Image reconstruction by conditional entropy maximisation for PET system

P.P. Mondal and K. Rajan

Abstract: The authors show that the conditional entropy maximisation algorithm is a generalised version of the maximum likelihood algorithm for positron emission tomography (PET). Promising properties of the conditional entropy maximisation algorithm are as follows: an assumption is made that the entropy of the information content of the data should be maximised; it is a consistent way of selecting an image from the very many images that fit the measurement data; this approach takes care of the positivity of the reconstructed image pixels, since entropy does not exist for negative image pixel values; and inclusion of prior distribution knowledge in the reconstruction process is possible. Simulated experiments performed on a PET system have shown that the quality of the reconstructed image using the entropy maximisation method is good. A Gibbs distribution is used to incorporate prior knowledge into the reconstruction process. The mean squared error (MSE) of the reconstructed images shows a sharp new dip, confirming improved image reconstruction. The entropy maximisation method is an alternative approach to maximum likelihood (ML) and maximum *a posteriori* (MAP) methodologies.

1 Introduction

Entropy maximisation is a general tool for solving a large number of inverse problems where the possibility of direct solution is ruled out. Defining an entropy measure in the context of image reconstruction was originally due to Friden [1, 2]. The maximum likelihood (ML) algorithm plays a vital role in emission tomography (ET) [3, 4]. Though a few generalisations of ML estimation, such as Bayesian approaches to ET do exist, we propose to take an entropy-based approach to ET. It is shown theoretically that the proposed algorithm leads to another generalisation of the ML algorithm for ET.

ET is an imaging technique potentially useful in the study of human physiology and organ function. The basic aim of ET is to obtain a quantitative map of the spatial and temporal distribution of radionuclides inside the human body which represent functional activity such as pain [5], motor action [6], etc. This is achieved by realistic modelling of a physical phenomenon involved along with the event measurement count of gamma rays either in coincidence or singly, depending upon the modality used. For positron emission tomography (PET), coincidence counts are detected due to the positron–electron annihilation process, while a single count is detected in single photon emission tomography (SPECT).

The maximum likelihood estimation (MLE) algorithm based on expectation maximisation (EM) produces a sequence of images that converge to the maximum likelihood estimate of the emission density. The quality of the image produced by MLE is found to be better than that

generated by deterministic algorithms such as convolution back projection (CBP) [3, 4, 7], especially with low measurement count data. The ML algorithm has not been found suitable for routine clinical application owing to its excessive computational requirements and the deteriorating nature of the reconstructed images [8]. Time constraint is not a serious concern because it can be reduced considerably by parallelising the algorithm [9, 10] and by using fast digital signal processor (DSP) chips as processing elements [11]. Various algorithmic methods have been suggested to accelerate the convergence of the ML algorithm [12–16]. In particular, one can accelerate the image reconstruction process using an ordered subset (OS) method [17]. Many variants of the ordered subset algorithm, such as OSEM [18], row–action maximum likelihood (RAMLA) [19], rescaled back iterative (RBI-EMML) [20] and RBI-MAP [21] have been proposed in the literature to reduce the number of iterations.

It is well known that as the iterations of the ML algorithm continue, estimates become more and more noisy. This has been termed ‘dimensional instability’ and occurs due to the application of unconstrained maximum likelihood estimation of density functions based on a point process [22]. To overcome this serious problem, two approaches have been suggested:

- (i) The iteration has to be stopped before the quality of the reconstructed image starts deteriorating [8].
- (ii) Inclusion of prior distribution knowledge of the emission densities in the reconstruction process [23, 24].

Obviously, approach (ii) is more promising, as this enables the possibility of improved reconstruction over the ML algorithm. In view of this possibility, a generalised version of the ML algorithm was developed [25, 26]. This algorithm considers conditional entropy as a cost function rather than a likelihood function. The motivation behind conditional entropy maximisation is that the resulting image has minimum configurational information so that there must be information in the measurement data for any structure to be seen [27, 28].

Maximum entropy estimation provides a consistent way of selecting a single image from the very many images that fit the measurement data. It is found that maximum entropy estimation does not introduce correlation in the image beyond that required by the data [29–31]. Also this addresses the important factor of the positivity of image pixels, since entropy does not exist for negative image pixels. Moreover, this method provides a handle for the possible inclusion of prior knowledge in the image reconstruction process.

2 Maximum entropy reconstruction for positron emission tomography

Byrne [32, 33] has shown that maximisation of objective functions such as likelihood, posterior, entropy etc. is equivalent to minimisation of functionals of the form $F(\lambda) = \alpha KL(y, Q\lambda) + (1 - \alpha)D(\lambda)$, where $KL(y, Q\lambda) = \sum_n y_n \log(y_n/[Q\lambda]_n) + [Q\lambda]_n - y_n$ is the Kullback–Leibler (KL) distance between two non-negative vectors, i.e. the actual count y and the expected counts $Q\lambda$, $D(\lambda)$ is some penalty function that forces a degree of smoothness upon λ . Minimising $F(\lambda)$ by choosing $D(\lambda) = KL(P, \lambda)$ for $\lambda \geq 0$ and $0 < \alpha \leq 1$ produces the iterative MAP algorithm of Lange *et al.* [14], where $P(\lambda)$ is a prior estimate of λ . Titterton [34] proved that maximising the likelihood function derived from a Poisson model is equivalent to minimising $KL(y, Q\lambda)$. Instead, given the prior estimate $P(\lambda)$, minimising the functional $KL(\lambda, P(\lambda))$, subject to non-negativity ($\lambda \geq 0$) and data consistency ($Q\lambda = y$) is equivalent to entropy maximisation [32, 33].

The basic idea behind the conditional entropy maximisation algorithm is as follows: let $\mathbf{Y}$ be the observable measurement space and $y \in \mathbf{Y}$ be the measurement data vector. The probability density function (or the likelihood function) for the measurement process $\mathbf{Y}$ is $P_Y(y; \lambda)$, defined over the space $\mathbf{Y}$ where $\lambda \in \mathbf{\Lambda}$ is an unknown parameter vector to be estimated. In general, $P_Y(y; \lambda)$ is complicated to analyse and the estimation of λ by direct maximisation of the likelihood function $P_Y(y; \lambda)$ is difficult. The entropy maximisation formulation postulates the existence of an unobservable space $\mathbf{X}$ which is related to the space $\mathbf{Y}$ by a many to one mapping $x \rightarrow y(x)$ taking each element of $x \in \mathbf{X}$ to an element $y(x) \in \mathbf{Y}$. It often happens that the number of parameters needed for an adequate description is considerably greater than the number of available measurements and so there are an infinite number of possible solutions. Given this formulation, we have to pick the best in some sense. The procedure begins with the assignment of entropy measure to indicate the uncertainty in the statistical environment. Each measurement or piece of prior information reduces the randomness and restricts statistical freedom. Once a given model is adopted, the maximum entropy approach proceeds by finding that reconstruction which maximises the entropy measure and at the same time remains compatible with both the measurement data and *a priori* available information. In situations where the data are insufficient to eliminate all uncertainty, the principle of maximum entropy can be used to assure that the admittedly incomplete description of the situation reflects only the statistical structure imposed by the available information. We propose to use this approach for the PET system, but this will remain valid for other modalities of emission tomography such as SPECT.

The measurements in PET, $y_j, j = 1, \dots, M$ are modelled as independent Poisson random variables with mean parameters $\sum_{i=1}^N \lambda_i p_{ij}$ for $j = 1, \dots, M$, where $\lambda_i, i = 1, \dots, N$ are the mean parameters of the emission process

and p_{ij} is the probability that an annihilation in the i th pixel is detected in the j th detector. The conditional probability for observing $\mathbf{Y} = \mathbf{y}$ given the emission parameter $\mathbf{\Lambda} = \lambda$ is the joint probability of the individual Poisson process expressed as

$$P(y/\lambda) = \prod_{j=1}^M \left[\frac{\exp(-\sum_{i=1}^N \lambda_i p_{ij}) (\sum_{i=1}^N \lambda_i p_{ij})^{y_j}}{y_j!} \right] \quad (1)$$

Now, the conditional entropy of the measurement process $\mathbf{Y} = \mathbf{y}$ given the object image $\mathbf{\Lambda} = \lambda$ is given by

$$H(y/\lambda) = - \sum_{\lambda} P(\lambda) \left[\sum_y P(y/\lambda) \log P(y/\lambda) \right] \quad (2)$$

Conditional entropy measures the information content of y when λ is known and our aim is to maximise the information content in the data set when the object image is given. So, the first derivative of the cost function with respect to $\lambda_i, i = 1, \dots, N$ produces

$$\begin{aligned} \frac{\partial H(y/\lambda)}{\partial \lambda_i} = & - \frac{\partial P(\lambda)}{\partial \lambda_i} \left[\sum_y P(y/\lambda) \log P(y/\lambda) \right] \\ & - \left[P(\lambda) \frac{\partial}{\partial \lambda_i} \left(\sum_y P(y/\lambda) \log P(y/\lambda) \right) \right] \end{aligned} \quad (3)$$

Expanding and simplifying (3) we obtain

$$\begin{aligned} \frac{\partial H(y/\lambda)}{\partial \lambda_i} = & - \sum_y \left[\frac{\partial P(\lambda)}{\partial \lambda_i} (P(y/\lambda) \log P(y/\lambda)) \right. \\ & \left. + P(\lambda) P(y/\lambda) (1 + \log P(y/\lambda)) \frac{\partial}{\partial \lambda_i} (\log P(y/\lambda)) \right] \end{aligned} \quad (4)$$

The Kuhn–Tucker condition [35] for λ_i to be the optimal solution of the nonlinear optimisation problem is

$$\lambda_i \frac{\partial H(y/\lambda)}{\partial \lambda_i} \Big|_{\lambda_i = \hat{\lambda}_i} = 0, \text{ if } \lambda_i > 0 \quad (5)$$

$$\frac{\partial H(y/\lambda)}{\partial \lambda_i} \Big|_{\lambda_i = \hat{\lambda}_i} \leq 0, \text{ if } \lambda_i = 0 \quad (6)$$

for $i = 1, \dots, N$.

Solution of the maximisation problem can be obtained by solving the set of nonlinear equations (5) and (6). Substituting the derivative of $H(y/\lambda)$ from (4) in (5), we obtain

$$\begin{aligned} \sum_y \hat{\lambda}_i \left[\frac{\partial P(\lambda_i)}{\partial \lambda_i} (P(y/\lambda) \log P(y/\lambda)) + P(\lambda_i) P(y/\lambda) \right. \\ \left. \times (1 + \log P(y/\lambda)) \frac{\partial}{\partial \lambda_i} (\log P(y/\lambda)) \right]_{\lambda_i = \hat{\lambda}_i} = 0 \end{aligned} \quad (7)$$

Since we have only one realisation of the measurement, (7) becomes

$$\begin{aligned} \hat{\lambda}_i \left[\frac{\partial P(\lambda_i)}{\partial \lambda_i} (P(y/\lambda) \log P(y/\lambda)) + P(\lambda_i) P(y/\lambda) \right. \\ \left. \times (1 + \log P(y/\lambda)) \frac{\partial}{\partial \lambda_i} (\log P(y/\lambda)) \right]_{\lambda_i = \hat{\lambda}_i} = 0 \end{aligned} \quad (8)$$

Differentiating (1) with respect to λ_i produces

$$\frac{\partial}{\partial \lambda_i} (\log P(y/\lambda)) = \left[-\sum_{j=1}^M p_{ij} + \sum_{j=1}^M \frac{y_j p_{ij}}{\sum_{i=1}^N \lambda_i p_{ij}} \right] \quad (9)$$

Substitution of (9) in (8) gives

$$\begin{aligned} P(y/\hat{\lambda}) \left[\lambda_i \frac{\partial P(\lambda_i)}{\partial \lambda_i} \log P(y/\lambda) + \lambda_i P(\lambda_i) (1 + \log P(y/\lambda)) \right. \\ \left. \times \left(\sum_{j=1}^M \frac{y_j p_{ij}}{\sum_{i=1}^N \lambda_i p_{ij}} \right) - \lambda_i P(\lambda_i) (1 + \log P(y/\lambda)) \sum_{j=1}^M p_{ij} \right]_{\lambda_i=\hat{\lambda}_i} = 0 \end{aligned} \quad (10)$$

Since, $P(y/\hat{\lambda}) \neq 0$, the second product term of (10) becomes zero. Rearranging the second term we obtain

$$\begin{aligned} \hat{\lambda}_i = \frac{\lambda_i}{P(\hat{\lambda}_i) (1 + \log P(y/\lambda_i)) \sum_{j=1}^M p_{ij}} \\ \times \left[\frac{\partial P(\lambda_i)}{\partial \lambda_i} \Big|_{\lambda_i=\hat{\lambda}_i} \log P(y/\hat{\lambda}_i) \right. \\ \left. + P(\hat{\lambda}_i) (1 + \log P(y/\hat{\lambda}_i)) \sum_{j=1}^M \frac{y_j p_{ij}}{\sum_{i=1}^N \hat{\lambda}_i p_{ij}} \right] \end{aligned} \quad (11)$$

A closed form formula for solving (11) does not exist; it is necessary to devise schemes which will estimate a solution iteratively. Heuristically, a simple iterative algorithm for solving (11) can be built by considering $\hat{\lambda}_i^{k+1}$ on the left-hand side of (11) as the new estimate and $\hat{\lambda}_i^k$ on the right-hand side as the old estimate.

$$\begin{aligned} \hat{\lambda}_i^{k+1} = \frac{\hat{\lambda}_i^k}{P(\hat{\lambda}_i^k) (1 + \log P(y/\hat{\lambda}_i^k)) \sum_{j=1}^M p_{ij}} \\ \times \left[\frac{\partial P(\lambda_i^k)}{\partial \lambda_i^k} \Big|_{\lambda_i=\hat{\lambda}_i^k} \log P(y/\hat{\lambda}_i^k) \right. \\ \left. + P(\hat{\lambda}_i^k) (1 + \log P(y/\hat{\lambda}_i^k)) \left(\sum_{j=1}^M \frac{y_j p_{ij}}{\sum_{i=1}^N \hat{\lambda}_i^k p_{ij}} \right) \right] \end{aligned} \quad (12)$$

We call (12) the entropy maximisation algorithm for PET, which recursively produces a series of reconstructed images that converge to the maximum entropy estimate.

It is instructive to see that as $P(\hat{\lambda}_i^k)$ tends to uniform distribution, the entropy maximisation algorithm becomes the ML algorithm

$$\hat{\lambda}_i^{k+1} = \frac{\hat{\lambda}_i^k}{\sum_{j=1}^M p_{ij}} \sum_{i=1}^N \frac{y_j p_{ij}}{\hat{\lambda}_i^k p_{ij}} \quad (13)$$

In other words, the entropy maximisation algorithm tends to the maximum likelihood algorithm when the prior distribution tends to a uniform distribution because a uniformly distributed prior gives equal probability for any estimate to be a solution.

Further simplification of (12) is possible if we make the following assumption:

$$1 + \log P(y/\hat{\lambda}_i^k) \approx \log P(y/\hat{\lambda}_i^k)$$

This assumption is justifiable on the basis that $P(y/\hat{\lambda})$ is a probability distribution bounded between 0 and 1. In a PET system $P(y/\hat{\lambda})$ has very small positive values, much smaller than 1. Hence, the assumption is valid whenever $P(y/\hat{\lambda}) \ll 1$.

With this assumption, (12) simplifies to the following simple iterative procedure for updating the estimates:

$$\hat{\lambda}_i^{k+1} = \frac{\hat{\lambda}_i^k}{P(\hat{\lambda}_i^k) \sum_{j=1}^M p_{ij}} \left[\frac{\partial P(\lambda_i)}{\partial \lambda_i} \Big|_{\lambda_i=\hat{\lambda}_i^k} + P(\hat{\lambda}_i^k) \sum_{j=1}^M \frac{y_j p_{ij}}{\sum_{i=1}^N \hat{\lambda}_i^k p_{ij}} \right] \quad (14)$$

For the rest of the paper we use the above simplified version of the proposed algorithm. It can be seen that this is slightly different from Green's MAP algorithm [36]. We cannot give a general proof for the convergence of the above algorithm as this will depend on the form of the prior function P .

3 Modelling prior distribution

The next step in the reconstruction process is the proper modelling of the prior distribution function, which incorporates pixel-pixel interaction in the reconstructed image. It is found that when the prior tends to a uniform distribution, the proposed algorithm tends to the ML algorithm. Only when the prior is chosen to be physically more accurate and meaningful, can one expect to obtain improved reconstruction over the ML approach.

One such property of an image is nearest neighbour interaction between the pixels. The Markov random field (MRF) has been extensively used to incorporate such interactions in an image lattice [37, 38]. The MRF defined on a lattice is the collection of random variables λ_j , corresponding to the site λ_i such that

$$P(\lambda_i | \lambda_j; i \neq j) = P(\lambda_i | \lambda_j; j \in N_i) \quad (15)$$

where N_i represents the subset of lattice sites in the neighborhood of the i th lattice site. The Hammersly-Clifford theorem [39] states that a random field over a lattice is a MRF if and only if its distribution function corresponds to a Gibbs distribution function and *vice versa*. So, a MRF model of image λ characterised by a Gibbs distribution is given by

$$P(\lambda) = \frac{1}{Z} e^{-\frac{1}{\beta} U(\lambda)} \quad (16)$$

where Z is the normalising constant for the distribution, β is the Gibbs hyper-parameter, and $U(\lambda)$ is the Gibbs energy.

For homogeneous MRF, the energy function has the following form:

$$U(\lambda) = \sum_i \sum_{j \in N_i} V(\lambda_i, \lambda_j) \quad (17)$$

where N_i are the nearest neighbours of pixel i , and $V(\lambda_i, \lambda_j)$ is termed the potential at the site containing the subset N_i .

Although a large number of potential functions have been suggested in the literature [37, 38, 40], we have used a simple quadratic potential for evaluation of the proposed algorithm. Consider the following potential function first suggested by Geman and Geman [41]:

$$V(\lambda_i, \lambda_j) = (\lambda_i - \lambda_j)^2 \quad (18)$$

The potential $V(\lambda_i, \lambda_j)$ utilises the correlation between neighbouring pixels and forces them to be of similar value. This is justifiable because we expect neighbouring pixels not to vary drastically in an image field.

It should be noted that the first term in (14) corresponds to the correction term due to the nearest neighbour contribution if the potential given by (18) is assumed, i.e.

$$I(\lambda_i^k) = \frac{\hat{\lambda}_i^k}{P(\hat{\lambda}_i^k) \sum_{j=1}^M p_{ij}} \left[\frac{\partial P(\lambda_i)}{\partial \lambda_i} \Big|_{\lambda_i = \hat{\lambda}_i^k} \right] \quad (19)$$

with $P(\lambda)$ given by (16). The second term of (14) is essentially the maximum likelihood contribution. For the simple potential defined by (18), the first correction term becomes

$$I(\lambda_i^k) = \frac{\hat{\lambda}_i^k}{P(\hat{\lambda}_i^k) \sum_{j=1}^M p_{ij}} \frac{2}{\beta} \sum_i \sum_{j \in N_i} (\hat{\lambda}_i^k - \hat{\lambda}_j^k) \quad (20)$$

As, the iteration increases, the numerator decreases since the value of $\hat{\lambda}_i$ slowly approaches the value of $\hat{\lambda}_j$ and in the limit $k \rightarrow \infty$, $I(\lambda_i^k) \rightarrow 0$. This is due to the smoothing nature of the potential. The contribution of $I(\lambda_i^k)$ is from $\sum_{j \in N_i} (\hat{\lambda}_i - \hat{\lambda}_j)$ and β . For small values of β , the contribution is more, whereas for large β it is essentially ML. So, we need to search for an appropriate value of β . It has been found that for $\beta > 2.5 \times 10^4$, the proposed algorithm converges.

4 Simulated experimental results

Implementation of the proposed algorithm was performed on the simulated PET system shown in Fig. 1. The PET system consists of a ring detector system with 64 detectors and the object space is decomposed into 64×64 square pixels. The object space is a square region inscribed within the circle of detectors. To exploit the eight-fold symmetry we have assumed the object to be bounded by a circle inscribed in this square object space [15, 16]. An electron–positron annihilation event occurring inside a pixel is governed by a number of physical phenomena such as attenuation, scattering, absorption, and detector characteristics. All these physical processes have a bearing on the probability matrix. Each element of the probability matrix defines the probability of a photon being detected in the detector j after emanating from the object pixel i . For simplicity, we assume that the probability of emission in pixel i that is detected in a detector tube j is proportional to the angle θ_{ij} seen by the centre of the pixel i into the detector tube j , i.e. $p_{ij} = \theta_{ij}/\pi$ (Fig. 2) [3, 4]. Before the reconstruction begins, the probability matrix $\mathbf{Q} = [p_{ij}]$, $i = 1, \dots, N$ and $j = 1, \dots, M$ is computed and stored. It is found that using the symmetry available in the system the data size of the $\mathbf{Q}$ matrix can be reduced considerably [16].

Fig. 1 Typical PET system with detector ring surrounding the head

γ -rays emanating from the pixels (shown as white squares) of the head phantom are subsequently detected

Fig. 2 Pixel-detector probabilities (p_{ij}^l) for simulated PET system computed from angle of view of the centre of pixel i into the tube j , ($p_{ij} = \theta_{ij}/\pi$)

For simulating measurement data, a Monte Carlo procedure is used in which each emission is simulated as follows [3, 15]. First, a random pixel is chosen in the test phantom. The concentration of the radionuclei at the given pixel is assumed to be proportional to the emission density of the pixel. For each of the accepted emission points, a randomly oriented line (between 0 and π radians) is selected and the pair of detectors this line intersects is found. The random line corresponds to the direction in which the pair of annihilated photons travel and the pixel corresponds to the annihilation point in the phantom. The detectors which this line intersects are assumed to detect this annihilation event and the count in the corresponding detector tube is incremented. In this way all emissions are simulated and counted in the respective tubes, and used as the measurement data. A Shepp–Logan mathematical phantom with 100 000 emissions has been used for simulation studies.

For iterative update of the estimates, the algorithm needs repeated passage from data space to image space and *vice versa* through the operator $\mathbf{P}$. So, it is convenient to write (14) in an additive form i.e.

$$\hat{\lambda}^{k+1} = \hat{\lambda}^k + \Delta(\hat{\lambda}^k) \quad (21)$$

Starting with a small positive initial value of the estimate, (21) updates the old estimate $\hat{\lambda}^k$, by adding the correction factor $\Delta(\hat{\lambda}^k)$ to produce the new estimate $\hat{\lambda}^{k+1}$.

The mean squared error (MSE) between the original Shepp–Logan phantom and the reconstructed phantom is considered to evaluate the performance of the proposed algorithm. The plot of MSE in Fig. 3 illustrates the general nature of the proposed algorithm for various β values. It is clear from the plot that only for restricted values of β does the algorithm produce superior quality reconstruction. We term this interval of β , the optimum interval. The image quality is found to be degraded outside this optimal interval, irrespective of the number of iterations performed. The reason for this is the smoothing behaviour of the potential. This potential increasingly penalises the difference between neighbouring pixels, thus forcing them to become the same.

In terms of entropy, it is clear that with increasing β the information content decreases, thereby resulting in

Fig. 3 Plot of mean square error against hyper-parameter β shown after different numbers of iterations

a 40 iterations
b 100 iterations

saturation of the mean square error (MSE) as shown in Fig. 3. On the other hand if β is in the optimal interval, then there is an increase in the information content gained at the cost of the local correlation invoked in the reconstruction procedure via prior knowledge. In the limit $\beta \rightarrow \text{large value}$, the maximum entropy algorithm approaches the ML algorithm. Below the optimal interval, the estimate becomes unstable due to the dominating nature of the prior. This basically removes the contribution of measurement data in the reconstruction algorithm. These can be visually verified from the reconstruction shown in Fig. 6. Once β is chosen in the optimal interval, estimates will produce better reconstruction as the iterations continue. This is evident from the lower MSE of the reconstructed images as shown in Fig. 3 for 40 and 100 iterations.

In Fig. 4, plots of the MSE against the number of iterations are shown with varying β values from 1.5×10^4 to 2.5×10^7 . The MSE is found to decrease with increasing iterations for $\beta = 1.5 \times 10^4$ to 2.5×10^5 (see Fig. 4). On the other hand, MSE increases with increasing iterations for $\beta \geq 2.5 \times 10^5$ (see Fig. 4k and Fig. 4l). The β value indicates the amount of prior to be introduced into the reconstruction procedure. When β is increased beyond a certain large value, then the reconstruction is effectively due to the likelihood contribution; the general nature of which is depicted in Fig. 4k and Fig. 4l. It is well known that the reconstruction improves during the earlier iterations of ML estimation, but becomes noisier as the iterations continued [8].

Fig. 4 Plots of mean square error against iterations

a $\beta = 1.5 \times 10^4$
b $\beta = 2.5 \times 10^4$
c $\beta = 3.5 \times 10^4$
d $\beta = 4.5 \times 10^4$
e $\beta = 5.5 \times 10^4$
f $\beta = 6.5 \times 10^4$
g $\beta = 7.5 \times 10^4$
h $\beta = 1.0 \times 10^5$
i $\beta = 2.0 \times 10^5$
j $\beta = 2.5 \times 10^5$
k $\beta = 2.5 \times 10^6$
l $\beta = 2.5 \times 10^7$

Fig. 5 Original 64×64 Shepp-Logan mathematical phantom

Fig. 6 Visual representation of image reconstruction after 100 iterations of entropy maximisation algorithm for varying β values

- a $\beta = 1.5 \times 10^4$
- b $\beta = 2.5 \times 10^4$
- c $\beta = 3.5 \times 10^4$
- d $\beta = 4.5 \times 10^4$
- e $\beta = 5.5 \times 10^4$
- f $\beta = 6.5 \times 10^4$
- g $\beta = 7.5 \times 10^4$
- h $\beta = 1.0 \times 10^5$
- i $\beta = 2.0 \times 10^5$
- j $\beta = 2.5 \times 10^5$
- k $\beta = 2.5 \times 10^6$
- l $\beta = 2.5 \times 10^7$

It is evident from the proposed algorithm that with an increase in the β value, the effect of the prior is nullified by the data and the maximum entropy estimation becomes equivalent to the EM estimation. This can be further seen in the reconstruction shown in Figs. 5 and 6. With increasing β values, the image makes a transition from smooth reconstruction to non-smooth. This non-smoothness in the reconstruction is an indicator of noise. It is clear from the reconstruction shown in Fig. 6 that noise artefacts are much more severe in Fig. 6k and Fig. 6l than Figs. 6a–6j, which are the reconstructions produced by β values in the

Fig. 7 Mean squared error against number of iterations showing the change in MSE for ML, MAP with different hyper-parameter values, and the proposed algorithm

optimal interval. Good reconstruction is found to occur for $2.5 \times 10^4 \leq \beta \leq 2.0 \times 10^5$. It is found that the entropy maximisation estimate becomes unstable for $\beta < 10^4$. The reconstructed images become increasingly noisy for large β values. This can be visually verified in the reconstructions shown in Fig. 6.

We also compare the results of the proposed algorithm with the well known maximum *a posteriori* (MAP) estimation [36, 40]. The potential used for the MAP estimation algorithm [36] and the proposed algorithm is the same (i.e. (18)) so as to compare the performance. Simulations were carried out for Gibbs hyper-parameter β ranging from 10^3 to 10^7 . For brevity, we have shown MSE against iterations plots for some of the key β values in Fig. 7. Comparing MSE values for the proposed algorithm and the MAP algorithm (Fig. 7), it is clear that the proposed algorithm performs slightly better than the MAP algorithm. It should be noted that more realistic potentials are capable of producing better estimates but require additional parameter estimation [37, 38]. The second potential used for reconstruction is similar to the Huber function:

$$V_H = \begin{cases} \frac{1}{2\mu}(\lambda_i - \lambda_j)^2, & \text{if } |\lambda_i - \lambda_j| \leq \mu \\ \mu|\lambda_i - \lambda_j|, & \text{otherwise} \end{cases} \quad (22)$$

where μ is an additional parameter to be chosen. Simulations are carried out using (22) as the potential and the results are shown in Fig. 8. The first row and second row correspond to $\mu = 1, 10, 25$ and 50 for MAP and the proposed algorithm, respectively. It is clear that quality decreases as μ increases because for small pixel differences, the penalty is quadratic while the penalty is linear when the differences are large.

Figure 9 shows images reconstructed using ML, MAP and the proposed algorithm with potential V (18), after 80 iterations. For visual inspection and quality of reconstruction the original test phantom is also shown in Fig. 9a. The ML reconstructed image is shown in Fig. 9b. The first and second row show reconstructed images for different Gibbs hyper-parameter values using MAP and the proposed algorithm, respectively. It can be seen that the images reconstructed using the proposed algorithm look sharper than the images reconstructed using the MAP algorithm.

Fig. 8 Reconstructions obtained using MAP algorithm and the proposed algorithm

- a* MAP, $\mu = 1$
- b* MAP, $\mu = 10$
- c* MAP, $\mu = 25$
- d* MAP, $\mu = 50$
- e* Proposed algorithm, $\mu = 1$
- f* Proposed algorithm, $\mu = 10$
- g* Proposed algorithm, $\mu = 25$
- h* Proposed algorithm, $\mu = 50$

Fig. 9 Images reconstructed after 80 iterations

- a* Original phantom
- b* Reconstructed by ML algorithm
- c* Reconstructed by MAP algorithm ($\beta = 10^4$)
- d* Reconstructed by MAP algorithm ($\beta = 10^5$)
- e* Reconstructed by MAP algorithm ($\beta = 10^7$)
- f* Reconstructed by proposed algorithm ($\beta = 10^4$)
- g* Reconstructed by proposed algorithm ($\beta = 10^5$)
- h* Reconstructed by proposed algorithm ($\beta = 10^7$)

5 Conclusions

A new algorithm is proposed for the PET system, which is capable of good reconstruction. This is based on the

maximisation of conditional entropy as a cost function rather than the likelihood or posterior function. The proposed method ensures positivity of the reconstructed image pixels since entropy is not defined for negative pixels.

Furthermore, the proposed methodology enables the inclusion of prior distribution knowledge of the object during the image reconstruction process.

The Gibbs distribution function is used for modelling the prior. The attractive feature of a Gibbs distribution prior is its ability to incorporate local correlation between neighbouring pixels in an image lattice. Simulated experimental results have shown much improved reconstruction for a broad range of Gibbs hyper-parameter values. It is shown that entropy is a better candidate as a cost function than the likelihood or posterior function.

6 Acknowledgments

The authors wish to thank the anonymous reviewers, who helped in improving the quality of this manuscript. The first author would like to thank Mr Hemanth, research scholar at the Indian Institute of Science for helping during the simulation experiment. Also, he wishes thanks to the Council of Scientific and Industrial research (CSIR), Government of India, for providing a Junior Research fellowship. He acknowledges that this work is partially supported by CSIR, New Delhi, India.

7 References

- Frieden, B.R.: 'Restoring with maximum likelihood and maximum entropy', *J. Opt. Soc. Am.*, 1972, **62**, pp. 511–518
- Frieden, B.R., and Wells, D.C.: 'Restoring with maximum entropy: Poisson sources and backgrounds', *J. Opt. Soc. Am.*, 1978, **68**, (1), pp. 93–103
- Vardi, Y., Shepp, L.A., and Kaufmann, L.: 'A statistical model for positron emission tomography', *J. Am. Stat. Assoc.*, 1985, **80**, pp. 8–37
- Shepp, L.A., and Vardi, Y.: 'Maximum likelihood estimation for emission tomography', *IEEE Trans. Med. Imaging*, 1982, **1**, pp. 113–121
- Craig, A.D., Reiman, E.M., Evans, A., and Bushnell, M.C.: 'Functional imaging of an illusion of pain', *Nature*, 1996, **384**, pp. 217–218
- Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., Woods, R., Mazziotta, J.C., and Fazio, F.: 'Mapping motor representations with positron emission tomography', *Nature*, 1994, **371**, pp. 600–602
- Chernoboy, E.S., Chen, C.J., Miller, M.I., Miller, T.R., and Snyder, D.L.: 'An evaluation of maximum likelihood reconstruction for SPECT', *IEEE Trans. Med. Imaging*, 1990, **9**, pp. 99–110
- Veclerov, E., and Llacer, J.: 'Stopping rule for MLE algorithm based on statistical hypothesis testing', *IEEE Trans. Med. Imaging*, 1987, **6**, pp. 313–319
- Chen, C.M., and Lee, S.Y.: 'A parallel implementation of 3-D CT image reconstruction on hypercube multi processor', *IEEE Trans. Nucl. Sci.*, 1990, **37**, (3), pp. 1333–1346
- Chen, C.M., and Lee, S.Y.: 'Parallelization of the EM algorithm for 3-D PET image reconstruction', *IEEE Trans. Med. Imaging*, 1991, **10**, (4), pp. 513–522
- Rajan, K., Patnaik, L.M., and Ramakrishna, J.: 'High speed computation of the EM algorithm for PET image reconstruction', *IEEE Trans. Nucl. Sci.*, 1994, **41**, (5), pp. 1–5
- Ranganath, M.V., Dhawan, A.P., and Mullani, N.: 'A multigrid expectation maximization algorithm for positron emission tomography', *IEEE Trans. Med. Imaging*, 1988, **7**, pp. 273–278
- Tanaka, E.: 'A fast reconstruction algorithm for stationary positron emission tomography based on a modified EM algorithm', *IEEE Trans. Med. Imaging*, 1987, **3**, (1), pp. 98–105
- Lange, K., Bahn, M., and Little, R.: 'A theoretical study of some maximum likelihood algorithm for emission and transmission tomography', *IEEE Trans. Med. Imaging*, 1987, **6**, (2), pp. 106–114
- Rajeevan, N., Rajgopal, K., and Krishna, G.: 'Vector-extrapolated fast maximum likelihood estimation algorithms for emission tomography', *IEEE Trans. Med. Imaging*, 1992, **11**, (1), pp. 9–20
- Kaufmann, L.: 'Implementing and accelerating the EM-algorithm for positron emission tomography', *IEEE Trans. Med. Imaging*, 1987, **6**, pp. 37–51
- Hudson, H.M., and Larkin, R.S.: 'Accelerated image reconstruction using ordered subsets of projection data', *IEEE Trans. Med. Imaging*, 1994, **13**, (4), pp. 601–609
- Hudson, H.M., Hutton, B.F., and Larkin, R.: 'Accelerated EM reconstruction using ordered subsets', *J. Nucl. Med.*, 1992, **33**, p. 960
- Browne, J., and De Pierro, A.: 'A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography', *IEEE Trans. Med. Imaging*, 1996, **15**, pp. 687–699
- Byrne, C.L.: 'Accelerating EMM algorithm and related iterative algorithm by rescaled block iterative methods', *IEEE Trans. Image Process.*, 1998, **7**, pp. 100–109
- Lalush, D.S., Frey, E.C., and Tsui, B.M.W.: 'Fast maximum entropy approximation in SPECT using RBI-MAP algorithm', *IEEE Trans. Med. Imaging*, 2000, **19**, pp. 286–294
- Snyder, D.L., and Miller, M.I.: 'The use of sieves to stabilize images produced with the EM-algorithm for emission tomography', *IEEE Trans. Nucl. Sci.*, 1985, **32**, (5), pp. 3864–3872
- Hebert, T., and Leahy, R.: 'A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors', *IEEE Trans. Med. Imaging*, 1989, **8**, (2), pp. 194–202
- Levitin, E., and Herman, G.T.: 'A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography', *IEEE Trans. Med. Imaging*, 1987, **6**, (3), pp. 185–192
- Mondal, P.P., and Rajan, K.: 'Entropy maximization algorithm for positron emission tomography'. Presented at 9th Int. Conf. on Neural Information Processing, Singapore, 18–22 Nov. 2002
- Mondal, P.P., and Rajan, K.: 'Conditional entropy maximization algorithm for PET'. Presented at IEEE Int. Conf. on Systems, Man and Cybernetics, Washington DC, USA, 5–8 October 2003
- Skilling, J., Strong, A.W., and Bennett, K.: 'Use of maximum entropy method in gamma-ray astronomy', *Mon. Not. R. Astron. Soc.*, 1979, **187**, pp. 145–152
- Skilling, J., and Bryan, R.K.: 'Maximum entropy image reconstruction: general algorithm', *Mon. Not. R. Astron. Soc.*, 1984, **211**, pp. 111–124
- Gull, S.F., and Daniel, G.J.: 'Image reconstruction from incomplete and noisy data', *Nature*, 1978, **272**, pp. 686–690
- Minerbo, G.: 'MENT: A maximum entropy algorithm for reconstructing a source from projection data', *Comput. Graph. Image Process.*, 1979, **10**, pp. 48–68
- Willingale, R.: 'Use of maximum entropy method in X-ray astronomy', *Mon. Not. R. Astron. Soc.*, 1981, **194**, pp. 359–364
- Byrne, C.L.: 'Likelihood maximization for list-mode emission tomographic image reconstruction', *IEEE Trans. Med. Imaging*, 2001, **20**, (10), pp. 1084–1092
- Byrne, C.L.: 'Iterative image reconstruction algorithms based on cross-entropy minimization', *IEEE Trans. Image Process.*, 1993, **2**, (1), pp. 96–103
- Titterton, D.: 'On the iterative image space reconstruction algorithm for ECT', *IEEE Trans. Med. Imaging*, 1987, **6**, pp. 52–56
- Zangwill, W.I.: 'Nonlinear programming: a unified approach' (Prentice Hall, Englewood Cliffs, NJ, USA)
- Green, P.J.: 'Bayesian reconstruction from emission tomography data using a modified EM algorithm', *IEEE Trans. Med. Imaging*, 1990, **9**, (1), pp. 84–93
- Zhou, Z., Leahy, R.M., and Qi, J.: 'Approximate maximum likelihood hyperparameter estimation for Gibbs prior', *IEEE Trans. Image Process.*, 1997, **6**, (6), pp. 844–861
- Hebert, T., and Leahy, R.: 'Statistic based MAP image reconstruction from Poisson data using Gibbs Priors', *IEEE Trans. Signal Process.*, 1992, **40**, (2), pp. 2290–2303
- Besag, J.: 'Spatial interaction and the statistical analysis of lattice systems', *J. R. Stat. Soc. B*, 1974, **36**, pp. 192–236
- Nuyts, J., Bequ, D., Dupont, P., and Mortelmans, L.: 'A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography', *IEEE Trans. Nucl. Sci.*, 2002, **49**, (1), pp. 56–60
- Geman, S., and Geman, D.: 'Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1984, **6**, pp. 721–741