AN ITERATIVE BAYES ALGORITHM FOR EMISSION TOMOGRAPHY USING A SMOOTHED SINOGRAM

Jun Ma

Department of Statistics, Macquarie University
NSW 2109, Australia

ABSTRACT

In this paper we formulate a new approach to medical image reconstruction from projections in emission tomography. This approach differs from traditional methods such as filtered back projection, maximum likelihood or maximum penalized likelihood. Our method is developed directly from the Bayes formula and the final result is an iterative algorithm, for which the maximum likelihood expectation-maximization of [1] (or [2]) is a special case.

1. INTRODUCTION

Maximum likelihood (ML) and maximum penalized likelihood (MPL) are widely used reconstruction methods in emission tomography. For ML image reconstructions, the iterative expectation maximization (EM) algorithm (see, e.g., [2]) is widely practiced. In the context of image and signal processing it is well known that the EM algorithm is indistinguishable from the Richardson-Lucy (RL) algorithm ([3], [4]). The RL method is a Bayesian based iterative algorithm; its development comes directly from the Bayes conditional probability formula.

Let $P(j)$ be the probability of a photon being emitted from voxel $j$ and $P(i)$ the probability of a photon being detected in camera bin $i$. Similarly, we use $P(i | j)$ to denote the probability of a photon being detected in bin $i$ given that it is released from voxel $j$ and $P(j | i)$ the probability of a photon being emitted from voxel $j$ given it arrives in bin $i$.

Probability $P(j)$ is not known; however if $x_j$ is available, then one can estimate $P(j)$ by $\hat{P}(j) \propto x_j$.

For $P(i | j)$, as $a_{ij}$ contains various information about projection from voxel $j$ to camera bin $i$, such as system geometry, attenuation, scattering etc, it is commonly accepted to estimate $P(i | j)$ by $P(i | j) \propto a_{ij}$.

Probability $P(j | i)$ can be evaluated from $P(j)$ and $P(i | j)$ using the Bayes formula, namely,

$$ P(j | i) = \frac{P(i | j)P(j)}{\sum_{i=1}^{p} P(i | t)P(t)}. \quad (1) $$

As the $P(j)$’s are unknown, we may compute $P(j | i)$ iteratively. From the current estimate of $x$ (denoted by $x^k$), first obtain the following update to $P(j | i)$ from (1):

$$ \hat{P}^k(j | i) = \frac{a_{ij}x^k_j}{\sum_{t=1}^{p} a_{it}x^k_t} = \frac{a_{ij}x^k_j}{\mu^k_i}, \quad (2) $$

where $\mu^k_i = \sum_{j=1}^{p} a_{ij}x^k_j$, then from $\hat{P}^k(j | i)$ we can develop $x^{k+1}$. Note for any emitted photon from voxel $j$ the probability of it being detected is $\sum_{i=1}^{n} a_{ij}$; thus,

$$ \left( \sum_{i=1}^{n} a_{ij} \right) x^{k+1}_j = \sum_{i=1}^{n} \hat{P}^k(j | i)\mu_i, \quad (3) $$

The rest of this paper is organized as follows. Section 2 formulates the IB algorithm. Section 3 discusses how to estimate the $\mu_i$’s by smoothing the measurements $y_i$. Section 4 provides an ordered subset formulation of the algorithm. Simulation studies are given in Section 5 and discussions in Section 6.

2. IMAGE RECONSTRUCTION BY THE ITERATIVE BAYES ALGORITHM

Let $P(j)$ be the probability of a photon being emitted from voxel $j$ and $P(i)$ the probability of a photon being detected in camera bin $i$. Similarly, we use $P(i | j)$ to denote the probability of a photon being detected in bin $i$ given that it is released from voxel $j$ and $P(j | i)$ the probability of a photon being emitted from voxel $j$ given it arrives in bin $i$.

Probability $P(j)$ is not known; however if $x_j$ is available, then one can estimate $P(j)$ by $\hat{P}(j) \propto x_j$.

For $P(i | j)$, as $a_{ij}$ contains various information about projection from voxel $j$ to camera bin $i$, such as system geometry, attenuation, scattering etc, it is commonly accepted to estimate $P(i | j)$ by $P(i | j) \propto a_{ij}$.

Probability $P(j | i)$ can be evaluated from $P(j)$ and $P(i | j)$ using the Bayes formula, namely,

$$ P(j | i) = \frac{P(i | j)P(j)}{\sum_{i=1}^{p} P(i | t)P(t)}. \quad (1) $$

As the $P(j)$’s are unknown, we may compute $P(j | i)$ iteratively. From the current estimate of $x$ (denoted by $x^k$), first obtain the following update to $P(j | i)$ from (1):

$$ \hat{P}^k(j | i) = \frac{a_{ij}x^k_j}{\sum_{t=1}^{p} a_{it}x^k_t} = \frac{a_{ij}x^k_j}{\mu^k_i}, \quad (2) $$

where $\mu^k_i = \sum_{j=1}^{p} a_{ij}x^k_j$, then from $\hat{P}^k(j | i)$ we can develop $x^{k+1}$. Note for any emitted photon from voxel $j$ the probability of it being detected is $\sum_{i=1}^{n} a_{ij}$; thus,

$$ \left( \sum_{i=1}^{n} a_{ij} \right) x^{k+1}_j = \sum_{i=1}^{n} \hat{P}^k(j | i)\mu_i, \quad (3) $$

The rest of this paper is organized as follows. Section 2 formulates the IB algorithm. Section 3 discusses how to estimate the $\mu_i$’s by smoothing the measurements $y_i$. Section 4 provides an ordered subset formulation of the algorithm. Simulation studies are given in Section 5 and discussions in Section 6.
where $\mu_i$ is the expected counts (unobservable) for bin $i$.

Equation (3) cannot be implemented yet as the $\mu_i$‘s are
unknown. We estimate $\mu_i$ exploiting the smoothness of the
$\mu_i$‘s, as neighboring $\mu_i$ values are believed to be similar.

Denote the estimate of $\mu_i$ by $\hat{\mu}_i$. Substituting $\hat{\mu}_i$ into (3)
gives an algorithm for estimating the emission intensity vector $x$:

$$ x_j^{k+1} = \frac{x_j^k \sum_{i=1}^{n} a_{ij} \hat{\mu}_i}{\sum_{i=1}^{n} a_{ij} \hat{\mu}_i}. $$

(4)

We call this the Iterative Bayes (IB) algorithm. Note when
$\hat{\mu}_i = y_i$, this algorithm coincides with the EM algorithm of
Vardi et al.[6] for emission tomography. Discussions on how
to obtain $\hat{\mu}_i$ are given in Section 3.

Following the convergence proof in [6] for the EM algo-
rithm we can show that, under certain regularity conditions,
the IB algorithm converges (with the initial $x^0$ > 0) to one of
the global solutions maximizing

$$ d(x) = -\sum_{i=1}^{n} \mu_i + \sum_{i=1}^{n} \hat{\mu}_i \log \mu_i. $$

(5)

for $x \geq 0$. Note $d(x)$ is a concave function, thus there exist
global maxima (may not be unique).

Let $\mu$ and $\hat{\mu}$ be the vector for all $\mu_i$ and $\hat{\mu}_i$. The Kullback-
Leibler (KL) distance between $\hat{\mu}$ and $\mu$ is

$$ KL(\hat{\mu}, \mu) = \sum_{i=1}^{n} \left( \hat{\mu}_i \log \frac{\hat{\mu}_i}{\mu_i} - \hat{\mu}_i + \mu_i \right). $$

(6)

Maximizing $d(x)$ is equivalent to minimizing $KL(\hat{\mu}, \mu)$. Thus
the IB algorithm finds a positive minimum $KL(\hat{\mu}, \mu)$ (minKL)
estimate.

3. NONPARAMETRIC SMOOTHING OF THE MEASUREMENTS

Estimation of $\mu$ is equivalent to the problem of nonparametric
regression, with an identity link function and Poisson noise in
the context of generalized linear models (GLM); see [7]. Gy-
orf et al.[8] covers almost all known nonparametric regres-
sion methods.

For emission tomography, as the likelihood function is
readily available, it is natural to use the penalized likelihood
method to estimate $\mu$.

As in [5], we estimate $\mu$ on an angle-by-angle basis. That
is, only those $\mu_i$ on the same projection angle are estimated
simultaneously; for different projection angles estimation pro-
cedures are independent. Suppose there are $M$ projection an-
gles. Let $S_t$ be the index set for projection angle $t$ and let $y_t$ and $\mu_t$ be, respectively, the vector for measurements and expected measurements of projection angle $t$. Given $y_t$, our aim is to find non-parametrically a smoothed estimate of $\mu_t$ using MPL.

From set $y_t$, the penalized log-likelihood function is:

$$ g(\mu_t) = -\sum_{i \in S_t} \mu_i + \sum_{i \in S_t} y_i \log \mu_i - \frac{1}{2} \lambda b(\mu_t), $$

(7)

where $\lambda > 0$ is the smoothing parameter and $b(\mu_t)$ is the
penalty function. We call $\lambda$ the projection smoothing para-
eter.

The function $b(\mu_t)$ is chosen to penalize local differences
between neighboring $\mu_i$. In this paper we let $b(\mu_t)$ be a
penalty function known as the “roughness penalty” defined
below. In this context the maximum of $g(\mu_t)$ is a natural cu-
bic spline [7].

Let $f_t(\xi)$ be a smooth function connecting all $\mu_i (i \in S_t)$
along angle $t$. If we assume each angle comprises the same
number $D$ of camera bins, the roughness penalty remains un-
changed for different projection angles as the supporting in-
terval for $f_t(\xi)$ is always $[1, D]$.

The roughness penalty is $b(f_t) = \int_1^D (f_t''(\xi))^2 d\xi$. Ac-
cording to [7, Chapter 2], given $D \geq 2$, if $f_t(\xi)$ is the cubic
spline interpolating $\mu_i (i \in S_t)$ with knots $\{1, 2, \ldots, D\}$, then

$$ \int_1^D (f_t''(\xi))^2 d\xi = \mu_t^T K \mu_t \overset{\text{def}}{=} b(\mu_t), $$

(8)

where $K$ is a $D \times D$ penalty matrix. Construction of the
matrix $K$ is very simple; details are given in [7].

For each angle $t$ our aim is to estimate $\mu_t$ maximizing
$g(\mu_t)$ of (7), with $b(\mu_t)$ given by (8), subject to the non-
negativity constraints $\mu_i \geq 0$ ($i \in S_t$). This is a special
case of the general non-negatively constrained inverse prob-
lem described in [9]; hence the multiplicative iterative inver-
sion method (MIIL) of [9] can be implemented directly.

4. ORDERED SUBSET FORMULATIONS

The ordered subset EM (OSEM) [10] algorithm has attracted
many research activities recently. We develop ordered subset
IB (OSIB) algorithms by adopting the idea of OSEM.

Let $\{S_u, u = 1, \ldots, L\}$ be a partition of the measure-
ments index set $\mathcal{N} = \{1, \ldots, n\}$, namely, $\bigcup_{u=1}^{L} S_u = \mathcal{N}$
and $S_u \cap S_v = \emptyset$ for $u \neq v$. Let $x^{k, u}$ denote the update of $x$
at sub-iteration $u$ of iteration $k$. At iteration $k$ of the OSIB
algorithm, $x$ is updated through $L$ sub-iterations using the
corresponding $\mu$ and $A$ subsets. More specifically, starting with
$x^{k, 0} = x^{k-1}$,

$$ x^{k, u}_j = \frac{1}{\sum_{i \in S_u} a_{ij} \hat{\mu}_i} \sum_{i \in S_u} a_{ij} \hat{\mu}_i, $$

(9)

where $\hat{\mu}_i$ is computed for $i \in S_u$, only, and is given by

$$ \hat{\mu}_i = \frac{a_{ij} x^{k, u-1}_j}{\sum_{t=1}^{p} a_{it} x^{k, u-1}_t}. $$

(10)

1213
Clearly, OSIB is unsatisfactory. In each sub-iteration, it estimates the entire $x$ by only a single subset and, at the same time, it attempts to match $\hat{\mu}_i$ with $\mu_i$ for this subset.

A more reliable approach is offered by the following modification. It updates $x$ using all the data rather than only the current subset:

$$x^{k,u}_j = \frac{1}{\sum_{i=1}^{m} a_{ij}} \sum_{i=1}^{n} \tilde{P}^{k,u}(j \mid i) \hat{\mu}_i,$$  \hspace{1cm} (11)

where

$$\tilde{P}^{k,u}(j \mid i) = \begin{cases} a_{ij}x_j^{k,u-1}/\mu_i^{k,u-1} & \text{for } i \in S_u \\ \tilde{P}_j^{k,u-1}(j \mid i) & \text{for } i \notin S_u \end{cases}$$  \hspace{1cm} (12)

with $\tilde{P}_j^0(j \mid i) = a_{ij}x_j^{k-1}/\mu_i^{k-1}$. Note that the $\tilde{P}(j \mid i)$’s for the current subset are updated and the others are unchanged.

This algorithm appears similar to the complete data OSEM (COSEM) algorithm of [11], and hence is named the COSIB algorithm.

Convergence properties of OSEM and COSEM-ML are studied in [10] and [11] respectively; these results apply directly to the OSIB and COSIB algorithms. In particular, we can show that the COSIB algorithm converges to one of the global maxima of $d(x)$.

5. SIMULATION STUDIES

In this section we explore the performance of IB, OSIB and COSIB for the positive minKL estimate, in 2D SPECT, by applying them to a simulated image. We will also compare them with the EM algorithm of Shepp and Vardi [1] for positive ML estimate and the multiplicative iterative inversion algorithm (MIIL) of Ma [9] for positive MPL estimate.

The simulation used an elliptical phantom of size $64 \times 64$ pixels. There were 64 attenuated parallel beam projections uniformly spaced over $360^\circ$, and each projection contained 64 measurements. Attenuation coefficients were $0.15 / \text{cm}$ (water) within the body, except for within the two lungs, where coefficients were $0.375 / \text{cm}$.

The projection matrix $A$ (with dimension $64^2 \times 64^2$) was pre-determined by the geometry of pixels, adjusted according to the attenuation coefficients. Poisson noises were added to $\mu = Ax$ to form the observed measurements vector $y$. The total measurements in projections was 400,605.

We used the Root Mean Square (RMS), defined as:

$$\text{RMS} = \sqrt{\frac{\sum_{j=1}^{p} (\hat{x}_j - x_j)^2}{p}},$$  \hspace{1cm} (13)

to evaluate the quality of estimates; for each algorithm, we computed the RMS in each iteration against the phantom image. Note that RMS alone is not a perfect measure for assessing image differences; a better measure, which is not available in a closed from, is given by replacing $\hat{x}_j$ by $E(\hat{x}_j)$ in RMS.

We used a quadratic penalty $J(x) = \frac{1}{2}x^TRx$ in the MPL estimate with $R$ given by $r_{jt} = -0.25$ for $j \neq t$ and $t$ in the first-order neighborhood of $j$, $r_{jj} = 1$ for all $j$ and $r_{jt} = 0$ for all other $j$ and $t$. The “optimal” smoothing parameter, which was found by minimizing the RMS, was $1.053 \times 10^{-5}$.

We implemented IB, COSIB with 8 and 64 (denoted by COSIB-8 and COSIB-64) subsets and OSIB with 8 (denoted by OSIB-8) subsets. We wish to exam their convergence performance.

Figure 1 exhibits various reconstructions. The EM results show that the ML estimates at high iterations are less attractive. The MPL solution, even with the optimal smoothing parameter, is also less satisfactory as it loses boundary details. In contrast, IB and its OS algorithms offer the best estimate; for this subset.

Figure 2 (b) displays $d(x^{(k)}) - d(x^{(\infty)})$ against the iteration number $k$, where $x^{(\infty)}$ represents the true positive image.
minKL solution and is approximated by 5000 iterations of IB. For COSIB, the subset number has negligible influence on convergence speed; COSIB-64 appears only about 3 to 4 iterations ahead of COSIB-8. However, COSIB (both 8 and 64 subsets) converges faster than IB; COSIB doubles the convergence speed of IB. OSIB-8 has initial fast convergence, but then is moved away from the minKL solution.

6. DISCUSSION

In this article we have developed a new image reconstruction method for emission tomography using simply the Bayes formula. When the measurements are unsmoothed then it becomes the EM algorithm.

The simulation studies show that, even with non-optimal projection smoothing, this method achieves lower RMS than the optimal MPL solution. Also, it enhances boundary preservation when compared with MPL.

Although we have developed the method in the context of histogram mode observations, it can be easily extended to the list mode. Hence it can be treated as a unifying approach for both type of measurements.

7. REFERENCES


