{i=1}^N c_{ijk} x_{ik} - y_{jk} \log \left(\sum_{i=1}^N c_{ijk} x_{ik} \right) \right).$$

The problem of minimizing $\phi(x)$ subject to the constraints (4) is feasible and admits optimal solutions since the objective clearly tends to infinity if at least one of the $x_{ik} \to \infty$ on the feasible set Ω. On the other hand, solutions are typically not unique. Any solution $x \in \Omega$ must satisfy the following Kuhn–Tucker conditions (cf. [10]). There exist multipliers $\lambda_{ik} \geq 0$, $i = 1, \ldots, N$, and $k = 1, \ldots, S$ such that for each

$i = 1, \ldots, N$ we have the system (KT_i)

$$\tau_{i1} - \sum_{j=1}^m \left(c_{ij1} y_{j1} \bigg/ \sum_{i'=1}^N c_{i'j1} x_{i'1} \right) - \lambda_{i1} \qquad = 0$$

$$\tau_{i2} - \sum_{j=1}^M \left(c_{ij2} y_{j2} \bigg/ \sum_{i'=1}^N c_{i'j2} x_{i'2} \right) + \lambda_{i1} - \lambda_{i2} \qquad = 0$$

$$\cdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \cdots$$

$$\tau_{iS-1} - \sum_{j=1}^M \left(c_{ijS-1} y_{jS-1} \bigg/ \sum_{i'=1}^N c_{i'jS-1} x_{i'S-1} \right)$$
$$+ \lambda_{iS-2} - \lambda_{iS-1} = 0$$

$$\tau_{iS} - \sum_{j=1}^M \left(c_{ijS} y_{jS} \bigg/ \sum_{i'=1}^N c_{i'jS} x_{i'S} \right) + \lambda_{iS-1} - \lambda_{iS} = 0$$

$$x_{ik} \geq x_{ik+1}, \qquad \lambda_{ik}(x_{ik} - x_{ik+1}) = 0$$
$$k = 1, \ldots, S-1, \qquad \lambda_{iS} x_{iS} = 0$$

where we used the abbreviation

$$\tau_{ik} := \sum_{j=1}^M c_{ijk} > 0. \tag{5}$$

Clearly, any solution x has $\sum_i c_{ijk} x_{ik} > 0$ for fixed j, k. Notice also that any two solutions $\bar{x}, \hat{x} \in \Omega$ must satisfy $\sum_i c_{ijk} \bar{x}_{ik} = \sum_i c_{ijk} \hat{x}_{ik}$ for every fixed j, k, since the functions $\eta \to \eta - y_{jk} \log \eta$ are strictly convex for $y_{jk} > 0$. As the operator given by $(Cx)_{jk} = \sum_{i=1}^N c_{ijk} x_{ik}$ is typically not one to one, (ML_1) may have multiple solutions. Ensuring that C is only moderately defective is of relevance for the numerics and may be influenced by the discretization we choose (cf. [5]).

Remark: We mention that the major difference with the widely known probabilistic model of Shepp and Vardi [34], [8] for static emission tomography is that, in the dynamic case, the model is extremely under fitted. For instance, in the situation of Section II, the unknown dynamic image has 64^3 degrees of freedom per slice, to which we fit of the order of 64^2 data (that is, 64 stops and 64 bins per camera cross-section). The situation improves if multiple camera heads are available, but the problem remains under fitted. To render it tractable, a prior model for the dynamics such as (2), (3), or (4) is required, represented by the constraint set Ω.

V. THE EM ALGORITHM

The EM algorithm is an iterative procedure to calculate maximum likelihood estimates. It first made its appearance in 1976 in a paper by Dempster *et al.* [12] (see also [22]) and its original intention was maximum-likelihood estimation with incomplete data. Since then, the EM algorithm has been applied in a much wider context, including situations in which the incomplete and the complete data space are defined in a somewhat artificial way.

In our present situation, the maximum likelihood problem (ML_1) involving the joint probability distribution $g(y; x)$ of Y represents the incomplete data space. To introduce a complete data space we consider the random variables Z_{ijk} which present that part of the activity in voxel i that radiates toward the receptor bin j during stop k. These are Poisson random

variables with mean $c_{ijk}x_{ik}$ and joint probability density $f(z; x)$ depending on the parameter $x \in \Omega$.

The EM algorithm is now the following. By induction define a sequence of parameter estimates $x^\alpha = \{x_{ik}^\alpha\} \in \Omega$, $\alpha = 1, 2, \ldots$ according to the following rules.

1) *E step:* Given $x^\alpha = \{x_{ik}^\alpha\} \in \Omega$ and the data y calculate the conditional expectation $z^\alpha = E(Z \mid Y = y, x^\alpha)$.
2) *M step:* Calculate the new estimate $x^{\alpha+1} \in \Omega$ by maximizing $E(\log f(z^\alpha; x) \mid y; x^\alpha)$ over $x \in \Omega$.

The E step is fully explicit here. Namely, given the data y and the current x^α we find that

$$z_{ijk}^\alpha = y_{jk} \cdot \frac{c_{ijk}x_{ik}^\alpha}{\sum_{i'=1}^N c_{i'jk}x_{i'k}^\alpha}. \tag{6}$$

This may, in fact, be obtained from the following.

Lemma 1: Let $X_1, \ldots, X_n$ be independent Poisson distributed random variables with mean $E(X_k) = \mu_k$. Then the conditional expectation

$$E((X_1, \ldots, X_n) \mid X_1 + \cdots + X_n = d) =: (x_1, \ldots, x_n)$$

is obtained as

$$x_k = d \cdot \frac{\mu_k}{\mu_1 + \cdots + \mu_n}.$$

For the M step, we maximize the log-likelihood function $\log f(z^{\alpha+1}; x)$ over the $x \in \Omega$. Up to constant terms, that entails minimizing

$$\psi(x) := \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^S \left(c_{ijk}x_{ik} - z_{ijk}^\alpha \log(c_{ijk}x_{ik}) \right)$$

over all $x \in \Omega$. Obviously, this function is of the form $\psi = \sum_i \psi_i$ with each

$$\psi_i(x^i) = \sum_{j=1}^M \sum_{k=1}^S \left(c_{ijk}x_{ik} - z_{ijk}^\alpha \log(c_{ijk}x_{ik}) \right)$$

depending only on the variable $x^i = (x_{i1}, \ldots, x_{iS}) \in \Omega_i$, the set of x^i satisfying (4) for fixed i. Notice that the data y and the previous iterate x^α do not enter directly into the calculation of $x^{\alpha+1}$. Rather, is it z^α which makes connection with the previous step.

We have

$$\frac{\partial \psi_i}{\partial x_{ik}} = \sum_{j=1}^M \left(c_{ijk} - \frac{z_{ijk}^\alpha}{x_{ik}} \right) = \tau_{ik} - \frac{\sigma_{ik}^\alpha}{x_{ik}}$$

with τ_{ik} as in (5) and

$$\sigma_{ik}^\alpha := \sum_{j=1}^M z_{ijk}^\alpha > 0. \tag{7}$$

We now consider the maximum-likelihood problem for the complete data space which perforce splits into N problems of size S

$$(ML_2) \quad \begin{array}{ll} \text{maximize} & E(\log f(z^\alpha; x) \mid y; x^\alpha) \\ \text{subject to} & x \in \Omega. \end{array}$$

Setting $\psi_{ik}(x^i) := -x_{ik} + x_{i,k+1}$, and with ψ_i defined as before, the ith problem becomes

$$(ML_{2,i}) \quad \begin{array}{ll} \text{minimize} & \psi_i(x^i) \\ \text{subject to} & \psi_{ik}(x^i) \leq 0 \quad (k = 1, \ldots, S-1). \end{array}$$

Notice that in contrast with the original maximum-likelihood problem (ML_1), we do not control the constraints $x_{iS} \geq 0$ here, since they are built in through the objective due to $\sigma_{iS}^\alpha > 0$. In fact, for fixed i, k, consider j such that both $c_{ijk} > 0$ and $z_{ijk}^\alpha > 0$. Then the corresponding term in $\psi_i(x^i)$ works as a barrier function which forces the variable x_{ik} to take strictly positive values. For the same reason, it is now reasonably clear that the optimal solutions $x^{i,\alpha+1}$ for $(ML_{2,i})$ are unique since the ψ_i are strictly convex and coercive.

The Kuhn–Tucker conditions (cf. [10]) for the optimal solutions $x^{i,\alpha+1}$ of $(ML_{2,i})$ imply the existence of multipliers $\lambda_{ik}^\alpha \geq 0$, $k = 1, \ldots, S-1$, giving the system $KT_{2,i}$

$$\begin{array}{llll}
\tau_{i1} & -\dfrac{\sigma_{i1}^\alpha}{x_{i1}^{\alpha+1}} & -\lambda_{i1}^\alpha & = 0 \\[2mm]
\tau_{i2} & -\dfrac{\sigma_{i2}^\alpha}{x_{i2}^{\alpha+1}} & +\lambda_{i1}^\alpha - \lambda_{i2}^\alpha & = 0 \\[1mm]
\cdots & & & \cdots \\[1mm]
\tau_{i,S-1} & -\dfrac{\sigma_{i,S-1}^\alpha}{x_{i,S-1}^{\alpha+1}} & +\lambda_{i,S-2}^\alpha - \lambda_{i,S-1}^\alpha & = 0 \\[2mm]
\tau_{iS} & -\dfrac{\sigma_{iS}^\alpha}{x_{iS}^{\alpha+1}} & +\lambda_{i,S-1}^\alpha & = 0
\end{array}$$

$$x_{ik}^{\alpha+1} \geq x_{ik+1}^{\alpha+1} \qquad \lambda_{ik}^\alpha \left(x_{ik}^{\alpha+1} - x_{ik+1}^{\alpha+1} \right) = 0,$$
$$k = 1, 2, \ldots, S-1.$$

Based on the optimality conditions $(KT_{2,i})$, we may now enter a detailed analysis of the EM algorithm.

VI. ALTERNATING PROJECTIONS

We show that the EM algorithm presented in the previous section may be interpreted as an alternating projection procedure, in the sense of von Neumann, provided that the Euclidean metric is replaced with the Kullback–Leibler distance. The interested reader is referred to the Appendix for a brief account of von Neumann's alternating projections.

Let $n = NS$ and $m = MS$, $p = NMS$. The Kullback–Leibler distance of $a, b \in \mathbb{R}_+^p$ is defined as

$$d(a, b) = \sum_{i=1}^p a_i \log(a_i/b_i) - a_i + b_i.$$

Now d has properties which resemble those of a metric, but obviously lacks symmetry and is only defined for $a \geq 0$ and $b > 0$. In analogy with the orthogonal projection we define projection operators $P_A^{\rightarrow}$ and $P_A^{\leftarrow}$ associated with d. Given a closed convex subset A of $\mathbb{R}_+^p$ and a point $x \in \mathbb{R}_+^p$, the forward, respectively backward, projections of x onto A are defined as

$$P_A^{\rightarrow}(x) = \arg\min_{a \in A} d(a, x)$$

and

$$P_A^{\leftarrow}(x) = \arg\min_{a \in A} d(x, a).$$

With these notions, the M step presented in Section V turns out to be a Kullback–Leibler backward projection.

Proposition 1: Let $A := \{v \in \mathbb{R}^{NMS} : v_{ijk} = c_{ijk}x_{ik}$ for some $x \in \Omega\}$. Given $z^\alpha \in \mathbb{R}^{NMS}$ let $v^{\alpha+1}$ be the backward projection of z^α onto A, that is

$$v^{\alpha+1} = P_A^{\leftarrow}(z^\alpha).$$

Then the solution $x^{\alpha+1}$ of the M step satisfies $v_{ijk}^{\alpha+1} = c_{ijk}x_{ik}^{\alpha+1}$, for short, $v^{\alpha+1} = \Gamma x^{\alpha+1}$.

Proof: Indeed, for the proof we simply observe that with $v_{ijk} = c_{ijk}x_{ik}$, for short, $v = \Gamma x$, the objective function ψ of the M-step satisfies $\psi(x) = \sum_{ijk}(v_{ijk} - z_{ijk}^\alpha \log v_{ijk})$ which up to a constant term equals $d(z^\alpha, v)$. Minimizing $\psi(x)$ over $x \in \Omega$ is therefore equivalent to minimizing $d(z^\alpha, v)$ over $v \in A$. Existence and unicity of the projection is guaranteed since $z^\alpha > 0$ and the set A contains points in $\mathbb{R}_{++}^{NMS}$. Notice also that A is closed, as it is the linear image of the closed convex and polyhedral set Ω under Γ, and is therefore itself polyhedral. $\square$

Based on this observation we now ask whether, similarly, the E step may be interpreted as a projection onto a convex set.

Proposition 2: Let $B := \{z \in \mathbb{R}^{NMS} : z \geq 0, \sum_{i=1}^N z_{ijk} = y_{jk} \forall j, k\}$. Let x^α be the current EM iterate and let $v^\alpha = \Gamma x^\alpha$. Then the conditional expectation z^α defined by the E step (6) is the Kullback–Leibler projection of v^α onto B in the forward and the backward sense

$$z^\alpha = P_B^{\rightarrow}(v^\alpha) = P_B^{\leftarrow}(v^\alpha).$$

Proof: First, consider the case of the backward projection. Clearly, $z = P_B^{\leftarrow}(v^\alpha)$ exists and satisfies the Kuhn–Tucker conditions. There exist multipliers λ_{jk} such that

$$-\frac{v_{ijk}^\alpha}{z_{ijk}} + 1 + \lambda_{jk} = 0 \quad \forall i, j, k$$

in tandem with the constraints

$$\sum_{i=1}^N z_{ijk} = y_{jk} \quad \forall j, k.$$

Summing over i for fixed j, k gives $(1 + \lambda_{jk})y_{jk} = \sum_i v_{ijk}^\alpha$. This readily implies (6).

Next, consider the case of the forward projection $z = P_B^{\rightarrow}(v^\alpha)$. Here the Kuhn–Tucker conditions provide multipliers λ_{jk} satisfying

$$\log(z_{ijk}/v_{ijk}^\alpha) + \lambda_{jk} = 0 \quad \forall i, j, k$$

and with constraints as above. Taking exponentials and summing over i for fixed j, k gives the same type of relation and again leads to (6). $\square$

It remains to observe that any limit point of the sequence x^α must be a Kuhn–Tucker point for the problem (ML_1). This is a consequence of a general fact (see [38]), but could equally well be checked using the conditions (KT_1) in tandem with (6) and (7) and the $(KT_{2,i})$. Indeed, convergence of the iterates v^α, z^α implies convergence of x^α, σ^α. Formula (KKT_i) implies

$$\lambda_{ik}^\alpha = \sum_{j=1}^k \left(\tau_{ij} - \sigma_{ij}^\alpha / x_{ij}^\alpha\right)$$

so the multipliers converge as well. Passing to the limit $\alpha \to \infty$ therefore implies the corresponding conditions (KT_1) and it remains to check that part of the Kuhn–Tucker conditions (KT_1), which concerns the constraints $x_{iS} \geq 0$. Clearly, however, if any such constraint is active $x_{iS} = 0$, the corresponding conditions in (KT_1) are satisfied. In case $x_{iS} > 0$, the multiplier $\lambda_{iS} = 0$ will do.

In consequence, we may now state the principal convergence result for the present Poisson EM algorithm.

Theorem 1: The sequences $v^\alpha = \Gamma x^\alpha$ and z^α generated by the Poisson EM algorithm based on the constraint set (4) converge to limit points $v^\alpha \to \bar{v} \in A$ $z^\alpha \to \bar{z} \in B$. Here, $\bar{v}$ is of the form $\bar{v} = \Gamma \bar{x}$, with $\bar{x} \in \Omega$ a solution to the maximum-likelihood problem (ML_1).

Proof: First, observe that the problem (ML_1) is convex and, due to the structure of the constraint set Ω, admits optimal solutions. We claim that the sequences of alternating Kullback–Leibler projections

$$z^\alpha = P_B^{\rightarrow}(v^\alpha) = P_B^{\leftarrow}(v^\alpha) \quad v^{\alpha+1} = P_A^{\leftarrow}(z^\alpha)$$

converge to certain limit points $v^\alpha \to \bar{v} \in A$, $z^\alpha \to \bar{z} \in B$. In the case where $A \cap B \neq \emptyset$, the alternating sequence converges to a common limit point $\bar{v} = \bar{z} \in A \cap B$. This is a known fact proved by Bregman (see, e.g., [2, Th. 8.1]). In case $A \cap B = \emptyset$, which is the more realistic one, the convergence proof is more complicated and only special cases have been considered. (See, for instance, [16], [17], [34], and [7], where the case of positivity constraints $x_{ik} \geq 0$ was discussed.) Our present case, involving the constraint set (4), may be settled using the same type of reasoning. We skip over the details, which are tedious.

Finally, notice that the point $\bar{x}$ satisfying $\bar{v} = \Gamma \bar{x}$ is uniquely determined by the injectivity of Γ and by closedness lies in Ω. Inspecting the Kuhn–Tucker conditions for (ML_1), we see that $\bar{x}$ must be a Kuhn–Tucker point and, by convexity, therefore solves (ML_1). $\square$

Remark: Consequently, the EM algorithm is best understood in the space of v variables, where it may be recast as an alternating Kullback–Leibler projection onto closed convex sets A and B. In fact, B is an affine subspace restricted to $\mathbb{R}_+^p$, while $A = \Gamma(\Omega)$ depends on the parameter space Ω and may therefore be quite complicated. The whole procedure may be generalized to any comparable situation with a closed convex parameter space Ω.

Remark: Replacing the constraint set Ω by $\tilde{\Omega}$, representing the nonlinear conditions (2), we obtain a nonconvex set $\tilde{A} = \Gamma(\tilde{\Omega})$. The EM algorithm could then still be formulated as an alternating projection between $\tilde{A}$ and B, but with the obvious problems when projecting onto nonconvex sets.

Remark: Iusem [16], among others, considers the case where we impose only positivity conditions $\hat{\Omega} = \{x : x_{ik} \geq 0\}$, $\hat{A} = \Gamma(\hat{\Omega})$. Here, the backward projection onto $\hat{A}$ is particularly pleasant to calculate and leads to an explicit formula.

VII. THE LEAST SQUARES APPROACH

In [20], the authors propose a different approach to the dynamic SPECT problem, which is based on the dynamic

model (2) and is solved using nonlinear least squares

$$
\text{(NLS)} \quad
\begin{aligned}
&\text{minimize} && \|Cx - y\| \\
&\text{subject to} && x_{ik} = A_i e^{-\lambda_i t_k} + B_i e^{-\mu_i t_k} + C_i \\
& && A_i \geq 0, \quad B_i \geq 0, \quad C_i \geq 0, \quad \lambda_i \geq 0,
\end{aligned}
$$

Here, the operator $C : \mathbb{R}^n \to \mathbb{R}^m$ is defined as before

$$
(Cx)_{jk} := \sum_{i=1}^N c_{ijk} x_{ik}.
$$

Problem (NLS) was found difficult to treat in practice due to the highly nonlinear dependence of the x_{ik} on the parameters (cf. [20], [5]). Following the ideas presented in Section III, we replace the parametric model (2) by (4). This leads to a linear least squares problem with inequality constraints

$$
\text{(LLS)} \quad
\begin{aligned}
&\text{minimize} && \|Cx - y\| \\
&\text{subject to} && x \in \Omega
\end{aligned}
$$

where Ω is, as before, the set of x satisfying (4). Our aim now is to present a probabilistic model for both (NLS) and (LLS). This will also provide clues to possible numerical approaches based on a version of the EM algorithm.

Following the pathway in Section V, we let the random variables Z_{ijk} and Y_{jk} be defined accordingly, but now assume that the Z_{ijk} are independent and normal variables with unknown mean $c_{ijk} x_{ik}$ and common variance $\sigma^2 > 0$. Consequently, the Y_{jk} are independent normal variables with mean $\sum_{i=1}^N c_{ijk} x_{ik}$ and variance $N\sigma^2$.

With these agreements, (LLS) is just the maximum-likelihood estimation problem

$$
\text{(ML}_1\text{)} \quad
\begin{aligned}
&\text{maximize} && E(\log g(y; x)) \\
&\text{subject to} && x \in \Omega
\end{aligned}
$$

where $g(y; x)$ denotes the density of the normal law Y with mean Cx and covariance matrix $N\sigma^2 I_m$. Similarly, problem (NLS) is the maximum likelihood problem with the constraint set $\tilde{\Omega}$ given by (2).

Returning to the outline of Section V, we apply the EM algorithm to the present situation, letting Z, respectively Y, represent the complete, respectively incomplete, data spaces. As in the Poisson case, this generates sequences x^α and z^α, $\alpha = 1, 2, \ldots$ according to the following rules.

1) *E step.* Given the current iterate $x^\alpha \in \Omega$, calculate the conditional expectation $z^\alpha = E(Z \mid Y = y; x^\alpha)$.
2) *M step.* Calculate the new iterate $x^{\alpha+1} \in \Omega$ by maximizing $E(\log f(z^\alpha; x) \mid y; x^\alpha)$ over $x \in \Omega$.

Here, $f(z; x)$ denotes the normal law with mean Γx and covariance matrix $\sigma^2 I_p$, which is the joint density of Z.

Proposition 3: Let $A = \{v \in \mathbb{R}^p : v = \Gamma x \text{ for some } x \in \Omega\}$. Then, given the result z^α of the previous E step, the next M step reduces to calculating the orthogonal projection $v^{\alpha+1}$ of z^α onto A and taking $x^{\alpha+1}$ to satisfy $v^{\alpha+1} = \Gamma x^{\alpha+1}$.

Proof: Indeed, the negative log-likelihood function used for the M step, up to constant terms, equals $\|z - \Gamma x\|^2 / (N\sigma^2)$ and is to be minimized over $x \in \Omega$. This establishes the statement. $\square$

Let us now pass to the E step, which we would like to reveal as an orthogonal projection. This involves calculating

the conditional expectation $E(Z \mid \sum_{i=1}^N Z_{ijk} = y_{jk}; x^\alpha)$. Due to independence of Z, this may be obtained from the following.

Lemma 2: Let $X = (X_1, \ldots, X_n)$ be a vector of independent normal variables with mean $E(X_i) = \mu_i$ and variance $\sigma^2 = 1$. Then the conditional expectation $x = E(X \mid X_1 + \cdots + X_n = y)$ satisfies

$$
x_i = \frac{1}{n} y + \mu_i - \frac{1}{n} \sum_{i'=1}^n \mu_{i'}.
$$

More precisely, x is the orthogonal projection of $(\mu_1, \ldots, \mu_n)$ onto the set of $(x_1, \ldots, x_n)$ satisfying $x_1 + \cdots + x_n = y$.

Proof: In consequence of [29, Th. 2.1(viii)], a normal vector

$$
\begin{pmatrix} a \\ b \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} U & W \\ W^* & V \end{pmatrix} \right)
$$

gives the conditional expectation

$$
E(b \mid a) \sim \mathcal{N}(\beta + W^* U^{-1}(a - \alpha), V - W^* U^{-1} W).
$$

We apply this in the case $b = (X_1, \ldots, X_n)$, $a = X_1 + \cdots + X_n$ which gives the stated formula. $\square$

With this observation, we find that the formula replacing (6) in the case of the normal variables is

$$
z_{ijk}^\alpha = \frac{1}{N} y_{jk} + c_{ijk} x_{ik}^\alpha - \frac{1}{N} \sum_{i'=1}^N c_{i'jk} x_{i'k}^\alpha \tag{8}
$$

and the E step is in fact realized as an orthogonal projection:

Proposition 4: Let $B = \{z \in \mathbb{R}^p : \sum_{i=1}^N z_{ijk} = y_{jk} \ \forall j, k\}$. Let x^α be the current iterate generated by the previous M step and let $v^\alpha = \Gamma x^\alpha$. Then the result z^α of the next E step is the orthogonal projection of v^α onto B.

We are now in the position to state the convergence result for the EM algorithm in the case of normal variables. It is based on von Neumann's classical method of alternating projections.

Theorem 2: Let A, B be defined as above and let $z^\alpha, v^\alpha = \Gamma x^\alpha$ be the sequences generated by the Gaussian EM algorithm. Then z^α, v^α converge to limit points $v^\alpha \to \bar{v} \in A$, $z^\alpha \to \bar{z} \in B$. Here, $\bar{v}$ is of the form $\bar{v} = \Gamma \bar{x}$ with $\bar{x} \in \Omega$ a solution to the least squares problem (LLS).

Proof: In the case where $A \cap B \neq \emptyset$, it is well known that the sequence of alternating projections converges to a common limit point $\bar{v} = \bar{z} \in A \cap B$ which, by definition of A, is then of the form $\Gamma \bar{x}$.

The more involved case occurs when $A \cap B = \emptyset$. Following [3], a dichotomy appears. Either the alternating sequences converge to limit points $v^\alpha \to \bar{v} \in A$, $z^\alpha \to \bar{z} \in B$, with $\bar{v}, \bar{z}$ realizing the distance between A, B, or there are no points realizing this distance, in which case the sequences tend to infinity $\|v^\alpha\| \to \infty$, $\|z^\alpha\| \to \infty$, with $\|v^\alpha - z^\alpha\|$ approaching the distance between A, B. We will argue that in our situation the second case is impossible.

Indeed, the distance $d(A, B)$ not being attained implies the existence of a common asymptotic direction for A, B: there exist $w \neq 0$ and $\hat{a} \in A$, $\hat{b} \in B$ having $\hat{a} + \mathbb{R}_+ w \subset A$, $\hat{b} + \mathbb{R}_+ w \subset B$. The second inclusion implies $\sum_i w_{ijk} = 0$ for fixed j, k, while the first gives $w = \Gamma x$ for some $x \in \Omega$.

The c_{ijk} being nonnegative, this is only possible when $x = 0$ and, hence, $w = 0$.

We deduce that $v^\alpha \to \bar{v} = \Gamma \bar{x}$ and $z^\alpha \to \bar{z}$. It is now routine to check that $\bar{x}$ is a Kuhn–Tucker point for the original problem (LLS). □

Remark: Replacing Ω by $\hat{\Omega} = \{x : x_{ik} \geq 0\}$ gives an important special case. The orthogonal projection onto $\hat{A} = \Gamma(\hat{\Omega})$ is explicit and, in tandem with (8), leads to a formula for updating x^α. We then have an iterative method for calculating nonnegative least squares solutions. One may compare this to other iterative methods for calculating nonnegative least squares solutions such as ISRA (see [27]).

Let us analyze the M step a little further. Recall that up to a constant additive term and a constant factor, the negative log-likelihood function $-\log f(z^\alpha, x)$ equals

$$\psi(x) = \frac{1}{2} \sum_{i,j,k} \left(z^\alpha_{ijk} - c_{ijk} x_{ik} \right)^2$$

and is to be minimized over Ω. Setting

$$\gamma_{ik} := \sum_{j=1}^{M} c_{ijk}^2 > 0 \quad \text{and} \quad \delta_{ik} := \sum_{j=1}^{M} c_{ijk} z^\alpha_{ijk} > 0 \qquad (9)$$

we have

$$\psi(x) = \frac{1}{2} \sum_{ik} \left(\gamma_{ik} x_{ik}^2 - 2\delta_{ik} x_{ik} \right) + \text{constant terms.}$$

The Kuhn–Tucker conditions therefore imply the existence of Lagrange multipliers $\lambda_{ik} \geq 0$, $i = 1, \ldots, N$ and $k = 1, \ldots, S$ satisfying (LKT_i)

$$
\begin{array}{llll}
\gamma_{i1} x_{i1} & -\delta_{i1} - \lambda_{i1} & & = 0 \\
\gamma_{i2} x_{i2} & -\delta_{i2} + \lambda_{i1} - \lambda_{i2} & & = 0 \\
\cdots & \cdots & & \cdots \\
\gamma_{iS-1} x_{iS-1} & -\delta_{iS-1} & + \lambda_{iS-2} - \lambda_{iS-1} & = 0 \\
\gamma_{iS} x_{iS} & -\delta_{iS} & + \lambda_{iS-1} - \lambda_{iS} & = 0
\end{array}
$$

and

$$x_{ik} \geq x_{ik+1}, \quad \lambda_{ik}(x_{ik} - x_{ik+1}) = 0, \quad \text{for } k = 1, \ldots S - 1$$
$$x_{iS} \geq 0, \quad \lambda_{iS} x_{iS} = 0.$$

Therefore, analogously to the Poisson case, the original problem of size NS splits into N problems of size S. Notice that in contrast with the Poisson case $(ML_{1,i})$, we do have to control the constraints $x_{iS} \geq 0$ here, which leads to the additional multipliers λ_{iS}. However, as we shall see, these constraints will generally be inactive and the multipliers will vanish.

VIII. NUMERICAL APPROACH

In this section we shall discuss the practical aspects of both EM algorithms. Clearly, the algorithms may be expected to be slow since alternating projections are known to converge no better than with a linear rate. Nevertheless, the additional numerical stability gained may often justify the EM scheme, in particular, if some speed is recovered, e.g., by parallelizing the M step. The crucial question to be addressed before proposing this scheme is the following. Observe that in the dynamic case the M-step is intrinsically more complicated than in

the stationary case, because it requires solving a nonlinear optimization problem $(ML_{2,i})$ where the static case goes with an explicit formula, (cf. [17], [7]). We therefore have to assure that a reasonably fast numerical solution is possible. In order to answer this question, we shall analyze the M steps of both the Poisson and the Gaussian model.

The first surprising observation is that both M steps, although coming from completely different out sets, lead to exactly the same procedure. Recall the Kuhn–Tucker conditions $(KT_{2,i})$ for the Poisson M step presented in Section V. To simplify notation we omit the indexes i and α which are fixed, writing $\tau_k := \tau_{ik}$ and $\sigma_k := \sigma^\alpha_{ik}$ and, similarly, $x_k := x^\alpha_{ik}$ for the solution and the multipliers. We then have the following.

Lemma 3: For fixed i the problem $(ML_{2,i})$ has a unique solution $(x_1, \ldots, x_S)$ satisfying (4). This solution is of the following form:

$$
\begin{aligned}
x_1 = \cdots = x_{r_1} &> x_{r_1+1} = \cdots = x_{r_2} \\
&> \cdots > x_{r_{t-1}+1} = \cdots = x_{r_t} \qquad (10)
\end{aligned}
$$

for appropriate $0 = r_0 < r_1 < \cdots < r_t = S$ with

$$x_{r_{j-1}+1} = \cdots = x_{r_j} =: x^{(j)} = \frac{\sigma_{r_{j-1}+1} + \cdots + \sigma_{r_j}}{\tau_{r_{j-1}+1} + \cdots + \tau_{r_j}} \quad (11)$$

the $x^{(j)}$ being strictly decreasing. In particular, $x_1 = x^{(1)} \geq \frac{\sigma_1}{\tau_1}$ and $x_S = x^{(t)} \leq \frac{\sigma_S}{\tau_S}$.

Proof: Uniqueness of the solution of $(ML_{2,i})$ follows from strict convexity of the objective. Clearly, any solution, since it satisfies (4), has the form (10) and it only remains to establish (11).

The complementarity condition in $(KT_{2,i})$ implies that $\lambda_{r_j} = 0$ for all $j = 1, \ldots, t$, while the remaining multipliers may be strictly positive. Now, summing the Kuhn–Tucker equations $(KT_{2,i})$ in each block $r_{j-1} + 1, \ldots, r_j$ separately gives (11). □

Remark: This result is interesting since it tells us that given any sequences $\sigma_k > 0, \tau_k > 0$, $k = 1, \ldots, S$, there exists a unique subdivision (10) such that $x^{(j)}, j = 1, \ldots, t$, defined through (11), is strictly decreasing. (Taking, for instance, $\tau_k = 1$ we obtain the amusing observation that, given any sequence $\sigma_k > 0$, there exist a unique subdivision (10) such that the arithmetic means of the σ_i over each block are strictly decreasing and satisfy the boundary conditions $x^{(1)} \geq \sigma_1$ and $x^{(t)} \leq \sigma_S$.)

Let us now pass to the Gaussian M step obtained in Section VII. Here we have the same observation which, again on suppressing the indexes i, α, is the following:

Lemma 4: The solution $(x_1, \ldots, x_S)$ of the Kuhn–Tucker conditions (LKT_i) is unique and of the form (10), possibly with a different t. It admits a representation of the form (11) with δ_i replacing σ_i and γ_i replacing τ_i. The sequence $x^{(j)}$ is strictly decreasing and, again, the limit conditions $x_1 = x^{(1)} \geq \frac{\delta_1}{\gamma_1}$, $x_S = x^{(t)} \leq \frac{\delta_S}{\gamma_S}$ are satisfied.

Proof: Indeed, starting out with the blocks (10), possibly with another t, we find again that the multipliers λ_{r_j} must vanish. Summing the Kuhn–Tucker conditions (LKT_i) in each
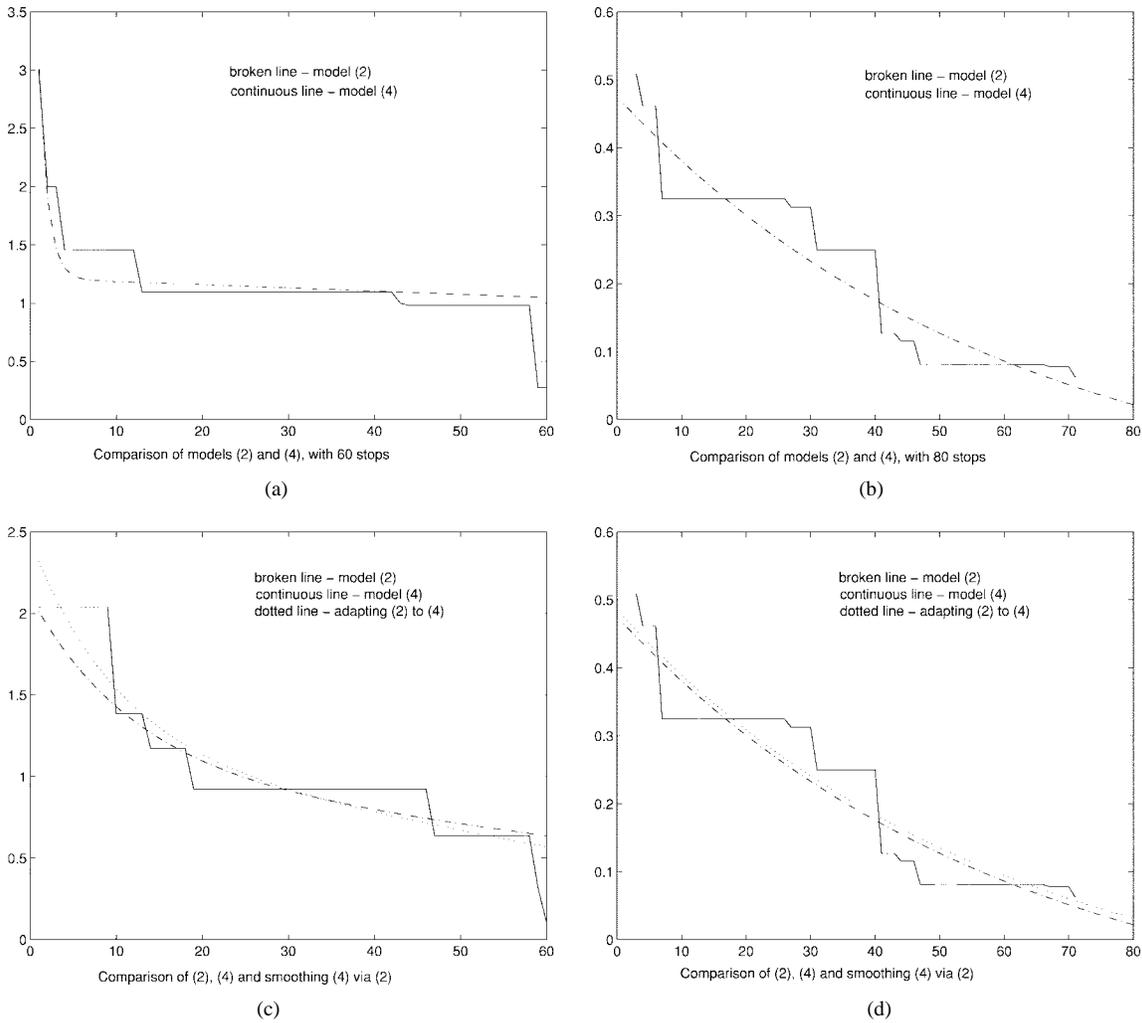
Fig. 3. Comparisons of models (2) and (4) using typical dynamic pixels from two simulated reconstructions on the basis of 60 (left column) and 80 (right column) camera positions. In the lower line an *a posteriori* fit of (2) to the data of (4) shows little difference with model (2).

block would suffice, provided we knew that the final multiplier λ_S, belonging to the constraint $x_S \geq 0$, had vanished. Clearly, this is the case if $x_S > 0$, which is consequently what remains to be checked.

Assume to the contrary that the last block in (10) is zero, $x^{(t)} = x_{r_{t-1}+1} = \cdots = x_S = 0$. Then the Kuhn–Tucker conditions give

$$\lambda_S = -\sum_{k=r_{t-1}+1}^{S} \delta_k$$

which, in view of $\lambda_S \geq 0$ and $\delta_k \geq 0$, is only possible when $\delta_{r_{t-1}+1} = \cdots = \delta_S = 0$. However, by construction (9) we have $\delta_i > 0$ so $\lambda_S = 0$, $x_S > 0$ and the result follows. $\square$

The observation that both M steps are essentially equivalent, although with different data $\sigma_k > 0, \tau_k > 0$ versus $\delta_k > 0, \gamma_k > 0$, gives us choice on how to perform the M step. In fact, the Gaussian M step uses nonnegative least squares, and for a moderate size S works faster than the Poisson M step ($\mathrm{ML}_{2,i}$). On the other hand, for a really large S, if required, the Poisson M step could more conveniently be solved by an interior point method. In fact, a logarithmic barrier term for

the constraints (4) leads to the objective

$$\phi_\mu(x) = \sum_{k=1}^{S} \tau_k x_k - \sigma_k \log x_k - \mu \sum_{k=1}^{S-1} \log(x_k - x_{k+1}).$$

One might perform only a few steps toward minimizing ϕ_μ for a fixed $\mu > 0$ before passing to another E step. Increasing the number of steps and reducing μ should then be controlled by testing decrease of the likelihood function in (ML_1), respectively in (LLS), which, for theoretical reasons (see [12]), is known to decrease at each iteration if an exact EM step is performed.

Experiment: Using the Poisson EM algorithm for the dynamic model (4), with an M step based on a quadratic program (justified through Lemma 3 and Lemma 4), we reconstruct a typical slice of 64×64 pixels using 60–80 stops within the order of 40–50 min CPU, with up to 100 EM iterations. Performance can often be improved if a good prior guess is available. A considerable speedup could, however, be obtained by parallelizing the M step. A more detailed evaluation of our present approach including comparison to other dynamic methods, shall be presented in [21].

The experiment reported in Fig. 3 is to demonstrate that replacing the mixed-exponential decay model (2) by the more general decay model (4) usually does not lead to a loss of information. The four pictures show randomly chosen dynamic pixels from two simulated reconstructions on the basis of 60 [Fig. 3(a) and (c)], respectively 80 [(Fig. 3(b) and (d)], camera positions. The plots in the top line show comparisons of typical decay curves produced by the models (2) (smooth semicolon curves) versus (4) (step-like curves).

The Fig. 3(c) and (d) indicates that, typically, no information is lost on replacing (2) with (4). Namely, the smooth dotted lines show model (2) curves which *a posteriori* have been fitted to the step-like functions (4). Typically, these show little difference from the direct fitting of (2) (smooth semicolon type curves as before) so (2), if desired, could be retrieved from (4). As (4) has numerical advantages, the EM algorithm should in fact be built on (4). Fitting a model (2) decay curve to the step functions (4) may be deferred to the end of the procedure.

IX. CYCLIC PROJECTIONS

The fact that the Kullback–Leibler, respectively orthogonal, projections onto the set A cannot be calculated explicitly is clearly a drawback of our present approach, accounting for the fact that the speed is still inconvenient for clinical applications. In addition to parallelizing the M step, this problem may be addressed by the following idea. Split the projection onto A into two projection steps which are easier to compute and combine the two resulting points to approximate $P_A^{\leftarrow}(z)$. More precisely, let us consider the sets

$$A_1 := \{v \in \mathbb{R}^p : v = \Gamma x, x_{i1} \geq x_{i2}, x_{i3} \geq x_{i4}, \ldots\}$$

and

$$A_2 := \{v \in \mathbb{R}^p : v = \Gamma x, x_{i2} \geq x_{i3}, x_{i4} \geq x_{i5}, \ldots\}.$$

Then, $A = A_1 \cap A_2$ and instead of projecting onto A we wish to use the projections onto the A_i which, as we shall see, are easier to perform.

Inspecting a simple case where the sets A_1, A_2, B are the edges of an equilateral triangle shows that projecting cyclically onto the sets is not exactly what we want. There are, however, various possible ways in which the projected information onto the three sets A_1, A_2, B may be used to approach the point $a \in A, b \in B$, realizing the distance between the two sets. We propose the following scheme, which applies to both the Poisson and the Gaussian case with the corresponding interpretations.

1) Given an iterate x^{α}, respectively $v^{\alpha} = \Gamma x^{\alpha}$, do an E step using (6), respectively, (8). The result is z^{α}.
2) Replace the M step by the following. Project z^{α} onto A_i, which gives $v^i = P_{A_i}^{\leftarrow}(z^{\alpha})$, $i = 1, 2$. To form the new iterate $x^{\alpha+1}$, take $v^{\alpha+1} = (v^1 + v^2)/2$, $v^{\alpha+1} = \Gamma(x^{\alpha+1})$.
3) At the end of the procedure do a few correction steps by projecting v^{α} onto the true set A.

Let us see why this scheme, leading to explicit formulas, is expected to be faster than the original EM algorithm. The crucial observation is the following.

Lemma 5: With the notation (7) and (5) let $v^1 = P_{A_1}^{\leftarrow}(z^{\alpha})$ be the Kullback–Leibler backward projection of z^{α} onto A_1. Then $v^1 = \Gamma x^1$ and x^1 is given by the following alternative. Either $\sigma_{i,2k-1}/\tau_{i,2k-1} \leq \sigma_{i,2k}/\tau_{i,2k}$, in which case

$$x_{i,2k-1}^1 = x_{i,2k}^1 = \frac{\sigma_{i,2k-1} + \sigma_{i,2k}}{\tau_{i,2k-1} + \tau_{i,2k}}$$

or $\sigma_{i,2k-1}/\tau_{i,2k-1} > \sigma_{i,2k}/\tau_{i,2k}$, in which case we have

$$x_{i,2k-1}^1 = \sigma_{i,2k-1}/\tau_{i,2k-1}, \quad \text{and} \quad x_{i,2k}^1 = \sigma_{i,2k}/\tau_{i,2k}.$$

A similar formula is obtained for the backward projection $v^2 = P_{A_2}^{\leftarrow}(z^{\alpha})$, $v^2 = \Gamma x^2$.

X. CONCLUSION

We have presented two versions of an EM algorithm for dynamic SPECT where the recorded events y_{jk} are assumed to be either Poisson or normally distributed. The EM algorithm is recognized as an alternating projection scheme in the sense of von Neumann for normal laws, and with respect to the Kullback–Leibler distance in the Poisson case. Our simulations indicate that a parametric model, based on experiments in myocardial viability studies (2) used in previous experiments [20], may be replaced by a less biased model (4) which, in addition, seems to have computational advantages. Modified versions of our approach (in the spirit of [28]), accounting for measurement noise in real clinical data, could easily be formulated. Related regularization techniques are discussed in [25] and [26]. Finally, we mention that our approach may be extended to a Baysian model, including prior information about the expected dynamic reconstructions.

APPENDIX

Let us briefly outline von Neumann's idea of alternating projections, originally formulated for hyperplanes, but later extended to more general convex sets.

Suppose we are given two closed convex sets A and B and we are interested to find a point in the intersection $C := A \cap B$, in short, a solution. Denote the orthogonal projections (also called nearest point mappings) onto set A, B by P_A, P_B and fix a starting point x_0. The method of alternating projections generates a sequence

$$x_0 \mapsto x_1 := P_B x_0 \mapsto x_2 := P_A x_1 \mapsto x_3 := P_B x_2 \mapsto \cdots.$$

If $C \neq \emptyset$, then the sequence (x_n) converges to some solution in C. Otherwise, it diverges. In the latter case, either the distance between the sets is attained, then the subsequences $(x_{2n}), (x_{2n+1})$ converge to points realizing this distance, that is, $x_{2n} \to a \in A$, $x_{2n+1} \to b \in B$ and $\|a - b\| = \text{dist}(A, B)$, or there are no points realizing the distance between A and B, in which case $\|x_n\| \to +\infty$.

The von Neumann scheme has often served as a numerical tool to calculate solutions to linear or nonlinear systems, in particular, if extended to the case of more than two sets (see [4] for more details and further references).

As the present paper highlights, the Gaussian EM algorithm is a particular case of von Neumann's scheme where the projection P_B onto B, the E step, has an explicit formula (8), while the projection P_A onto A, the M step, is more involved and usually requires solving a number of optimization problems. In the static case, the M step P_A is equally convenient since it admits an explicit formula. Notice that in the case of the EM algorithm, the intersection $A \cap B$ is typically empty and the first part of the above dichotomy occurs, i.e., the distance between A and B is attained.

In the case of Poisson distributed events, the same scheme is valid if the Euclidean distance is replaced with the Kullback–Leibler distance (see Section IV and also [16] and [11]). For readers familiar with the Shepp–Vardi version of the EM algorithm for static emission tomography, we mention that in their case both projections $P_B^\leftarrow$, the E step, and $P_A^\leftarrow$, the M step, are equally convenient to perform since they lead to explicit formulas. It is then possible to calculate the product $P_A^\leftarrow \circ P_B^\leftarrow$ and it is the iterative scheme based on this operator which, in medical imaging, is widely known as *the* EM algorithm.

REFERENCES

[1] C. Bai, G. L. Zeng, G. T. Gullberg, F. DiFilippo, and S. Miller, "Slab-by-slab blurring model for geometric point response correction and attenuation correction using iterative reconstruction algorithms," in *Proc. IEEE Medical Imaging Conf.*, Albuquerque, NM, Nov. 1997.

[2] H. H. Bauschke and J. M. Borwein, "Convex functions of Legendre type and the method of cyclic Bregman projections onto convex sets in Euclidean space," *J. Convex Anal.*, vol. 4, no. 1, pp. 27–67, 1997.

[3] H. H. Bauschke and J. M. Borwein, "On the convergence of von Neumann's alternating projection algorithm for two sets," *Set-Valued Analy.*, vol. 1, no. 2, pp. 185–212, 1993.

[4] H. H. Bauschke, J. M. Borwein, and A. S. Lewis, "Convex sets and the cyclic projection algorithm," in *Recent Developments in Optimization Theory and Nonlinear Analysis*, Y. Censor and S. Reich, Eds. Jerusalem, Israel, 1995.

[5] J. M. Borwein and W. Sun, "The stability analysis of dynamic SPECT systems," *Numerische Mathematik*, vol. 77, no. 3, pp. 283–298, 1997.

[6] A. V. Bronnikov, "Degradation transform in tomography," *Pattern Recognit. Lett.*, vol. 15, pp. 527–532, 1994.

[7] C. L. Byrne, "Iterative image reconstruction algorithms based on cross-entropy minimization," *IEEE Trans. Image Processing*, vol. 2, pp. 96–103, 1993.

[8] R. E. Carson and K. Lange, "The EM parametric image reconstruction algorithm," *J. Amer. Statist. Assoc.*, vol. 80, no. 389, pp. 20–22, 1985.

[9] A. Celler, A. Sitek, and R. Harrop, "Reconstruction of multiple line source attenuation maps," *IEEE Trans. Nucl. Sci.*, vol. 44, pp. 1503–1506, 1997.

[10] F. Clarke, "Optimization and nonsmooth analysis," *SIAM Classics Appl. Math.*, vol. 5, 1990.

[11] I. Csiszár and G. Tusnády, "Information geometry and alternating minimization procedures," *Statist. Decisions*, vol. 1, pp. 205–237, 1984.

[12] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist Soc. Ser. B*, vol. 39, pp. 1–38, 1977.

[13] V. Dicken, "Simultaneous activity and attenuation reconstruction in SPECT, a nonlinear ill-posed problem," Ph.D. thesis, Univ. Potsdam, 1997.

[14] M. P. J. Hudon, D. M. Lyster, W. R. E. Jamieson, A. K. Qayumi, C. Sartori, and H. A. Dougan, "The metabolism of (^{123}I)-iodophenyl pentadecanoic acid in a surgically induced canine model of regional ischemia," *European J. Nucl. Med.*, vol. 16, pp. 199–204, 1990.

[15] H. Iida, H. Itoh, M. Nakazawa, J. Hatazawa, H. Nishimura, Y. Onish, and K. Uemura, "Quantitative mapping of regional cerebral blood flow using Iodine-123-IMP and SPECT," *J. Nucl. Med.*, vol. 35, pp. 2019–2030, 1994.

[16] A. N. Iusem, "A short convergence proof of the EM algorithm for a specific Poisson model," *Revista Brasileira de Probabilidade e Estatística*, vol. 6, pp. 57–67, 1992.

[17] A. N. Iusem, "Convergence analysis for a multiplicatively relaxed EM algorithm," *Math. Methods Appl. Sci.*, vol. 14, pp. 573–593, 1991.

[18] M. A. King, B. M. W. Tsui, and T. P. Pan, "Attenuation compensation for cardiac single photon emission computed tomographic imaging: Part 1. Impact of attenuation and methods of estimating attenuation maps," *J. Nucl. Cardiol.*, vol. 2 pp. 513–524, 1995.

[19] M. King, B. Tsui, T. P. Pan, S. J. Glick, and E. J. Soares, "Attenuation compensation for cardiac single-photon emission computed tomographic imaging: Part 2. Attenuation compensation algorithms," *J. Nucl. Cardiol.*, vol. 3, pp. 55–64, 1996.

[20] M. N. Limber, A. Celler, J. S. Barney, M. A. Limber, and J. Borwein, "Direct reconstruction of functional parameters for dynamic," *IEEE Trans. Nucl. Sci.*, vol. 42, pp. 1249–1256, 1995.

[21] Maeght, J. D. Noll, A. Celler, and T. Farncombe, "Methods for dynamic emission tomography, Laboratoire Mathématiques pour l'Industrie et le Physique (MIP)," Tech. Rep. LAO 97-08.

[22] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Mag.*, vol. 13, pp. 47–60, June 1996.

[23] K. Nakajima, J. Taki, H. Bunko, M. Matsudaira, A. Muramori, I. Matsunari, K. Hisadaa, and T. Ichihara, "Dynamic acquisition with tree-headed SPECT system: Application to technetium 99m-SQ30217 myocardial imaging," *J. Nucl. Med.*, vol. 32, pp. 1273–1277, 1991.

[24] F. Natterer, "Determination of tissue attenuation in emission tomography of optically dense media," *Inverse Problems*, vol. 9, pp. 731–736, 1993.

[25] D. Noll, "Restoration of degraded images with maximum entropy," *J. Global Optimization*, vol. 10, pp. 91–103, 1997.

[26] D. Noll, "Reconstruction with noisy data—An approach via eigenvalue optimization," *SIAM J. Optimization*, vol. 8, no. 1, pp. 82–104, 1998.

[27] A. R. De Pierro, "On the relation between ISRA and the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 12, pp. 328–333, 1993.

[28] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imag.*, vol. 14, pp. 132–137, 1995.

[29] G. A. F. Seber, *Multivariate Observations* (Wiley Series Probability and Mathematical Statistics). New York: Wiley, 1984.

[30] A. M. Smith, G. T. Gullberg, P. E. Christian, and F. L. Datz, "Kinetic modeling of teboroxime using dynamic SPECT imaging of a canine model," *J. Nucl. Med.*, vol. 35, pp. 484–495, 1994.

[31] P. Tan, D. Bailey, S. Meikle, S. Eberl, R. Fulton, and B. Hutton, "A scanning line source for simultaneous emission and transmission measurements in SPECT," *J. Nucl. Med.*, vol. 34, pp. 1752–1760, 1993.

[32] C. H. Tung, G. T. Gullberg, G. L. Zeng, P. E. Christian, F. L. Datz, and H. T. Morgan, "Non-uniform attenuation correction using simultaneous transmission and emission converging tomography," *IEEE Trans. Nucl. Sci.*, vol. 339, pp. 1134–1143, 1992.

[33] L. Van Elmbt and S. Walrand, "Simultaneous correction of attenuation and distance-dependent resolution in SPECT: An analytical approach," *Phys. Med. Biol.*, vol. 38, pp. 1107–1217, 1993.

[34] Y. Vardi, L. A. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *J. Amer. Statist. Assoc.*, vol. 80, no. 389, pp. 8–20, 1985.

[35] A. Welch, R. Clack, F. Natterer, and G. T. Gullberg, "Toward accurate attenuation correction in SPECT without transmission measurements," *IEEE Trans. Med. Imag.*, vol. 16, no. 5, pp. 532–541, 1997.

[36] R. G. Wells, A. Celler, and R. Harrop, "Experimental validation of an analytical method of calculation of photon distributions," *IEEE Trans. Nucl. Sci.*, vol. 44, pp. 1283–1290, 1997.

[37] K. P. Whittall and A. L. MacKay, "Quantitative interpretation of NMR relaxation data," *J. Magn. Resonance*, vol. 84, pp. 134–152, 1989.

[38] C. F. J. Wu, "On the convergence properties of the EM-algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, 1983.