An EM-algorithm for dynamic SPECT tomography

Heinz H. Bauschke\textsuperscript{1}, Dominikus Noll\textsuperscript{2}, Anna Celler\textsuperscript{3} and Jonathan M. Borwein\textsuperscript{4}

1 Introduction

Single Photon Emission Computed Tomography (SPECT) is a nuclear medicine diagnosis technique which measures the three dimensional distribution of a radioactively labeled pharmaceutical injected in the body. As compared to standard imaging techniques like 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), [19].

A SPECT camera works by rotating around the patient an array of photo multipliers that detect gamma rays emitted by the patient. A collimator placed in front of the camera rejects rays that are not perpendicular to the camera face (see Figure 1). The images resulting in the camera are 2D projections of the original 3D 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 not more than 20 minutes). However, physiological processes in the body are usually dynamic, and some organs (kidney, heart) show a significant decay of activity due to wash-out. Being able to trace activity in space and time is therefore of importance and expected to significantly enhance diagnostic possibilities. In particular, a dynamic SPECT reconstruction might be presented as a movie rather than a static image.

With a significant decay of activity over the acquisition period, the filtered back projection method (FBP) may no longer be reliably used, and new approaches have to be

\textsuperscript{1}University of Waterloo, Department of Combinatorics and Optimization, N2L 3G1 Waterloo, Ontario, Canada

\textsuperscript{2}Université Paul Sabatier, Laboratoire Approximation et Optimisation, 118, route de Narbonne, 31062 Toulouse, France

\textsuperscript{3}Vancouver Hospital and Health Centre, Division of Nuclear Medicine, Vancouver, BC, Canada, V5Z 1M9

\textsuperscript{4}Center for Experimental and Constructive Mathematics, Simon Fraser University, Burnaby, BC, Canada, V5A 1S6
developed. The purpose of the present paper is to present 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 [13, 9].

2 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 $180^\circ$, with $S$ stops, $60 \leq S \leq 180$, say, each stop taking $20/S$ minutes. The camera plane is an array of photo multipliers of typical size $30\text{cm} \times 40\text{cm}$. Depending on its resolution, the camera is divided into $M$ receptor bins, a typical choice being for instance some $6\text{mm} \times 6\text{mm}$, in which case we have $M \approx 250$ receptor elements. The camera rotates about a fixed axis, with a radius of $20\text{cm} - 30\text{cm}$.

![Diagram](image)

**Figure 1.** Photons radiating from the region of interest. a) misses the camera, b) is absorbed by the collimator, c) passes the collimator and hits the camera.

The gamma rays originate from a region of interest of approximately $30 \times 30 \times 30\text{cm}^3$. This region of interest is divided into $N$ little cubes called voxels. In our example, assuming a
spatial resolution of .5cm - 1cm leads to $N = 30^3$ up to $N = 60^3$ elements. During the $k$th stop the camera is at a fixed position with angle $\theta_k = (k - 1) \cdot \frac{180}{S}$, $k = 1, \ldots, S$. 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 $j = 1, \ldots, M$. While the camera head is at position $\theta_k$, only gammas traveling along a line with angle $\theta_k$ are allowed to pass the collimator and hit some photo cell (see Figure 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 period of 20 minutes. Here activity $x_{ik}$ of voxel $i$ during the $k$th stop is proportional to the number of photons that leave the voxel during this time interval, radiating in any possible direction.

**Figure 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 $\theta_k$.

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 $\theta_k$ 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 besides geometry, they may as well include probabilistic factors like gamma ray scattering, and more importantly, attenuation (see [6, 20] for details on those). In a simplified model, where attenuation is 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 $\theta_k$, see Figure 2. Based on the geometry, the expected values of the $Y_{jk}$ are

$$E(Y_{jk}) = \sum_{i=1}^{N} c_{ijk} x_{ik} > 0.$$ (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 $(C x)_j = \sum_{i=1}^{N} c_{ijk} x_{ik}$, and $\Gamma : \mathbb{R}^{NS} \to \mathbb{R}^{NMS}$ by $(\Gamma x)_{ijk} = c_{ijk} x_{ik}$. We observe that in practice $\Gamma$ is usually injective, while $C$ is typically not.

3 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 [10], the following parametric model of decay

$$x_i(t) = A_i e^{-\lambda_i t} + B_i e^{-\mu_i t} + C_i$$ (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 [13, 3]. Discretizing (2) according to the $S$ stops, $t_k = (k - 1)S/20$, leads to a parametric model with 5 degrees of freedom for each $i$. Another approach suggested in the investigation of NMR relaxation data (cf. [21]) uses a model of the form

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

$$x_{i1} \geq x_{i2} \geq x_{i3} \geq \ldots \geq x_{iS} \quad \text{and} \quad x_{iS} \geq 0.$$ (4)

As compared to (2), this increases the flexibility of modeling decay, but increasing the degrees of freedom bares the risk of producing an underfitted model. Fortunately, experiments indicate (see section 8) that the model works well, and that its obvious 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 $\Omega$. 

4
4 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

\[
(ML_1) \quad \text{maximize} \quad E(\log g(y; x)) \\
\text{subject to} \quad x \in \Omega
\]

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

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

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 \( \Omega \). On the other hand, solutions are typically not unique. Any solution \( x \in \Omega \) must satisfy the following Kuhn-Tucker conditions (cf. [7]): There exist multipliers \( \lambda_{ik} \geq 0 \), \( i = 1, \ldots, N \) and \( k = 1, \ldots, S \) such that for each \( i = 1, \ldots, N \),

\[
\begin{align*}
\tau_{i1} & - \sum_{j=1}^{M} \left( c_{ij1} y_{j1} / \sum_{i'\neq 1}^{N} c_{ij1} x_{i'1} \right) - \lambda_{i1} = 0 \\
\tau_{i2} & - \sum_{j=1}^{M} \left( c_{ij2} y_{j2} / \sum_{i'=1}^{N} c_{ij2} x_{i'i2} \right) + \lambda_{i1} - \lambda_{i2} = 0 \\
(KT_1) & \quad \text{...} \quad \text{...} \quad \text{...} \\
\tau_{iS-1} & - \sum_{j=1}^{M} \left( c_{ijS-1} y_{jS-1} / \sum_{i'=1}^{N} c_{ijS-1} x_{i'S-1} \right) + \lambda_{iS-1} - \lambda_{iS-1} = 0 \\
\tau_{iS} & - \sum_{j=1}^{M} \left( c_{ijS} y_{jS} / \sum_{i'=1}^{N} c_{ijS} x_{i'S} \right) + \lambda_{iS-1} - \lambda_{iS} = 0 \\
x_{ik} & \geq x_{ik+1} \quad \lambda_{ik} (x_{ik} - x_{ik+1}) = 0 \quad k = 1, \ldots, S - 1, \quad \lambda_{iS} x_{iS} = 0
\end{align*}
\]

where we used the abbreviation

\[
\tau_{ik} := \sum_{j=1}^{M} c_{ijk} > 0.
\]

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 \((C x)_{jk} = \sum_{i=1}^{N} c_{ijk} x_{ik}\) is typically not injective (\(ML_1\)) may have multiple solutions. (Notice that ensuring that \( C \) is only moderately defective is of relevance for the numerics and may be influenced by the discretization we choose; cf. [3]).
5 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, Laird and Rubin [8], 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 $\mathcal{ML}_1$ 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 towards 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 $x^\alpha = \{x^\alpha_{ik}\} \in \Omega, \alpha = 1, 2, \ldots$ according to the following rules:

1. **E-step**: Given $x^\alpha = \{x^\alpha_{ik}\} \in \Omega$, 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 $x^\alpha$, we find that

$$z^\alpha_{ijk} = y_{jk} \frac{c_{ijk}x^\alpha_{ik}}{\sum_{i' = 1}^N c_{i'jk}x^\alpha_{i'k}}. \quad (6)$$

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

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

is obtained as

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

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

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

6
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 \( x^i = (x_{i1}, \ldots, x_{iS}) \in \Omega_i \), where \( \Omega_i \) are the \( x^i \) satisfying (4) for fixed \( i \). Notice that the data \( y \) and the previous iterate \( x^\alpha \) do not enter directly into the calculation of \( x^{\alpha+1} \). Rather it is \( z^\alpha \) 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 \( \tau_{ik} \) 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 \( S \):
\[
(ML_2) \quad \text{maximize} \quad E(\log f(y; x) | y, x^\alpha) \\
\text{subject to} \quad x \in \Omega
\]

Setting \( \psi_{ik}(x^i) := -x_{ik} + x_{i,k+1} \), and with \( \psi_i \) defined as before, the \( i \)th problem becomes:
\[
(ML_{2,i}) \quad \text{minimize} \quad \psi_i(x^i) \\
\text{subject to} \quad \psi_{ik}(x^i) \leq 0 (k = 1, \ldots, S - 1)
\]

Notice that in contrast with the original maximum likelihood problem \((ML_1)\), we do not control the constraint \( x_{iS} \geq 0 \) here, since it is built in through the objective due to \( \sigma_{ik}^{\alpha} > 0 \). 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 \( \psi_i \) are strictly convex and coercive.

The Kuhn-Tucker conditions (cf. [7]) for the optimal solutions \( x^{i,\alpha+1} \) of \((ML_{2,i})\) imply the existence of multipliers \( \lambda^\alpha_{ik} \geq 0 \), \( k = 1, \ldots, S - 1 \) such that:
\[
\begin{align*}
\tau_1 &= -\frac{\sigma_{i1}^{\alpha}}{x_{i1}^{\alpha+1}} - \lambda^\alpha_{i1} = 0 \\
\tau_2 &= -\frac{\sigma_{i2}^{\alpha}}{x_{i2}^{\alpha+1}} + \lambda^\alpha_{i1} - \lambda^\alpha_{i2} = 0 \\
&\vdots \\
\tau_{iS-1} &= -\frac{\sigma_{i,S-1}^{\alpha}}{x_{i,S-1}^{\alpha+1}} + \lambda^\alpha_{i,S-2} - \lambda^\alpha_{i,S-1} = 0 \\
\tau_{iS} &= -\frac{\sigma_{iS}^{\alpha}}{x_{iS}^{\alpha+1}} + \lambda^\alpha_{i,S-1} = 0 \\
x_{ik}^{\alpha+1} &\geq x_{ik}^{\alpha+1}_{ik+1} + \lambda^\alpha_{ik}(x_{ik}^{\alpha+1} - x_{ik+1}^{\alpha+1}) = 0 \quad k = 1, 2, \ldots, S - 1.
\end{align*}
\]
Based on the optimality conditions ($KT_{2,i}$), we may now enter a detailed analysis of the present EM algorithm.

6 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.

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 $P_A^-$ and $P^-_A$ associated with $d$. Given a closed convex subset $A$ of $\mathbb{R}_+^p$ and a point $x \in \mathbb{R}_+^p$, the forward resp. backward projections of $x$ onto $A$ are defined as

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

and

$$P^-_A(x) = \arg\min_{a \in A} d(x, a).$$ 

With these notions, the M-step presented in Section 5 turns out to be a Kullback-Leibler backward projection.

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

$$v^{\alpha+1} = P^-_A(z^\alpha).$$ 

Then the solution $x^{\alpha+1}$ of the M-step satisfies $v^{\alpha+1}_{ijk} = c_{ijk} x^{\alpha+1}_{ik}$, 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 satisfies $\psi(x) = \sum_{i,j,k} \left(v_{ijk} - z^\alpha_{ijk} \log v_{ijk}\right)$, 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 here 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 $\Omega$ under $\Gamma$, and is therefore itself polyhedral. 

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

$$z^\alpha = P_B^-(v^\alpha) = P_B^+(v^\alpha).$$
Proof. First consider the case of the backward projection. Clearly \( z = P_{\mathcal{B}}^{\alpha}(v^\alpha) \) exists and satisfies the Kuhn-Tucker conditions: There exist multipliers \( \lambda_{jk} \) such that

\[
- \frac{v_{ijk}^\alpha}{\hat{z}_{ijk}} + 1 + \lambda_{jk} = 0 \quad \forall i, j, k
\]

in tandem with the constraints

\[
\sum_{i=1}^{N} \hat{z}_{ijk} = y_{jk} \quad \forall j, k.
\]

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

Next consider the case of the forward projection \( z = P_{\mathcal{A}}^{\alpha}(v^\alpha) \). Here the Kuhn-Tucker conditions provide multipliers \( \lambda_{jk} \) satisfying

\[
\log\left( \frac{z_{ijk}}{v_{ijk}^\alpha} \right) + \lambda_{jk} = 0 \quad \forall i, j, k
\]

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 formula (6).

\( \square \)

It remains to observe that any limit point of the sequence \( x^\alpha \) must be a Kuhn-Tucker point for the problem \((ML_1)\). This is a consequence of a general fact (see [22]), but could equally well be checked using the conditions \((KT_1)\) in tandem with (6), (7) and the \((KT_{2,i})\). Indeed, convergence of the iterates \( v^\alpha, z^\alpha \) implies convergence of \( x^\alpha, \sigma^\alpha \). Formula \((KKT_1)\) implies

\[
\lambda_{ij}^\alpha = \sum_{j=1}^{k} (\tau_{ij} - \sigma_{ij}^\alpha/x_{ij}^\alpha),
\]

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 \( x_{iS} \geq 0 \). But clearly, 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 \( z^\alpha \) 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 \( \Omega \), admits optimal solutions. We claim that the sequences of alternating Kullback-Leibler projections

\[
\begin{align*}
    z^\alpha &= P_{\mathcal{B}}^{\alpha}(v^\alpha) = B_{\mathcal{B}}(v^\alpha) \\
v^{\alpha+1} &= P_{\mathcal{A}}^{\alpha}(z^\alpha)
\end{align*}
\]
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. [1, Thm. 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 [11, 12, 19, 4], 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 \( \Gamma \), and by closedness lies in \( \Omega \). 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 \( \mathbb{R}_+^n \), while \( A = \Gamma(\Omega) \) depends on the parameter space \( \Omega \), and may therefore be quite complicated. The whole procedure may indeed be generalized to any comparable situation with a closed convex parameter space \( \Omega \).

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

**Remark.** The case considered by Iusem [11] and others becomes the case where we impose only positivity conditions, \( \hat{\Omega} = \{ x : x_{ik} \geq 0 \} \), \( \hat{A} = \Gamma(\hat{\Omega}) \). In this case the backward projection onto \( \hat{A} \) is particularly pleasant to calculate and leads to an explicit formula.

### 7 The Least Squares Approach

In [9], 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:

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

Here the operator \( C : \mathbb{R}^n \to \mathbb{R}^n \) 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. [13, 3]). Following the ideas presented in Section 3, we replace the parametric model (2) by (4). This leads to a linear least squares problem with inequality constraints

\[
\begin{align*}
\text{(L1S)} \quad &\text{minimize} \quad \|Cx - y\| \\
&\text{subject to} \quad x \in \Omega
\end{align*}
\]
where $\Omega$ 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 5, 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 $c_{ijk}x_{jk}$ 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

$$
(ML_1) \quad \text{maximize} \quad E(\log g(y; x))
$$

subject to $x \in \Omega$

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

Returning to the outline of Section 5, we apply the EM-algorithm to the present situation, letting $Z$ resp. $Y$ represent the complete resp. incomplete data spaces. As in the Poisson case, this generates sequences $x^\alpha$ and $z^\alpha$, $\alpha = 1, 2, \ldots$ according to the following rules:

1. **E-step.** Given the current iterate $x^\alpha \in \Omega$, calculate the conditional expectation $z^\alpha = E(Z|Y = y; x^\alpha)$.

2. **M-step.** Calculate the new iterate $x^{\alpha+1} \in \Omega$ by maximizing $E(\log f(z^\alpha; x)|y; x^\alpha)$ over $x \in \Omega$.

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

**Proposition 3.** Let $A = \{v \in \mathbb{R}^p : v = \Gamma x$ for some $x \in \Omega\}$. Then, given the result $z^\alpha$ of the previous E-step, the next M-step reduces to calculating the orthogonal projection $v^{\alpha+1}$ of $z^\alpha$ 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 $E(Z|\sum_{i=1}^{N} Z_{ijk} = y_{jk}; x^\alpha)$. 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 $x = E(X|X_1 + \ldots + X_n = y)$ satisfies

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

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

11
Proof. In consequence of [18, Thm. 2.1(viii)], a normal vector

\[
\left( \begin{array}{c}
 a \\
 b
\end{array} \right) \sim \mathcal{N}\left( \left( \begin{array}{c}
 a \\
 \beta
\end{array} \right), \left( \begin{array}{cc}
 U^* & W \\
 W & V
\end{array} \right) \right)
\]

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 \( b = (X_1, \ldots, X_n), a = X_1 + \ldots + X_n \), which gives the stated formula.

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

\[
z_{ijk}^\alpha = \frac{1}{N}y_{jk} + c_{ijk}x_{ik}^\alpha - \frac{1}{N} \sum_{i'=1}^N c_{i'jk}x_{i'k}^\alpha,
\]

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 \( x^\alpha \) be the current iterate generated by the previous M-step, and let \( \nu^\alpha = \Gamma x^\alpha \). Then the result \( z^\alpha \) of the next E-step is the orthogonal projection of \( \nu^\alpha \) 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 \( z^\alpha, \nu^\alpha = \Gamma x^\alpha \) be the sequences generated by the Gaussian EM-algorithm. Then \( z^\alpha, \nu^\alpha \) converge to certain limit points \( \nu^\alpha \rightarrow \bar{\nu} \in A, z^\alpha \rightarrow \bar{z} \in B \). Here \( \bar{\nu} \) is of the form \( \bar{\nu} = \Gamma \bar{x} \), with \( \bar{x} \in \Omega \) a solution to the least squares problem \( (L S) \).

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 \( \bar{\nu} = \bar{z} \in A \cap B \), which by definition of \( A \) is then of the form \( \Gamma \bar{x} \).

The more involved case occurs when \( A \cap B = \emptyset \). Following [2], a dichotomy appears: Either the alternating sequences converge to limit points \( \nu^\alpha \rightarrow \bar{\nu} \in A, z^\alpha \rightarrow \bar{z} \in B \), with \( \bar{\nu}, \bar{z} \) realizing the distance between \( A, B \), or there are no points realizing this distance, in which case the sequences tend to infinity, \( \|\nu^\alpha\| \rightarrow \infty, \|z^\alpha\| \rightarrow \infty \), with \( \|\nu^\alpha - z^\alpha\| \) 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 \( \bar{a} \in A, \bar{b} \in B \) having \( \bar{a} + \mathbb{R}_+w \subset A, \bar{b} + \mathbb{R}_+w \subset B \). 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 \( \nu^\alpha \rightarrow \bar{\nu} = \Gamma \bar{x} \) and \( z^\alpha \rightarrow \bar{z} \). It is now routine to check that \( \bar{x} \) is a Kuhn-Tucker point for the original problem \( (L S) \).
Remark. Replacing $\Omega$ by $\hat{\Omega} = \{ x : x_{ik} \geq 0 \}$ again gives an important special case. The orthogonal projection onto $\hat{A} = \Gamma(\hat{\Omega})$ may be calculated explicitly, and in tandem with (8) leads to an explicit formula for updating $x^\alpha$. 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 like ISRA (see [16]).

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

$$\psi(x) = \frac{1}{2} \sum_{i,j,k} (z_{ijk}^\alpha - c_{ijk} x_{ik})^2,$$

and is to be minimized over $\Omega$. Setting

$$\gamma_{ik} := \sum_{j=1}^{M} c_{ijk}^2 > 0 \quad \text{and} \quad \delta_{ik} := \sum_{j=1}^{M} c_{ijk} z_{ijk}^\alpha > 0,$$

we have

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

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

$$
\begin{align*}
\gamma_{i1} x_{i1} &- \delta_{i1} - \lambda_{i1} = 0 \\
\gamma_{i2} x_{i2} &- \delta_{i2} + \lambda_{i1} - \lambda_{i2} = 0 \\
\vdots & \quad \quad \ldots \\
\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{align*}

(LKT$_i$)

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

So 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 $\lambda_{iS}$. However, as we shall see, these constraints will generally be inactive, and the multipliers will vanish.

8 Our 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 using an 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 whether the M-step, which is intrinsically more complicated than for the stationary case
([11, 4]), allows for a fast numerical solution. Let us therefore analyze the M-steps of both the Poisson and the Gaussian models.

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 5. To simplify notation we omit the indices i and α which are fixed, writing τ_k := τ_{ik} and σ_k := σ_{ik}, and similarly x_k := x_{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

\[ x_1 = \ldots = x_{r_1} > x_{r_1+1} = \ldots = x_{r_2} > \ldots > x_{r_{t-1}+1} = \ldots = x_{r_t}, \quad (10) \]

for appropriate 0 = r_0 < r_1 < \ldots < r_t = S, with

\[ x_{r_{j-1}+1} = \ldots = x_{r_j} =: x^{(j)} = \frac{\sigma_{r_{j-1}+1} + \ldots + \sigma_{r_j}}{\tau_{r_{j-1}+1} + \ldots + \tau_{r_j}}, \quad (11) \]

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 formula (11). \( \square \)

**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 the sequence \( x^{(j)}, j = 1, \ldots, t \) 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 \( \sigma_i \) 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 7. Here we have the same observation, which, again on suppressing the indices 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 \( \delta_i \) replacing \( \sigma_i \), and \( \gamma_i \) replacing \( \tau_i \). 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 \( \lambda_j \) must vanish. Again summing the Kuhn-Tucker conditions (LKT_i) in each block would suffice, provided we knew that the final multiplier \( \lambda_S \) belonging to the constraint \( x_S \geq 0 \) 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: \( x^{(t)} = x_{r_i+1} = \ldots = x_S = 0 \). Then the Kuhn-Tucker conditions give

\[ \lambda_S = - \sum_{k=r_i+1}^{S} \delta_k, \]

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

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 \((ML_{2,i})\). 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 towards minimizing \( \phi_{\mu} \) for a fixed \( \mu > 0 \), before passing to another E-step. Increasing the number of steps and reducing \( \mu \) should then be controlled.
by testing decrease of the likelihood function in \((ML_1)\) resp. \((LLS)\), which for theoretical reasons (see \([8]\)) is known to decrease at each iteration if an exact EM-step is performed.

**Experiment.** Figure 3 shows typical results of the Gaussian M-step for randomly generated data on the basis of 60 (left column) resp. 80 (right column) camera positions. The plots in the top line show comparisons of typical decay curves produced by the models (2) (smooth curves) versus (4) (steplike curves).

The lower pictures indicate 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 steplike functions (4). Typically, these show little difference from the direct fitting of (2) (smooth broken curves as before), so (2), if desired, could be retrieved from (4). As (4) has obvious 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.

## 9 Cyclic Projections

The fact that the Kullback-Leibler resp. orthogonal projections onto the set \(A\) cannot be calculated explicitly is clearly a drawback of our present approach. This may be overcome by splitting the alternating projections based on the EM-algorithm into three or even more projection steps which are easier and may be performed in a cyclic ordering. To be more precise 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 show that projecting cyclically onto the sets is not exactly what we want. The are, however, various possible ways in which the projected information onto the three sets may be used to approach the points \(a \in A, b \in B\) 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 \(x^\alpha\) resp. \(v^\alpha = \Gamma x^\alpha\), do an E-step using (6) resp. (8). The result is \(x^\alpha\).

2. Replace the M-step by the following: Project \(z^\alpha\) onto \(A_i\), which gives \(v^i = P_{A_i}(z^\alpha), i = 1, 2\). 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 \(v^\alpha\) 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 \( v^1 = P_{A_1}^{-}(z^n) \) be the Kullback-Leibler backward projection of \( z^n \) 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

\[
x^1_{i,2k-1} = x^1_{i,2k} = \frac{\sigma_{i,2k-1} + \sigma_{i,2k}}{\tau_{i,2k-1} + \tau_{i,2k}} \]

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

\[
x^1_{i,2k-1} = \sigma_{i,2k-1}/\tau_{i,2k-1}, \quad \text{and} \quad x^1_{i,2k} = \sigma_{i,2k}/\tau_{i,2k}.
\]

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

### 10 Conclusion

We have presented two versions of an EM algorithm for dynamic SPECT based on a Poisson resp. a normal distribution. The EM algorithm is recognized as an alternating projection scheme either in the sense of von Neumann, or with respect to the Kullback-Leibler distance. Our simulations indicate that a parametric model based on experiments in myocardial viability studies (2) used in previous experiments [9] should be replaced by a less biased model (4), which in addition has computational advantages. Modified versions of our approach (in the spirit of [17]) accounting for measurement noise in real clinical data could easily be formulated. Related regularization techniques are discussed in [14, 15].

### References


