EM FILTER FOR TIME-VARYING SPECT RECONSTRUCTION

Joe Qranfal[1][§], Charles Byrne[2]

[1]Department of Mathematics
Simon Fraser University
British Columbia, CANADA
[2]Department of Mathematical Sciences
University of Massachusetts at Lowell
Lowellm, USA

Abstract: The new filtering algorithm EM (expectation maximization) filter is introduced and is validated numerically by applying it to solve the ill-posed inverse problem of reconstructing a time-varying medical image. A linear state-space stochastic approach based on a Markov process is utilized to model the problem, while no precise a-priori information about the underlying dynamics of the physical process is required. The method is tested for the case of image reconstruction from noisy data in dynamic single photon emission computed tomography (SPECT), where the signal strength changes substantially over the time required for the noisy data acquisition. Numerical results corroborate the effectiveness of the reconstruction method.

© 2011 Academic Publications, Ltd.

[§]Correspondence author

1. Introduction

We introduce a new filtering algorithm to find a nonnegative estimate $\hat{x}_k$, $k = 1, \ldots, S$, to the nonnegative unknown x_k of the problem given by the two linear space-state equations,

$$
\begin{aligned}
x_k &= A_k x_{k-1} + \mu_k \\
z_k &= H_k x_k + \nu_k
\end{aligned}
$$

μ_k is the error vector, $\mathbb{E}(\mu_k) = 0$ and $\mathbb{E}(\mu_k \mu_k^\top) = Q_k$ is the covariance of the error in modeling the transition from x_{k-1} to x_k. $\mathbb{E}(\nu_k) = 0$ and $\mathbb{E}(\nu_k \nu_k^\top) = R_k$ are the mean and covariance respectively of the noise vector ν_k. Entries of the vector z_k and of both matrices A_k and H_k are nonnegative. We know also that we deal with white error and noise so that,

$$
\begin{aligned}
Q_k &= \sigma_k^2 I \\
R_k &= diag(z_k)
\end{aligned}
$$

where I is the identity matrix of order N, $diag(a)$ denotes the square matrix that has the a_i in its main diagonal and 0 otherwise. Our new algorithm that we refer as the *SMART filter* is then numerically tested to solve a reconstruction problem arising in medical imaging, namely dynamic/time-varying SPECT.

Recent advances in high speed computing and image processing have contributed significantly to the progress of treatment planning including radiation therapy, hyperthermia, surgical procedure, and cryosurgery. Emission tomography, as in nuclear medicine, is a medical imaging modality which uses detection of electromagnetic radiation for diagnostic purposes. In case of SPECT or PET (positron emission tomography), a judiciously designed chemical radiopharmaceutical tagged with a radioactive isotope is administered to the patient, usually by intravenous injection. It is chosen to amass in a targeted organ or region of the body, the heart or the brain for instance. The radioactive isotope emits photons which are detected by an external device, the gamma camera, at several angular positions around the patient body. Data from these 2D angular views/projections are reconstructed into a 3D image. This reconstructed radionuclide distribution from measured data is a useful tool to clinically interpret and diagnose unhealthy tissue. Nuclear medicine is concerned in investigations of the dynamics of the human body's physiological processes and biochemical function, thus the interest in the dynamic SPECT.

The image reconstruction of the dynamic SPECT is an ill-posed inverse problem. This ill-posedness is further amplified by physical degradation of the

acquired data caused by camera blurring, photon scattering, or attenuation. Most of the reconstruction algorithms presented so far have mainly focused on the case where the activity of the object is time-invariant within the time taken to acquire a full set of independent measurement data. Noniterative approaches, such as FBP (filtered back-projection) approach, and iterative ones, such as EM algorithm, provide acceptable results in the static case when the activity is time independent. However, they will break down when the activity is dynamic.

In the static case, the EM algorithm has been extensively applied to medical image reconstruction since its introduction by Dempster et al [1] in 1977. The first application in emission tomography began with Shepp and Vardi [2] in 1982. Its efficiency is well established; it converges to an optimal solution. A variant, called OSEM (ordered subsets EM), converges faster and is due to Hudson and Larkin [3] in 1994. The EM algorithm and its variants are easy to implement and perform very well to reconstruct a static/time-invariant emission tomography when a nonnegative solution is needed. An extension to a convex constraint set was studied by Bauschke et al [4] for the dynamic/time-varying SPECT problem. The authors describe the expectation step (E step) and the minimization step (M step) as a KL projection onto a convex set and they emphasized that there is no explicit formula for the M step. The M step, indeed, requires solving a nonlinear optimization problem. This does not provide an explicit formula as in the static case. As we shall see, by using a temporal regularization through a cross-entropy minimization, we are able to derive an explicit EM like formula for the dynamic SPECT problem.

The remainder of the paper is organized as follows. First, we describe in Section 2 the stochastic modeling of the state evolution and projection in space. The state evolution and space models are the basis of the HMM (hidden Markov model) to apply our method. Then Section 3 reviews the EM and Kalman filter algorithms. Section 4 introduces the EM filter which is derived through a weighted Kullback-Leibler (KL) distance. The application of EM filter to time-varying SPECT is covered in Section 5. Section 6 details the simulations of dynamic SPECT as numerical experiments; these corroborate the effectiveness of our algorithm in terms of convergence and cpu time. Section 7 concludes the paper in summing up the findings.

2. Problem Formulation

2.1. Stochastic Modeling of Dynamic SPECT

We consider a physiological process where the distribution of the radioactive tracer in an organ or a specific region is time dependent. This region is divided into small parts called dynamic voxels in 3D or doxels and dynamic pixels in 2D or dixels. A SPECT camera, that could have one, two or three heads, is used to register the number of photons emitted by the patient. Let t_k, $k = 1, \ldots, S$ be an index of a sequence of acquisition times, N the total number of voxels and M the total number of bins, we denote by $x_k \in \mathbb{R}^N$ and $z_k \in \mathbb{R}^M$ the spatial distribution of the activity and the measured data during the $k\Delta t$ time. The observations $z_1, z_2 \ldots, z_S$ are random vectors. The ith entry in the vector z_k measures the number of photons registered at the ith bin during the time t_k. The jth entry in the vector x_k measures the number of photons emitted from the jth dixel/voxel during the time t_k. Furthermore, each observation z_k depends on x_k only. The nonnegative activities sequence $x_1, x_2 \ldots, x_S$ satisfy Markov property with unknown time varying transition matrix $A_k \in \mathbb{R}^{N \times N}$. That is

$$x_k = A_k x_{k-1} + \mu_k \tag{1}$$

where, μ_k is the error random vector in modeling the transition from x_{k-1} to x_k with $\mathbb{E}(\mu_k)$ zero and covariance matrix $\mathbb{E}(\mu_k \mu_k^\top) = Q_k$. The random variable μ_k does not have to be a Gaussian process. In many applications the unknown transition matrix is approximated by a random walk or a discrete diffusion-transport operator.

Let $(h_k)_{ij}$ be the conditional probability that an emission from voxel j during the acquisition time $k\Delta t$ will be detected in bin i. We call projection or observation matrix the time varying matrix defined by $H_k = [(h_k)_{ij}]$. It is assumed to be known from the geometry of the detector array and may include attenuation correction. We shall assume throughout that the matrix $H_k \in \mathbb{R}^{M \times N}$ has been constructed so that H_k, as well as any submatrix C obtained from H_k by deleting columns, has full rank. In particular, if C is $M \times L$ and $L \geq M$, then C has rank M. This is not an unrealistic assumption. The columns of H_k are vectors in the nonnegative orthant of M-dimensional space. When attenuation, detector response, and scattering are omitted from the design of the projection matrix H_k, it can sometimes happen that H_k or some submatrices C can fail to be full rank. However, the slightest perturbation of the entries of such a H_k will almost surely produce a new H_k having the

desired full-rank properties. The observation and activity vectors are related by the following

$$z_k = H_k x_k + \nu_k \tag{2}$$

where $\mathbb{E}(\nu_k) = 0$ and $\mathbb{E}(\nu_k \nu_k^\top) = R_k$ are the mean and covariance respectively of the noise vector ν_k. The random variable ν_k does not have to be a Gaussian process or Poisson distributed. We only need to know its mean and covariance matrix. The observation noise is not additive to the measurements in a strict physical sense. However, feasible solutions can also be obtained using this approximate noise model. Multiplicative noise is generally more difficult to remove than additive noise [5], because the intensity of the noise varies with the signal intensity, thus violating the linearity of the observation model. The linear model we choose requires then the noise to be additive. Otherwise we would not have an unbiased estimator, as in the Kalman filter (KF) case for instance.

The operation that concerns itself in the extraction of information about a quantity of interest at time k by using noisy data measured up to and including k is called *filtering*. In our setting, the determination of the activity x_k from the noisy data z_k is a filtering problem. Stochastic filtering is an inverse problem: One needs to find the optimal $\hat{x}_k$, given the data z_k, the evolution matrix A_k and the observation system matrix H_k at each time step k. Equations (1) and (2) are the state-space form of a particular case of a more general filtering problem [6, 7]. The actual model is a linear dynamic system for which the analytic filtering solution is given by the KF [8]. The KF can be viewed as a temporal regularization technique for solving dynamic inverse problems.

We presume that covariance matrices Q_k and R_k are diagonal in the case of time-varying SPECT. We deal then with white error and noise respectively. We also presume that we have a system that has nonnegative entries. Diagonal covariance matrices, white noise and error, and nonnegative system matrices are three more assumptions that we shall adopt throughout this paper. In case one or more of these three is violated, refer to [9, 10] in how we could bring a general setting to an equivalent one presented here in this paper.

3. Filtering and Nonnegative Reconstruction

3.1. The Expectation Maximization Algorithm

The EM algorithm, when the activity is assumed to be static, has been extensively applied to medical image reconstruction since its introduction by Dempster et al [1] in 1977. The first application in emission tomography began with Shepp and Vardi [2] in 1982. Its efficiency is well established; it converges to an optimal solution. A variant, called OSEM (ordered subsets EM), converges faster and is due to Hudson and Larkin [3] in 1994. The EM algorithm and its variants are easy to implement and perform very well to reconstruct a static/time-invariant emission tomography when a nonnegative solution is needed. An extension to a convex constraint set was studied by Bauschke et al [4] for the dynamic/time-varying SPECT problem. The authors describe the expectation step (E step) and the minimization step (M step) as a KL projection onto a convex set and they emphasized that there is no explicit formula for the M step. The M step, indeed, requires solving a nonlinear optimization problem. This does not provide an explicit formula as in the static case. As we shall see, by using a temporal regularization through a cross-entropy minimization, we are able to derive an explicit EM like formula for the dynamic SPECT problem.

The static emission tomography problem amounts to finding $x \in \mathbb{R}^N$ solution of the linear equation

$$z = Hx + \nu \qquad (3)$$

where, $z \in \mathbb{R}^M$, $H \in \mathbb{R}^{M \times N}$ are the observation data vector and the observation system matrix respectively. The vector $\nu \in \mathbb{R}^M$ represents the additive noise in recording the data z. We assume the entries of the vector z are nonnegative, the entries of the matrix H are nonnegative such as each one of its columns sums up to one; that is the system is normalized. We denote by support(x) the set of indexes j of the vector x for which $x_j > 0$. Central to our discussion is the notion of cross-entropy or Kullback-Leibler [11] distance between vectors with nonnegative entries. Recall that the KL distance between nonnegative numbers α and β is

$$KL(\alpha, \beta) = \alpha \log \frac{\alpha}{\beta} + \beta - \alpha$$

We also define $KL(\alpha, 0) = +\infty$, $KL(0, \beta) = \beta$, and $KL(0, 0) = 0$. Extending

to nonnegative vectors $a = (a_1, \cdots, a_N)^\top$ and $b = (b_1, \cdots, b_N)^\top$, we have

$$KL(a,b) = \sum_{j=1}^{N} KL(a_j, b_j) = \sum_{j=1}^{N} (a_j \log \frac{a_j}{b_j} + b_j - a_j)$$

We have $KL(a, b) = \infty$ unless support(a) is contained in support(b). Note how the KL distance is not symmetric that is $KL(a, b) \neq KL(b, a)$.

The EM algorithm is an iterative procedure for computing a nonnegative solution of the linear system (3) when we discard the additive noise ν. Using the cross-entropy/Kullback-Leibler (KL) distance, Titterington [12] noted in 1987 that maximizing the likelihood function is equivalent to minimizing KL(z, Hx) distance. Later on in 1993, Byrne [13] showed that by minimizing KL(Hx, z) we obtain the simultaneous version SMART of MART (multiplicative algebraic reconstruction technique) considered earlier by Gordon et al. [14] and others. Both algorithms are meant for the static state. Starting with an arbitrary $x^0 > 0$, for $\ell = 0, 1, \ldots$, an iteration of EM Algorithm is given by

$$x_j^{\ell+1} = x_j^\ell \sum_{i=1}^{M} \frac{H_{ij} z_i}{(Hx^\ell)_i} \tag{4}$$

When using EM or least squares techniques for dynamic emission tomography, the challenge is how to take into account the activity's dynamics. Some authors assume that the time activity curves (TACs) or the the activity temporal behavior in different regions of interest (ROI) are assumed to be known [15]. In this particular case, constrained least squares method can be effective. However such extra information is not usually available. An alternative approach is to assume nothing about the dynamics of the activity and just use a temporal regularization that would penalize high variations of the activity's changes over time. This has been done in [16] in 2008, where the authors present a Kalman based regularization technique. Recently, Qranfal and Byrne [9] developed a temporal regularization method based on cross-entropy minimization and analyzed a recursive reconstruction method called SMART filter. We introduce here, by the same token, a new method that we refer to as the EM filter. But first, let us review the KF.

3.2. Kalman Filter

Consider the problem of finding an estimator $\hat{x}$ as a linear function of the noisy data vector z that satisfies (3), that is $z = Hx + \nu$. The best linear

unbiased estimator (BLUE) for x from z is the vector $\hat{x}$, which minimizes the cost function

$$J(x) = \|z - Hx\|_{R^{-1}}^2$$

where $\|x\|_B^2$ denotes $x^\top Bx$, the weighted Euclidian norm. If H have full rank then $\hat{x} = V^\top z$, where $V = R^{-1}H(H^\top R^{-1}H)^{-1}$. Now suppose that, in addition to the data vector z, we have $y = x + \mu$. The vector y is a prior estimate of x, where μ is the mean-zero error of this estimate and Q is the known covariance matrix of μ. We want to estimate x as a linear function of both z and y. Applying the BLUE to the augmented system of equations, that is minimizing the cost function

$$F(x) = \|z - Hx\|_{R^{-1}}^2 + \|y - x\|_{Q^{-1}}^2$$

we find the solution to be

$$\hat{x} = y + W(z - Hy)$$

where

$$W = QH^\top (R + HQH^\top)^{-1}$$

We see that to obtain the estimate $\hat{x}$ of x, we first check to see how well y, the prior estimate of x, performs as a potential solution of the system $z = Hx$ and correct the estimate y, using the error $z - Hy$, to get the new estimate $\hat{x}$. If $z = Hy$, then $\hat{x} = y$. The KF involves the repeated application of this extension of the BLUE.

The KF is a recursive algorithm to estimate the state vector x_k during the time $k\Delta t$ as a linear combination of the vectors z_k and y_k. In this case $\hat{x}_k$ is the unique minimizer of the functional

$$F(x_k) = \|z_k - H_k x_k\|_{R_k^{-1}}^2 + \|y_k - x_k\|_{Q_k^{-1}}^2 \tag{5}$$

The KF update is the following: Given an unbiased estimate $\hat{x}_{k-1}$ of the state vector x_{k-1}, our prior estimate of x_k based solely on the activity dynamics is

$$y_k = A_k \hat{x}_{k-1} \tag{6}$$

The estimate $\hat{x}_k$ will have the form, refer for instance to [17],

$$\hat{x}_k = y_k + K_k(z_k - H_k y_k) \tag{7}$$

where

$$P_k = A_k P_{k-1} A_k^\top + Q_k \tag{8}$$

$$K_k \quad = \quad P_k H_k^\top (H_k P_k H_k^\top + R_k)^{-1} \tag{9}$$

P_k and P_{k-1} in (8) are the covariances of the estimated activity $\hat{x}$ at time k and $k-1$ respectively. Since the Kalman estimate is an unconstrained minimizer of $F(x_k)$, we will most likely end up with some negative entries in the vector solution $\hat{x}_k$. This is not a desirable solution in medical imaging. This has been remedied, for instance in [16, 17], where projection into the set of nonnegative vectors was used to cope with this situation. Another drawback of KF is the matrix-matrix multiplications involved in (8) and (9) and matrix inversion involved in (9). Attempts have been made to rectify these two shortcomings; see for instance [6, 7] for more details. Furthermore, KF needs to calculate, update, and store covariance matrices. Our goal is manyfold. We intent to find a substitute algorithm to KF that

1. filters out errors from modeling the dynamical system,

2. filters out the noise from the data,

3. insures temporal regularization,

4. is an optimal recursive estimate,

5. does not require the storage of past measurement data,

6. guarantees nonnegativity of the solution,

7. does not use matrix-matrix multiplications

8. does not necessitate any matrix inversion, and

9. does not need to calculate, update, or store any covariance matrix.

We aim then to keep the same first five properties of KF while improving it by requesting four more. Each recursive step in the new approach is an iterative reconstruction that involves only matrix-vector multiplication. These should then handle the problems of huge number of variables, such is the case in medical imaging, and would guarantee positive solutions. We mentioned earlier that the SMART filter [9] satisfies these properties. The main goal of this paper is to provide, yet another alternative, based on an extension of the EM algorithm to solve our filtering problem for the dynamic state.

4. EM Filtering

4.1. Towards a Nonnegative Filter

The problem we are trying to solve can be summarized as follows. From one/two/three tomographic projections entire image needs to be estimated for each time frame. This is usually, if not always, a highly underdetermined problem in SPECT and some constraints need to be used in order to get a reasonable estimate. In this paper, we impose few constraints, namely, filtering out errors from modeling the dynamical system and the noise from the data, imposing temporal regularization, and enforcing the nonnegativity of the solution.

A nonnegative approach, applicable to nonnegative vectors and matrices, might need to minimize a distance that applies only to nonnegative quantities; the Kullback-Leibler distance satisfies this constraint. Suppose now that, in addition to the data vector z as in the static situation, we have $y = x + \mu$. We find a new estimate $\hat{x}$ by minimizing the following cost function:

$$F(x) = KL(z, Hx) + KL(y, x)$$

where it is clear that if the prior estimate y of x satisfies $z = Hy$, then the new estimate is y. So the nonnegative filter would use a repeated application of the solution to the minimization problem. However, the covariances do not seem to play a role now, since this is not a least-squares or Gaussian theory. Nevertheless, the two covariance matrices R and Q play crucial roles in KF to filter out the noise from the data and the errors from our modeling of the dynamic system. We would like then to keep this filtering property by utilizing these two matrices. We introduce then a weighted KL distance that will handle the noise and error filtering part; this is covered in the next subsection.

4.2. Weighted Kullback-Leibler Distance

R and Q, being covariance matrices, are symmetric positive definite, so are their inverses R^{-1} and Q^{-1}. Therefore, the Cholesky decomposition implies that there exist upper triangular matrices D_1, and D_2 with strictly positive diagonal entries such that

$$\begin{aligned} R^{-1} &= D_1^\top D_1 & (10) \\ Q^{-1} &= D_2^\top D_2 & (11) \end{aligned}$$

with these decompositions, we have $\|x\|_{R^{-1}}^2 = \|D_1 x\|^2$ and $\|x\|_{Q^{-1}}^2 = \|D_2 x\|^2$. By the same token, we define the weighted KL distance w.r.t. to, for instance

R^{-1}, as follows,

$$KL_{R^{-1}}(a, b) = KL(D_1 a, D_1 b))$$ (12)

A nonnegative approach, applicable to nonnegative vectors and nonnegative matrix H, might be to minimize the following weighted cross-entropy/KL sum

$$KL(D_1 z, D_1 H x) + KL(D_2 y, D_2 x)$$ (13)

It goes without saying that first we should make sure that the four vectors $D_1 z$, $D_1 H x$, $D_2 y$, and $D_2 x$ have nonnegative entries. For our application, the matrix H is any of the observation matrices H_k and the vector z is any of the observation vectors z_k; $k = 1, \cdots, S$. Recall that we assume that we have diagonal covariance matrices, white noise and error, nonnegative z, and normalized system matrix H with nonnegative entries. On one hand, the matrix H and vectors z, y, and x have nonnegative values. On the other hand, matrices D_1 in (10) and D_2 in (11) could have off-diagonal negative entries to the extent that it is not ensured that the four vectors are all nonnegative coordinate-wise. Nonetheless, when we deal with white noise and error, D_1 and D_2 will be diagonal matrices with nonnegative values. This is exactly the case of most applications including dynamic SPECT where these four vectors $D_1 z$, $D_1 H x$, $D_2 y$, and $D_2 x$ are nonnegative. Otherwise, we should first convert a general system to a nonnegative one and then pre-whiten it, refer to [9]. Diagonal covariance matrices, white noise and error, nonnegative detected photons z_k, and normalized system matrices H_k $(k = 1, \cdots, S)$ with nonnegative entries are assumptions that we are adopting throughout this paper. We also assume that the matrix $H_k \in \mathbb{R}^{M \times N}$ has been constructed so that H_k, as well as any submatrix C obtained from H_k by deleting columns, has full rank. In particular, if C is $M \times L$ and $L \geq M$, then C has rank M.

4.3. Cross-Entropy Minimization

Byrne [13] considers the following regularization problem,

$$\min_{x \geq 0} F(x) = \alpha KL(d, Px) + (1 - \alpha)KL(y, x)$$ (14)

His alternating projections algorithm goes something like this. Let x^0 be a starting nonnegative point. Then having got the ℓth iterate x^ℓ, we obtain

$$r_{ij}^{\ell+1} = \frac{P_{ij} x_j^\ell d_i}{(P x^\ell)_i}$$ (15)

$$x_j^{\ell+1} \quad = \quad \alpha \sum_{i=1}^{M} r_{ij}^{\ell+1} + (1 - \alpha)y_j \qquad (16)$$

In this article, we also offer an alternative update formula by gathering both steps (15) and (16) into a compact convex combination form as follows,

$$x_j^{\ell+1} = \alpha x_j^{\ell} \sum_{i=1}^{M} \frac{P_{ij}d_i}{(Px^{\ell})_i} + (1 - \alpha)y_j \qquad (17)$$

For $\alpha = 1$, we obtain the well known EM/ML (expectation maximization/maximum likelihood) iteration (4). He states the following convergence lemma and proves it using the orthogonality conditions of a Pythagorean type.

Lemma 1. *The sequence $\{x^{\ell}\}$ converges to a limit x^{∞} for all M and N, for all staring $x^0 > 0$, for all $y > 0$, and for all $0 \leq \alpha \leq 1$. For $0 \leq \alpha < 1$, x^{∞} is the unique minimizer of $F(x)$. For $\alpha = 1$, x^{∞} is the unique nonnegative minimizer of $KL(d, Px)$ if there is no nonnegative solution of $d = Px$. If there are nonnegative solutions, then $d = Px^{\infty}$ and $x^{\sharp} = x^{\infty}$ is the only solution for which the inequalities $KL(x, x^{\sharp}) \leq KL(x, x^{\ell})$ hold for all $x \geq 0$ with $d = Px$ and all ℓ; it follows that support(x) is contained in support(x^{∞}) for all such x.*

Note that the term $KL(y, x)$ forces temporal smoothness with α as a temporal smoothing parameter. When recursively minimizing this function over x_k, there will be a trade off between fidelity to the data and to smoothness of the TAC. The positive parameter α is a smoothing parameter that controls the relative importance of the two criteria. A crucial step in regularization is the selection of the regularization parameter α.

5. Application to Dynamic SPECT

Our error and noise covariance matrices in the dynamic SPECT problem are usually modeled as diagonal matrices with nonnegative entries, refer to section 2.1.

$$Q_k \quad = \quad \sigma_k^2 I \qquad (18)$$
$$R_k \quad = \quad diag(z_k) \qquad (19)$$

where $diag(a)$ is the square matrix that has the a_i in its main diagonal and 0 otherwise. Our measurements z_k are modeled as Poisson random variables,

thus the choice of $diag(z_k)$ as their covariance matrices. Let

$$D_1 \quad = \quad diag(\frac{1}{\sqrt{z_k}}) \tag{20}$$

$$D_2 \quad = \quad \frac{1}{\sigma_k}I \tag{21}$$

where the vector $\sqrt{z_k}$ is the one having $\sqrt{(z_k)_i}$ in its entry i. In this case, we are sure that $D_1 z \geq 0$, $D_1 Hx \geq 0$, $D_2 y \geq 0$, and $D_2 x \geq 0$. For ease of notation, we drop for a while the subscript k. The weighted KL cost function is then

$$KL(\sqrt{z}, (\frac{1}{\sqrt{z}}).Hx) + KL(\frac{1}{\sigma}y, \frac{1}{\sigma}x) \tag{22}$$

or

$$\frac{\sigma-1}{\sigma}KL(\frac{\sigma}{\sigma-1}\sqrt{z}, \frac{\sigma}{(\sigma-1)\sqrt{z}}.Hx) + \frac{1}{\sigma}KL(y, x) \tag{23}$$

where $a.b$ and a/b designate the multiplication and division respectively of the vectors a and b component wise. We can obtain the same functional (23) by using a pre-whitening [9].

Recall that we aim to find a nonnegative estimate $\hat{x}_k$, $k = 1, \ldots, S$, to the nonnegative unknown x_k of the problem given by the two linear space-state equations (1) and (2),

$$x_k \quad = \quad A_k x_{k-1} + \mu_k$$
$$z_k \quad = \quad H_k x_k + \nu_k$$

μ_k is the error vector, $\mathbb{E}(\mu_k) = 0$ and $\mathbb{E}(\mu_k \mu_k^\top) = Q_k$ is the covariance of the error in modeling the transition from x_{k-1} to x_k. $\mathbb{E}(\nu_k) = 0$ and $\mathbb{E}(\nu_k \nu_k^\top) = R_k$ are the mean and covariance respectively of the noise vector ν_k. Entries of the vector z_k and of both matrices A_k and H_k are nonnegative. We know also from equations (18) and (19) that,

$$Q_k \quad = \quad \sigma_k^2 I$$
$$R_k \quad = \quad diag(z_k)$$

Minimizing the functional (23) at each recursion step k is the same as solving the functional $F(x)$ in (14). It suffices to use the change of variables given in step 2 of the EM filter algorithm that follows and to recall that the predicted activity state y_k is $A_k x_{k-1}$. We apply the iterative method based on alternating projections and given by the formulas (15) and (16) at each time step k. The

clustering point x^∞ will be the estimate state $\hat{x}_k$ we are solving for using equations (1) and (2). The matrix H_k is assumed to be a normalized system matrix $(\sum_{i=1}^{M}(h_k)_{ij} = 1, \forall j = 1, \cdots, N)$ with nonnegative entries. Thus we obtain the following EM filter algorithm,

Algorithm 2. EM Filter Algorithm

1. *Start with $\hat{x}_0 > 0$. For $k = 1, \cdots, S$ execute the following steps*

2. *Assume we have done the recursive step up to time $k - 1$, do the change of variables*

$$
\begin{aligned}
\alpha &= \frac{\sigma_k - 1}{\sigma_k} \\
P &= \frac{1}{\alpha} R_k^{-1/2} H_k = \frac{1}{\alpha} \text{diag}\left(\frac{1}{\sqrt{z_k}}\right) H_k \\
d &= \frac{1}{\alpha} R_k^{-1/2} z_k = \frac{1}{\alpha} \sqrt{z_k}
\end{aligned}
$$

3. *To get $\hat{x}_k$, start with $s^0 = \hat{x}_{k-1}$*

4. *Make $y_k = A_k \hat{x}_{k-1}$*

5. *Do $r_{ij}^{\ell+1} = \frac{P_{ij} s_j^\ell d_i}{(Ps^\ell)_i}$*

6. *Compute $s_j^{\ell+1} = \alpha \sum_{i=1}^{M} r_{ij}^{\ell+1} + (1 - \alpha)(y_k)_j, \qquad \ell = 0, 1, \ldots$*

7. *The update formula for the next estimate is $\hat{x}_k = s^\infty$, where s^∞ is the cluster point of the sequence $(s^\ell)_{\ell \in \mathbb{N}}$.*

Observe that we could combine the two steps 5 and 6 into one step as we did before in (17)

$$
s_j^{\ell+1} = \alpha s_j^\ell \sum_{i=1}^{M} \frac{P_{ij} d_i}{(Ps^\ell)_i} + (1 - \alpha)(y_k)_j, \qquad \ell = 0, 1, \ldots \tag{24}
$$

or in an even simplified form when we combine steps 2, 5, and 6

$$
s_j^{\ell+1} = s_j^\ell \sum_{i=1}^{M} \frac{(h_k)_{ij}}{(R_k^{-1/2} H_k s^\ell)_i} + (1 - \alpha)(y_k)_j, \qquad \ell = 0, 1, \ldots \tag{25}
$$

Note how the next iterate $s_j^{\ell+1}$ in (24) is formed as a sum of a convex combination of the predicted $(y_k)_j$, associated with the same coefficient $1 - \alpha$ as in problem (14), relying only on the evolution model and of the calculated EM iterate, in a similar form as in (4), associated with the same coefficient α as in problem (14) as well relying only on the observation model. The matrix R_k is diagonal, thus the EM filter does not involve any matrix-matrix multiplication or any matrix inversion. It involves only matrix-vector multiplication. The temporal regularization parameter α is well defined and takes values between 0 and 1 when σ_k varies between 1 and ∞. If $\alpha = 0$ at each time step k, that is $\sigma_k = 1$ and $\hat{x}_k = A_k \hat{x}_{k-1}$, then we are discarding completely the observations to the extend we rely only on our evolution model. This defeats the purpose of the experiment. When $\alpha = 1$ at each time step k, the predicted y_k in step 4 is not needed in step 6. Thus we retrieve the EM/ML iteration as in (4), which is only valid for the static case as mentioned before. Indeed, choosing $\alpha = 1$ means that $\frac{\sigma_k - 1}{\sigma_k} = 1$ or simply $\sigma_k = \infty$. That is the covariance matrix in the transition equation (1) is very huge; which implies we have no confidence at all in our evolution model. In other words, we discard the evolutionary state of the variable. Only the observations z_k are meaningful in finding the $\hat{x}_k = \hat{x}$; which is then a stationary state as it should be.

We have the following convergence result; this is a direct consequence of lemma 1

Theorem 3. *For all $k = 1, \cdots, S$, the sequence $\{s^\ell\}$ converges to a limit $\hat{x}_k = s^\infty$ for all M and N, for all staring $\hat{x}_0 > 0$, and for all $1 \leq \sigma_k \leq \infty$. For $1 \leq \sigma_k < \infty$, $\hat{x}_k$ is the unique minimizer of the functional (23). For $\sigma_k = \infty$, $\hat{x} = \hat{x}_k$ is the unique static nonnegative minimizer of $KL(z, Hx)$ if there is no nonnegative solution of $z = Hx$. If there are nonnegative solutions, then $z = H\hat{x}$ and $x^\sharp = \hat{x}$ is the only solution for which the inequalities $KL(x, x^\sharp) \leq KL(x, s^\ell)$ hold for all $x \geq 0$ with $z = Hx$ and all ℓ; it follows that support(x) is contained in support$(\hat{x})$ for all such x.*

Remark 4. In considering these particular mixture of KL distances in (23), we unified approaches commonly taken in the underdetermined and overdetermined cases, and we developed an iterative solution method within a single framework of alternating projections as well as establishing a convergence result.
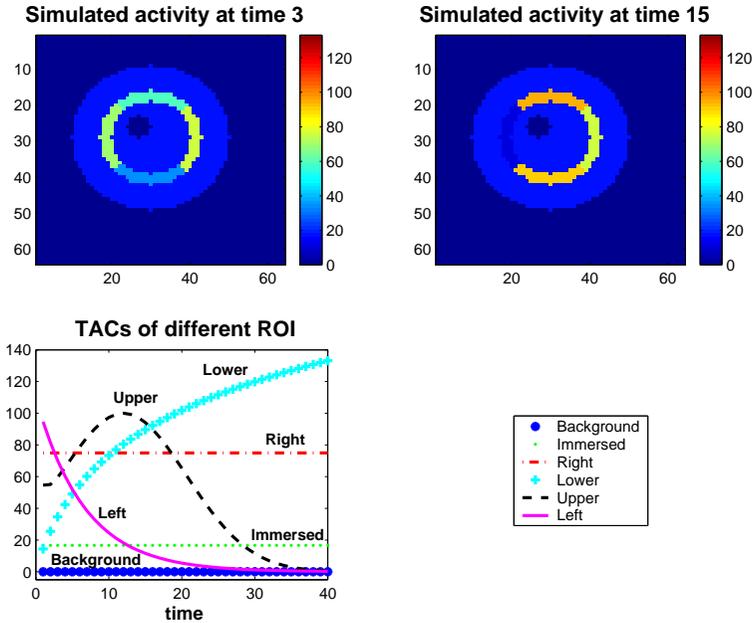
Figure 1: Simulated annulus with its different ROI and their TACs. Upper left: simulated activity at time 3, upper right: simulated activity at time 15, and lower left: TACs of the 6 different ROI.

6. Numerical Experiment

6.1. Simulation

Our phantom is composed of six regions of interest (ROI) or segments. A segment represents a spatial region with similar temporal behavior. In the present study we do not assume the segments to be known exactly. Different approaches to determine the segmentation are described in literature. Each ROI has a different time activity curve (TAC), see figure 1. The example investigated in this work is based on the teboroxime dynamics in the body during first hour post injection. The choice of the TACs is motivated by the behavior of liver, healthy myocardium, muscles, stenotic myocardium, and lungs. Only one slice is modeled; that is we simulate a 2D object. The star-like shape placed on the left ensures that the phantom is not entirely symmetrical. We simulate 120 projections over 360°, one projection for every 3°, with attenuation and a 2D Gaussian detector response.
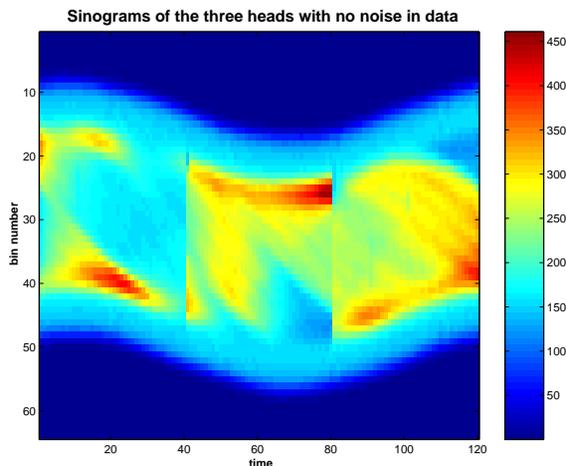
Figure 2: Noiseless sinogram or $2D$ projections without noise: y-axis has the bin number and the x-axis has the 40 time instances of the 3 heads. Time instances from 1 to 40 are for head 1, 41 to 80 for head 2, and 81 to 120 for head 3. A color intensity of a pixel is the number of detected photons by a certain bin at a certain time.

There are three camera heads consisting of 64 square bins each measuring 0.625 cm in each side. The distance from the annulus to the camera head rotation axis is 30 cm. We simulate $S = 40$ time instances for three heads; that is we have $3 \times 40 = 120$ projections for a camera rotating clock wise (CW) in a circular orbit. Head 1 starts at $-60°$, head 2 at $60°$, and head 3 at $180°$. A low energy high resolution (LEHR) collimator is used with a full width at half maximum (fwhm). We determine the blurred parallel strip/beam geometry system matrices for all projections with resolution recovery and attenuation correction [18].

We have 64 projection/measurement values for each head, which amounts to a total of $M = 192$ observations at each time frame. We run tests where we compare reconstructions without added noise to the data, see figure 2, and with noise included into the data/observations, see figure 3. That is, instead of working with the observation z as is in the case of noiseless data, we took rather a Poisson random observation with mean z in the case of noisy data. We noticed that there are very slight differences in the TACs and in the reconstructed images with and without noisy data. We present here only the results with the noise added to the observations. The size of the image we aim to reconstruct is $N = 4096 = 64 \times 64$ dixels. We have six kinds of TACs that
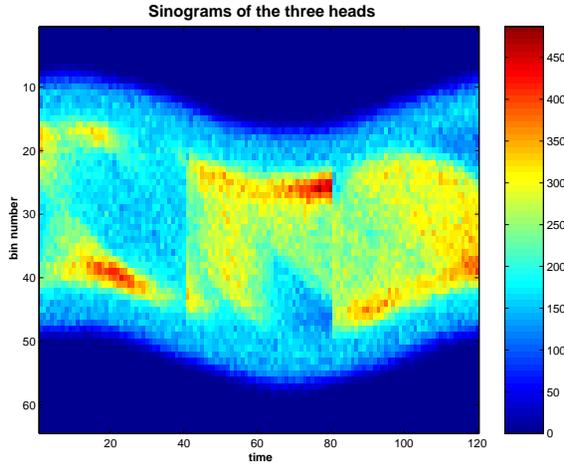
Figure 3: Noisy Sinogram: y-axis has the bin number and the x-axis has the 40 time instances of the 3 heads. Time instances from 1 to 40 are for head 1, 41 to 80 for head 2, and 81 to 120 for head 3. A color intensity of a pixel is the number of detected photons by a certain bin at a certain time. Note: We use this noisy data for the experiment.

are very representative for clinical applications. The annulus has four arcs that we name "Left", "Upper", "Right", and "Lower" according to their location. The activity is decreasing in the Left arc, increasing-decreasing in the Upper arc, constant in the Right arc, and increasing in the Lower arc; see figure 1. The star-like shape has zero activity within it and is called the "Star" region; we refer to it as "Background" too. The annulus is immersed within a region, called "Immersed", that has a constant activity. We have six ROI in total. The EM filter algorithm should work in both underdetermined and overdetermined settings, refer to remark 4. We aim then to test the algorithm 2 in the underdetermined and overdetermined cases. The undermined case happens when we reconstruct dixel by dixel; we possess $M = 192$ noisy data for $N = 4096$ unknowns or a ratio of about 1:21 data to unknowns. It is an ill-posed problem. The overdetermined consists in reconstructing the six ROI, when we assume full knowledge of their locations, so that we have $M = 192$ noisy data for $N = 6$ unknowns or a ratio of 32:1 data to unknowns. We should of course get better reconstructed images in the latter case than in the former one; this, indeed, will be confirmed shortly.

We provide quantitative analysis of the reconstructed images in order to compare the simulated activity with the reconstructed one. We define the rel-

ative deviation error τ of the reconstructed activity v^* from the truth x, refer to (26) through (28). Hence we compare the simulated count $x_{i,k}$ with the corresponding reconstructed one $v^*_{i,k}$ at each time frame k for every location i. We sum over a ROI containing J dixels normalized by the total simulated/true counts in order to diminish the effect of statistical fluctuations. We have a $\tau_{ROI,k}$ for every sector. These indicators allow us to see how the method performs under different dynamic behaviors. We could compare, for instance, sectors with fast washout with those with slow one [17]. We calculate similar τ_k over the total number of doxels (dynamic voxels) N then we average them over the total number S of time acquisitions; so that we have τ_{avg}. This is an objective comparison of the quality of reconstruction for different sets of parameters such as iteration stopping criteria, noise levels, etc. The closer τ_{avg} is to zero, the better the reconstructed images should be.

$$\tau^2_{ROI,k} \quad = \quad \frac{\sum_{j=1}^{J}(v^*_{j,k} - x_{j,k})^2}{\sum_{j=1}^{J} x^2_{j,k}} \tag{26}$$

$$\tau^2_k \quad = \quad \frac{\sum_{j=1}^{N}(v^*_{j,k} - x_{j,k})^2}{\sum_{j=1}^{N} x^2_{j,k}} \tag{27}$$

$$\tau_{avg} \quad = \quad \frac{1}{S}\sum_{k=1}^{S}\tau_k \tag{28}$$

6.2. Results

We assume that the system dynamics are unknown to us (1); therefore we use a random walk. In practical terms, we set $A_k = I$, for all $k = 1, \cdots, S$. For the state transition linear model, we proceeded as follows. We are not interested in the background and we assume that we know the locations of these zero activities; this is a common practice [19, 20, 21]. We have run experiments without this assumption and results are very comparable to when we have run them with this assumption. One interesting way to deal with this assumption is as this. Set to zero the values of the corresponding positions of the matrix H_k. The updating equations (8) and (9) ensure that the updated activities will remain equal to zero; thus the KF reconstructs perfectly the star/background region(s) [16, 17]. By having these values as zero while eliminating these columns and setting to zero the corresponding entries of the estimated activity [9, 10], we guarantee automatically that those entries stay at the value of zero when using our present algorithm EM filter.

We experiment with different initial guesses $\hat{x}_0$ such as $(1, \cdots, 1)^{\top}$ and $(10^{-6}, \cdots, 10^{-6})^{\top}$. We also start the algorithm with the static image given by OSEM; we call this initial guess *OSEM activtiy*. The average of the deviation error τ_{avg} combined with visual inspection show that there is no pronounced advantage in favor of any. We do not have much confidence in our transition model (1) so we choose covariance matrices to be pretty high, $10^3 \leq \sigma_k \leq 10^4$. In the underdetermined case, we do not assume the ROI to be known exactly. However, we make use of these segments only to interpret the results. As a consequence, there are some differences in intensity between pixels within the same region. To assess the effectiveness of the method and of the convergence result 3, we show the TACs averaged over the pixels within the same ROI and this is also valid for the overdetermined case. Follow are the results of both reconstruction cases, underdetermined and overdetermined, SMART filter [9] and their comparison with results obtained using the projected Kalman algorithm [17]. The projected Kalman is the classical KF followed by a projection into the positive octant, using a proximal approach, to ensure the feasibility of the activity. Recall that KF, see equations (8) and (9), necessitates inversion of huge matrices which is computationally time consuming and RAM memory hungry.

6.2.1. Underdetermined Case

We are solving the ill-posed inverse problem in reconstructing the dynamic images of the annulus, 192 noisy observations for 4096 unknowns. We use both algorithms, EM algorithm 2 and SMART filter algorithm [9], on a P4 3.00 GHz desktop. It takes less than 2 min to run the EM filter and about 8 mins, 4 times slower, to run the SMART filter. In contrast to the projected Kalman algorithm [17] which takes more that 2.5 hr, we witness improvements ranging from about 18 to 75 times faster. The SMART filter takes longer than the EM filter because of the many evaluations of the log and exp functions in its steps 5 and 6. It was mentioned in [9] that combining 2 steps of the SMART filter algorithm into one to avoid the evaluation of log and exp functions, evaluating rather a power function instead, might speed up the SMART filter algorithm. The τ_{avg} of both methods is about 0.52 which is the same as with projected Kalman. Images and TACs look fine with both, although the image with EM filter looks slightly better and TACs are somewhat smoother, see figure 4. Images and TACs of both methods are of the same quality as with projected Kalman [17].
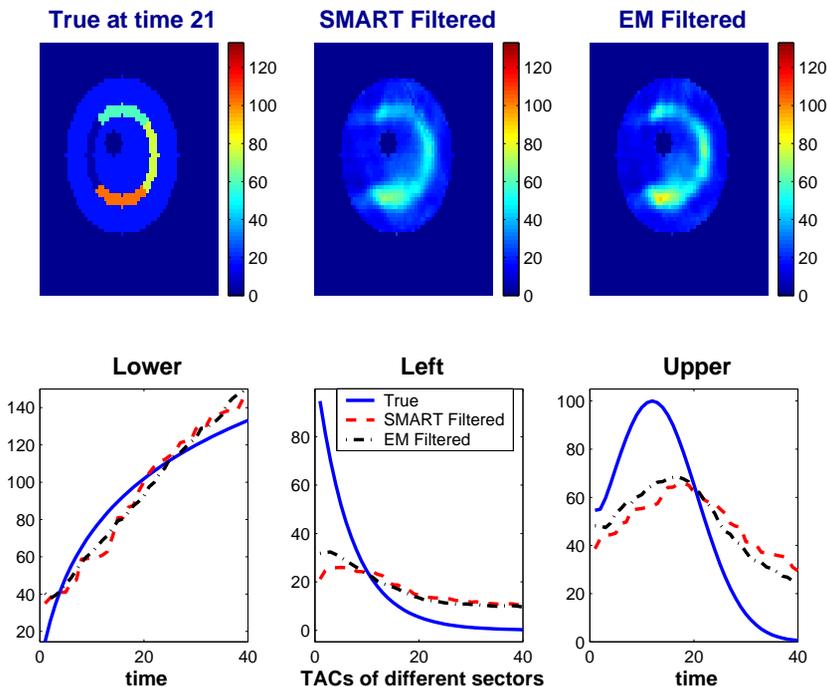
Figure 4: Reconstructed images at time 21 and TACs of 4096 dixels using EM filter and SMART filter.

6.2.2. Overdetermined Case

In medical imaging, we are sometimes not interested in individual intensities of each and every pixel/voxel but rather on some ROI intensities. We are then more concerned with a segmented reconstruction [21]. A CT scan for instance might give us an idea about the ROI. In this case we are solving the inverse problem in reconstructing the dynamic images of the annulus, 192 noisy observations for 6 unknown ROI. We use both algorithms, EM filter 2 and SMART filter algorithm [9], on a P4 3.00 GHz desktop. It takes about 1.5 sec to run the EM filter and about 15 sec to run the SMART filter. In contrast to the projected Kalman algorithm [17] which takes about 1.7 sec, we do not witness any improvement in using EM filter. Again, the SMART filter takes longer than the EM filter, 10 times slower but both times in the seconds, because of the many evaluations of the log and exp functions in its steps 5 and 6. This shortcoming could be remedied, see section 6.2.1 above. The τ_{avg} of both methods is about 0.03 which is half of the one with projected Kalman.
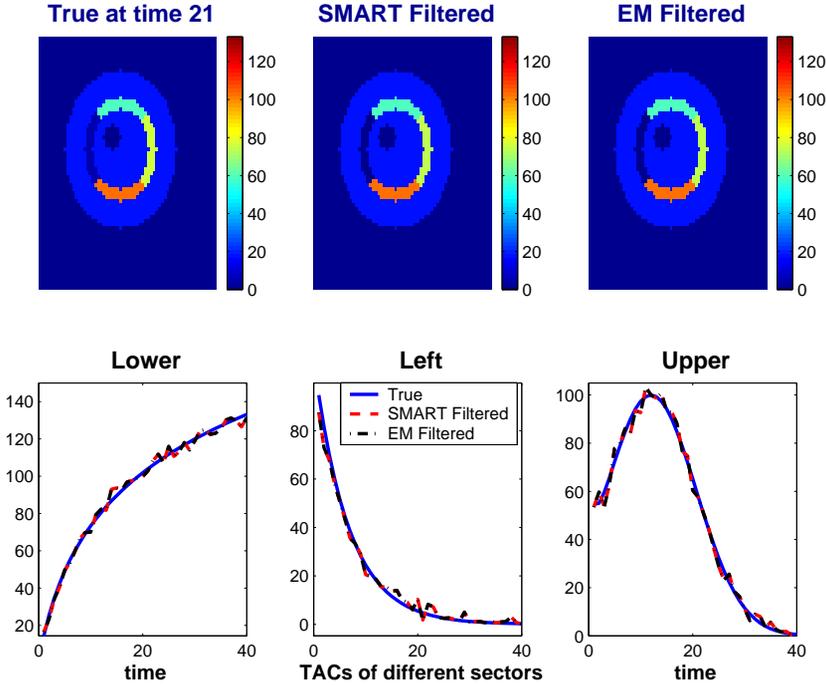
Figure 5: Reconstructed images at time 21 and TACs of 6 ROI using EM filter and SMART.

Hence we improve on convergence. As expected with this overdetermined case, we get much better images and TACs, compare figure 4 to figures 5 and 6. Images and TACs of both methods, EM filter and SMART filter, are of the same quality as with projected Kalman [17].

7. Conclusion

We presented here a novel algorithm that we refer to as EM filter. It applies to nonnegative normalized systems when a nonnegative solution is desired. Our algorithm guarantees this as well as a temporal smoothness. We ran experiments comparing EM filter with SMART filter and projected Kalman in both cases, underdetermined and overdetermined. Quality of images and TACs is about the same in the three reconstructed images, although EM filter performs slightly better than SMART filter. EM filter is 4 to 10 times faster than the SMART filter; still, both are 18 to 75 times faster than the projected Kalman
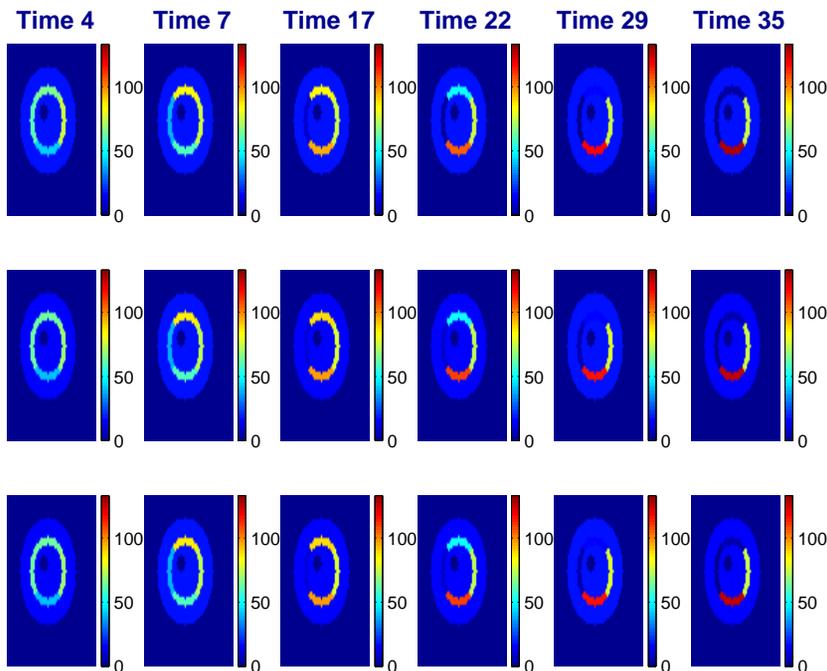
Figure 6: Reconstructed images at various time instances: simulated images in top row, SMART filter reconstructed images in middle row, and EM filter reconstructed images in bottom row.

in the underdetermined case, minutes instead of hours. EM filter performs about the same as projected Kalman time wise; nevertheless, it improves on convergence in the overdetermined case. EM filter algorithm filters out errors from modeling the dynamical system and the noise from the data. It insures temporal regularization and outputs an optimal recursive estimate. It also does not use any matrix-matrix multiplication and does not necessitate any matrix inversion. These last two properties make it very suitable for large scale systems such as the ones in medical imaging. EM filter could be used in any discipline which has used, for instance KF, or in any one that is interested in time-varying variables such as financial risk assesment/evaluation and forecasting or control, especially when there is concern with nonnegative outputs. To confirm corollary 3, we applied the algorithm to time-varying SPECT, a medical imaging modality in nuclear medicine. Our results substantiate the efficiency of this novel EM filter. Like two faces of the same coin, EM filter and SMART filter are two algorithms deduced from the same weighted KL distance.

Acknowledgments

References

[1] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)*, **39** (1977), 1-38.

[2] L.A. Shepp, Y. Vardi, Maximum likelihood reconstruction for emission tomography, *IEEE Trans. Med. Imag.*, **1** (1982), 113-122.

[3] A. Hudson, R.S. Larkin, Accelerated image reconstruction using ordered subsets of projection data, *IEEE Transactions on Medical Imaging*, **13** (1994), 601-609.

[4] H.H. Bauschke, D. Noll, A. Celler, J. M. Borwein, An EM-algorithm for dynamic SPECT tomography, *IEEE Trans. Med. Imag.*, **18** (1999), 252-261.

[5] M.A. Schulze, An edge-enhancing nonlinear filter for reducing multiplicative noise, In: *Proc. SPIE Vol. 3026*, (1997), 46-56.

[6] B.D.O. Anderson, J.B. Moore, *Optimal Filtering*, Printice-Hall, Englewood, Ciffs, NJ (1979).

[7] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*, Wiley-Interscience, (2006).

[8] R.E. Kalman, K.G. Smythe, A new approach to linear filtering and prediction problems, *Trans. of the ASME-Jour. of Basic Eng.*, **82** (1960).

[9] J. Qranfal, C. Byrne, SMART Filter and its Application to Dynamic SPECT Reconstruction, *Int. J. of Pure and Appli Math*, **to appear** (2011).

[10] C.L. Byrne, *Signal Processing, A Mathematical Approach*, A K Peters, Wellesley, MA (2005).

[11] S. Kullback, R. Leibler, On information and sufficiency, *J. Theoret. Biol.*, **22** (1951), 79-86.

[12] D. Titterington, On the iterative image space reconstruction algorithm for ECT, *IEEE Trans. Med. Imag.*, **1** (1987), 52-56.

[13] C.L. Byrne, Iterative image reconstruction algorithms based on cross-entropy minimization, *IEEE Trans. on Image Processing*, **2** (1993), 96-103.

[14] R. Gordon, R. Bender, G.T. Herman, Algebraic reconstruction technique (ART) for three-dimentional electron microscopy and X-ray photography, *Ann. Math. Statist*, **29** (1970), 471-481.

[15] T. Farncombe, A. Celler, C. Bever, D. Noll, J. Maeght, R. Harrop, The Incorporation of Organ Uptake into Dynamic SPECT (dSPECT) Image Reconstruction, *Nuclear Science, IEEE Transactions on*, **48** (2001), 3-9.

[16] J. Qranfal, G. Tanoh, Regularized Kalman filtering for dynamic SPECT, In: *J. Phys.: Conf. Ser.*, **124** (2008).

[17] J. Qranfal, *Optimal Recursive Estimation Techniques for Dynamic Medical Image Reconstruction*, Simon Fraser University (2009).

[18] G. Tanoh, *Algorithmes du point intérieur pour l'optimisation en tomographie dynamique et en mécanique du contact*, Université Paul Sabatier (2004).

[19] T. Farncombe, *Functional Dynamic SPECT Imaging Using a Single Slow Camera Rotation*, University of British Columbia (2000).

[20] M. Kervinen, M. Vauhkonen, J.P. Kaipio, P.A. Karjalainen, Time-varying reconstruction in single photon emission computed tomography, *Int. J. of imaging syst. tech*, **14** (2004), 186-197.

[21] B.W. Reutter, G.T. Gullberg, R.H. Huesman, Direct least Squares Estimation of Spatiotemporal Distribution from Dynamic SPECT Projections using Spatial Segmentation and Temporal B-splines, *IEEE trans. med. imaging*, **19** (2000), 434-450.

404