Information Processing in Medical Imaging, H. Barrett and A. Gmitro, editors, LNCS 687, pp. 322-338, Springer-Verlag, 1993

Bayesian Reconstruction for Emission Tomography via Deterministic Annealing

G. Gindi[¶#], A. Rangarajan[†§], M. Lee[‡], P. J. Hong[#], and I. G. Zubal[†]

[‡]Department of Electrical Engineering, Yale University, New Haven, CT 06520-2157
[†]Department of Diagnostic Radiology, Yale University, New Haven, CT 06510-3333
[§]Department of Computer Science, Yale University, New Haven, CT 06520-2158
[#]Department of Electrical Engineering, SUNY, Stony Brook, NY 11794-2350
[¶]Department of Radiology, SUNY, Stony Brook, NY 11794-8460

Abstract. In emission tomography, a principled means of incorporating a piecewise smooth prior on the source $\mathbf{f}$ is via a mixed variable objective function $E(\mathbf{f},\mathbf{l})$ defined on $\mathbf{f}$ and binary valued line processes $\mathbf{l}$. MAP estimation on $E(\mathbf{f},\mathbf{l})$ results in the difficult problem of minimizing an objective function that includes a nonsmooth Gibbs prior ϕ^* defined on the spatial derivatives of $\mathbf{f}$. Previous approaches have used heuristic Gibbs potentials ϕ that incorporate line processes, but only approximately. In this work, we present a *continuation* method in which the correct function ϕ^* is approached through a *sequence* of smooth ϕ functions. Our continuation method is implemented using a GEM–ICM procedure. Simulation results show improvement using our continuation method relative to using ϕ^* alone, and to conventional EM reconstructions. Finally, we show a means of generalizing this formalism to the less restrictive case of piecewise linear instead of piecewise flat priors.

1 Introduction

Bayesian reconstruction methods for emission tomography have been a topic of some interest in recent years. These methods allow for the introduction of prior information into the reconstruction problem thereby extending maximum likelihood (ML) approaches via maximum *a posteriori* (MAP) approaches. MAP solutions are usually found by minimization of an appropriate objective (energy) function. Many of the well-known difficulties with MAP reconstruction stem from essentially technical problems in the minimization of the MAP objective function. One of these problems is associated with the minimization of objective functions composed of both binary and continuous variables. The second problem concerns local minima. We offer a framework – deterministic annealing – that can potentially address both problems. In this paper, we focus on the first of these problems.

In the case of emission tomography, the prior distribution reflects assumptions regarding the nature of the underlying source distribution. In this paper, the source distribution is assumed to be piecewise smooth; in other words, the image consists mostly of smooth regions separated by edges or discontinuities.

We model this kind of prior via the use of binary-valued line processes [1] which are unobservables corresponding to image discontinuities. At locations where the line processes are "off", the prior encourages smoothness by penalizing the formation of high image-intensity gradients; at locations where the line processes are "on", this penalty is removed and high intensity gradients are permitted.

The prior consists of both continuous variables (the source intensities) and binary variables (the line processes), and we are interested in simultaneously estimating them. This estimation is a difficult problem. Unlike other methods, the deterministic annealing method offers a principled and efficient means of handling the problems associated with mixed continuous and binary variable objectives.

The application of the deterministic annealing method results in a *sequence* of objective functions (defined only on the continuous variables) whose sequence of solutions approaches that of the original mixed variable objective function. The sequence is indexed by a control parameter β ($\beta = \frac{1}{T}$, where T is the temperature). At each temperature, a standard descent optimization algorithm is used to find a solution which is then used as an initial condition for the next temperature setting. The energy functions at high temperatures are smooth approximations of the energy functions at lower temperatures. Consequently, it is easier to minimize the energy functions at high temperatures and then track the minimum through the variation of the temperature. This is the essence of a *continuation* method. In other domains such as computer vision, where some problems can be formulated as Bayesian estimation, continuation methods have been successfully used [2, 3] in finding good solutions.

2 A Bayesian Model

We first formulate our problem with the aid of binary-valued $\{0, 1\}$ line processes as proposed by Geman and Geman [1], though we shall eventually dispense with them. We may now write Bayes theorem:

$$\Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l} | \mathbf{G} = \mathbf{g}) = \frac{\Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l}) \Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})}{\Pr(\mathbf{G} = \mathbf{g})} \tag{1}$$

where $\mathbf{f}$, $\mathbf{l}$, and $\mathbf{g}$ are the source intensities, line processes, and projection data respectively, and $\mathbf{F}$, $\mathbf{L}$, $\mathbf{G}$ are the associated random fields. Given the posterior distribution in (1), one possible estimator is the MAP estimate.

$$(\hat{\mathbf{f}}, \hat{\mathbf{l}}) = \arg \max_{(\mathbf{f}, \mathbf{l})} \Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}) \Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l}). \tag{2}$$

The two terms on the right side of (2) are the likelihood and prior which we now describe.

2.1 The "Correct" Energy Function

We model both the likelihood and prior as Gibbs distributions. A Gibbs distribution is specified by an energy function E and the partition function Z

(a suitable normalization).

$$\Pr(\mathbf{Y}=\mathbf{y}|\mathbf{X}=\mathbf{x}) = \frac{1}{Z(\mathbf{x})}\exp(-E(\mathbf{y})). \tag{3}$$

Since the partition function is a normalization over all possible configurations of $\mathbf{Y}$ and not of $\mathbf{X}$, it is usually a function of $\mathbf{x}$.

We describe the process by which an energy function can be obtained from the two terms in (2) constituting the posterior distribution. We model the likelihood in a conventional way. Since the dominant noise source is photon noise, we use the well known fact that

$$\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f}) = \prod_{t\theta}\frac{\bar{g}_{t\theta}^{g_{t\theta}}\exp(-\bar{g}_{t\theta})}{g_{t\theta}!} \tag{4}$$

where $\bar{g}_{t\theta} \stackrel{\text{def}}{=} \sum_{ij}\mathcal{H}_{t\theta,ij}f_{ij}$. In (4), $g_{t\theta}$ is the number of counts and $\bar{g}_{t\theta}$ is the expected number of counts for a particular source $\mathbf{f}$ and a forward projection matrix $\mathcal{H}_{t\theta,ij}$. This projection is from source location (i,j) to detector bin t at angle θ. Using our definition of an energy function E,

$$\begin{aligned} E_D(\mathbf{f}) &\stackrel{\text{def}}{=} -\log\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f}) \\ &= E_L(\mathbf{f}) + \log(Z_L(\mathbf{f})) \end{aligned}$$

where

$$E_L(\mathbf{f}) = \sum_{t\theta}\left[-g_{t\theta}\log(\bar{g}_{t\theta}) + \log(g_{t\theta}!)\right] \tag{5}$$

and

$$\log(Z_L(\mathbf{f})) = \sum_{t\theta}\bar{g}_{t\theta}.$$

The prior is defined over the source intensities and the line processes. The corresponding energy function is

$$\begin{aligned} &E_P(\mathbf{f},\mathbf{l}) = -\log\Pr(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l}) \\ &\quad = \textstyle\sum_{ij}\left[f_v^2(i,j)(1-l_{ij}^h) + \alpha l_{ij}^h + f_h^2(i,j)(1-l_{ij}^v) + \alpha l_{ij}^v\right] \end{aligned}$$

where $f_v(i,j)$ and $f_h(i,j)$ are the vertical and horizontal source intensity gradients, respectively. The two terms involving f_v and f_h encourage smoothness except where discontinuities ($l_{ij}=1$) occur and include a penalty α for the creation of discontinuities. The combination of the smoothness prior on the intensities and the penalty term introduced above (E_P) is the *weak membrane* prior used in the computer vision literature [2] and [3]. This prior favors the formation of piecewise flat reconstructions.

We may thus write the overall energy function as

$$\begin{aligned} E(\mathbf{f},\mathbf{l}) &= -\log\Pr(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l}|\mathbf{G}=\mathbf{g}) \\ &= E_L(\mathbf{f}) + \log(Z_L(\mathbf{f})) + \lambda E_P(\mathbf{f},\mathbf{l}) \end{aligned} \tag{6}$$

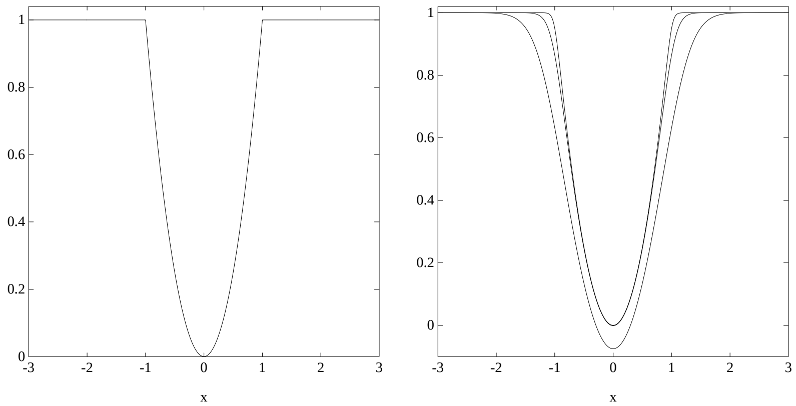

Figure 1: (a) Left: $\phi^*(x)$ (b) Right: $\phi_\beta(x)$

where λ weights the prior E_P with respect to the likelihood energy.

It is difficult to minimize the mixed, binary and continuous variable energy function in (6). However, the search over configurations in the weak membrane can be reduced to a search over only the $\mathbf{f}$. The new energy function is [4]

$$E(\mathbf{f}) = \sum_{t\theta} \left[-g_{t\theta} \log(\bar{g}_{t\theta}) + \bar{g}_{t\theta}\right] + \lambda \sum_{ij} \left[\phi^*(f_v(i,j)) + \phi^*(f_h(i,j))\right] \qquad (7)$$

where we have dropped the term $\log(g_{t\theta}!)$ from $E_L(\mathbf{f})$ in (5) since it does not involve $\mathbf{f}$ and

$$\phi^*(f_v(i,j)) = \begin{cases} f_v^2(i,j) & f_v^2(i,j) \leq \alpha \\ \alpha & f_v^2(i,j) > \alpha \end{cases} \qquad (8)$$

with a corresponding expression for $\phi^*(f_h(i,j))$. (For convenience, we write ϕ^* and related functions in terms of $f_v(i,j)$ only – the dependence on $f_h(i,j)$ should then be obvious). Figure 1(a) shows $\phi^*(f_v(i,j))$ as a "broken parabola". Though the global minima of (7) are the same as those of (6) and the use of binary line variables have been circumvented, there are problems associated with minimizing an energy function consisting of the likelihood and the broken parabola. The broken parabola makes the energy function non-convex and non-differentiable, thereby ruling out gradient-based descent methods.

2.2 The Deterministic Annealing Energy Function

In this section, we show that the broken parabola can be approached by a sequence of associated ϕ functions. The sequence approaches ϕ^* as a parameter β approaches infinity. The sequence can be obtained by first considering a transformation of the probability distributions in (3) to

$$\Pr_\beta(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) = \frac{1}{Z(\mathbf{x};\beta)} \exp(-\beta E(\mathbf{y})).$$

This new sequence of probability distributions, indexed by the real parameter β ($\beta > 0$ is the inverse temperature), is clearly related to the distributions in (3). As β approaches infinity, this new distribution becomes a highly peaked version of the distribution in (3). As far as MAP estimation is concerned, the location of the maximum is unchanged. In our deterministic annealing framework, β ranges from low to high values according to an annealing schedule.

The broken parabola was obtained by eliminating the line process from the original objective function. The line process can also be "eliminated" from the posterior distribution $\Pr_\beta(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l}|\mathbf{G}=\mathbf{g})$. First, the marginal posterior distribution of the intensities is obtained by *integrating out* the line process from the Gibbs distribution. Since the prior does not involve the data $\mathbf{g}$, the partitioning of the posterior distribution can be written as

$$\Pr_\beta(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l}|\mathbf{G}=\mathbf{g}) = \Pr_\beta(\mathbf{F}=\mathbf{f}|\mathbf{G}=\mathbf{g})\Pr_\beta(\mathbf{L}=\mathbf{l}|\mathbf{F}=\mathbf{f})$$

where $\Pr_\beta(\mathbf{F}=\mathbf{f}|\mathbf{G}=\mathbf{g})$ is the marginal posterior of $\mathbf{F}$. The energy function arising out of the marginal posterior of $\mathbf{F}$ is a function of $\mathbf{f}$ alone and not of $\mathbf{l}$. We now show that this function contains a smooth Gibbs prior that approaches the broken parabola as the parameter β is increased. First, we evaluate the marginal posterior:

$$\begin{aligned}
\Pr_\beta(\mathbf{F}=\mathbf{f}|\mathbf{G}=\mathbf{g}) &= \sum_{\{\mathbf{l}\}} \Pr_\beta(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l}|\mathbf{G}=\mathbf{g}) \\
&= \frac{1}{Z_{MP}(\mathbf{g};\beta)} \sum_{\{l^v_{ij}\},\{l^h_{ij}\}} \exp\Big[-\beta\big\{E_L(\mathbf{f}) + \log(Z_L(\mathbf{f})) + \lambda E_P(\mathbf{f},\mathbf{l})\big\}\Big] \\
&= \frac{1}{Z_{MP}(\mathbf{g};\beta)} \exp\Big[-\beta\big\{E_L(\mathbf{f}) + \log(Z_L(\mathbf{f}))\big\}\Big]\Big[\sum_{\{l^v_{ij}\},\{l^h_{ij}\}} \exp\big\{-\beta\lambda E_P(\mathbf{f},\mathbf{l})\big\}\Big]
\end{aligned}$$

where $\sum_{\{l^v_{ij}\},\{l^h_{ij}\}}$ is the summation over all configurations of $\mathbf{L}$. The partition function $Z_{MP}(\mathbf{g};\beta)$ is the partition function (normalization factor) corresponding to the marginal posterior of $\mathbf{F}$ and is therefore not a function of $\mathbf{f}$, but is a function of $\mathbf{g}$ and is parametrized by β. Now

$$\begin{aligned}
&\Pr_\beta(\mathbf{F}=\mathbf{f}|\mathbf{G}=\mathbf{g}) \\
&= \frac{1}{Z_{MP}(\mathbf{g};\beta)} \exp\Big[-\beta\big\{E_L(\mathbf{f}) + \log(Z_L(\mathbf{f}))\big\}\Big]\Big[\sum_{\{l^v_{ij}\},\{l^h_{ij}\}} \exp\big\{-\beta\lambda E_P(\mathbf{f},\mathbf{l})\big\}\Big] \\
&= \frac{1}{Z_{MP}(\mathbf{g};\beta)} \exp\Big[-\beta\sum_{t\theta}(-g_{t\theta}\log(\bar{g}_{t\theta}) + \bar{g}_{t\theta})\Big] \\
&\quad \sum_{\{l^v_{ij}\},\{l^h_{ij}\}} \exp\Big[-\beta\lambda\sum_{ij}\big(f_v^2(i,j)(1-l^h_{ij}) + \alpha l^h_{ij} + f_h^2(i,j)(1-l^v_{ij}) + \alpha l^v_{ij}\big)\Big] \\
&= \frac{1}{Z_{MP}(\mathbf{g};\beta)} \exp\Big[-\beta\Big\{\sum_{t\theta}(-g_{t\theta}\log(\bar{g}_{t\theta}) + \bar{g}_{t\theta})
\end{aligned}$$

$$+\sum_{ij}\frac{-1}{\beta}\log\left[\exp(-\beta\lambda f_v^2(i,j))+\exp(-\beta\lambda\alpha)\right]$$
$$+\sum_{ij}\frac{-1}{\beta}\log\left[\exp(-\beta\lambda f_h^2(i,j))+\exp(-\beta\lambda\alpha)\right]\Big\}\Big].$$

The last step above may be seen with the aid of the following identity:

$$\sum_{l\in\{0,1\}}\exp(-\beta Al)=1+\exp(-\beta A)=\exp\left[-\beta\left\{\frac{-1}{\beta}\log\left(1+\exp(-\beta A)\right)\right\}\right].$$

The smooth Gibbs prior is

$$\phi_\beta(f_v(i,j))\stackrel{\text{def}}{=}-\frac{1}{\beta\lambda}\log(\exp(-\beta\lambda f_v^2(i,j)+\exp(-\beta\lambda\alpha)).$$

At large values of β, if $\alpha > f_v^2(i,j)$, the first term inside the $\log(\cdot)$ expression dominates and the entire expression reduces to $f_v^2(i,j)$. If $\alpha < f_v^2(i,j)$, the second term inside the $\log(\cdot)$ expression dominates and the entire expression reduces to α. Inspection reveals that the ϕ_β function at large β is in fact the broken parabola of (8). At low values of β, the function is very smooth. As can be seen from Figure 1(b), the smooth ϕ_β approaches the broken parabola as β approaches infinity. The energy function associated with the marginal posterior is

$$E(\mathbf{f};\beta)=\sum_{t\theta}\left[-g_{t\theta}\log(\bar{g}_{t\theta})+\bar{g}_{t\theta}\right]+\lambda\sum_{ij}\left[\phi_\beta(f_v(i,j))+\phi_\beta(f_h(i,j))\right].$$

Our task now reduces to finding the minimum at a given β of this energy function. The minimum is to be tracked as β is increased. Despite the absence of the line processes from our final energy function, they are implicitly contained in the broken parabola that is reached at limiting values of the control parameter β.

3 Derivation of the Deterministic Annealing Algorithm

The energy function contains the likelihood energy which precludes a standard gradient descent minimization because of its global connectivity. Instead, we resort to the familiar incomplete/complete data formulation of the Generalized Expectation-Maximization (GEM) approach [5]. The maximization step of the GEM algorithm in our case results in the minimization of

$$M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta)\stackrel{\text{def}}{=}-Q(\mathbf{f}|\hat{\mathbf{f}}^n)+E_{GP}(\mathbf{f};\beta)\tag{9}$$

where

$$Q(\mathbf{f}|\hat{\mathbf{f}}^n)\stackrel{\text{def}}{=}\sum_{ij}\sum_{t\theta}\left[\mathcal{H}_{t\theta,ij}f_{ij}-g_{t\theta}\frac{\mathcal{H}_{t\theta,ij}\hat{f}_{ij}^n}{\sum_{kl}\mathcal{H}_{t\theta,kl}\hat{f}_{kl}^n}\log(f_{ij})\right]$$

and

$$
\begin{aligned}
E_{GP}(\mathbf{f};\beta) &\stackrel{\text{def}}{=} \sum_{ij}\left[-\frac{1}{\beta}\log\left(\exp(-\beta\lambda f_v^2(i,j))+\exp(-\beta\lambda\alpha)\right)\right] \\
&+ \sum_{ij}\left[-\frac{1}{\beta}\log\left(\exp(-\beta\lambda f_h^2(i,j))+\exp(-\beta\lambda\alpha)\right)\right].
\end{aligned}
$$

The original energy function has been transformed via the GEM approach to the energy function $M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta)$ that is to be decreased during each step before returning to the expectation step. Note that the expectation step simply involves substituting $\hat{\mathbf{f}}^{n+1}$ for $\hat{\mathbf{f}}^n$ in $M(\mathbf{f}|\hat{\mathbf{f}}^n)$. The maximization step is to be carried out at each value of β using any standard ascent algorithm.

The approach taken in this paper is to apply the method of Iterated Conditional Modes (ICM) [6], to the above energy function $M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta)$. ICM is a logical choice for the optimization method since it can be used to obtain convenient closed form solutions for the update equation. Step sizes necessary for gradient descent methods are also avoided. ICM minimizes the energy function by coordinate-wise descent. Coordinate-wise descent is performed by minimizing the energy function with respect to f_{ij} while keeping all other source intensities fixed. Then a new location is chosen and the same method is repeated. After a full sweep of the lattice, the procedure is repeated until convergence criteria are met. Then β is increased and ICM is repeated. The closed form solution with respect to each variable can be obtained by differentiating the energy function and solving for the variable of interest. Hence,

$$
\frac{\partial M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta)}{\partial f_{ij}} = -\frac{\partial Q(\mathbf{f}|\hat{\mathbf{f}}^n)}{\partial f_{ij}} + \frac{\partial E_{GP}(\mathbf{f};\beta)}{\partial f_{ij}}
$$

and

$$
-\frac{\partial Q(\mathbf{f}|\hat{\mathbf{f}}^n)}{\partial f_{ij}} = \sum_{t\theta}\mathcal{H}_{t\theta;ij} - \sum_{t\theta} g_{t\theta}\frac{\mathcal{H}_{t\theta;ij}\hat{f}_{ij}^n}{\sum_{kl}\mathcal{H}_{t\theta;kl}\hat{f}_{kl}^n}\frac{1}{f_{ij}}.
$$

Similarly,

$$
\begin{aligned}
&\frac{\partial E_{GP}(\mathbf{f};\beta)}{\partial f_{ij}} = \qquad (10)\\
&-2\lambda f_h(i,j)\tfrac{1}{1+\exp\left[\beta\lambda(f_h^2(i,j)-\alpha)\right]} + 2\lambda f_h(i,j-1)\tfrac{1}{1+\exp\left[\beta\lambda(f_h^2(i,j-1)-\alpha)\right]}\\
&-2\lambda f_v(i,j)\tfrac{1}{1+\exp[\beta\lambda(f_v^2(i,j)-\alpha)]} + 2\lambda f_v(i-1,j)\tfrac{1}{1+\exp[\beta\lambda(f_v^2(i-1,j)-\alpha)]}.
\end{aligned}
$$

We would like to obtain a closed-form solution for f_{ij} from $\frac{\partial M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta)}{\partial f_{ij}} = 0$. However, upon examining (10), it is clear that such an explicit solution is impossible due to the transcendental nature of the exponential term. One possible solution to this problem is to consider the exponential term to be a new, albeit dependent (on f_{ij}), variable and to descend on each such variable separately. The following transformation performs the trick:

$$z^v_{ij} = \frac{1}{1+\exp(-\beta\lambda(f_h^2(i,j)-\alpha))} \tag{11}$$

with a similar expression for z^h_{ij}. With this transformation, $M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta)$ becomes $M(\mathbf{f},\mathbf{z}|\hat{\mathbf{f}}^n;\beta)$. Having done this, we can perform ICM by keeping f_{ij} frozen while descending on z_{ij} and keeping z_{ij} frozen while descending on f_{ij}. Using (16) in the Appendix, the update equations for f_{ij} can be shown to be

$$f_{ij} = \frac{-(\sum_{t\theta}\mathcal{H}_{t\theta;ij} - 2\lambda X_3) + \sqrt{(\sum_{t\theta}\mathcal{H}_{t\theta;ij} - 2\lambda X_3)^2 + 8\lambda X_2 X_1}}{4\lambda X_2} \tag{12}$$

where

$$X_1 \stackrel{\text{def}}{=} \sum_{t\theta} g_{t\theta}\frac{\mathcal{H}_{t\theta,ij}\hat{f}^n_{ij}}{\sum_{kl}\mathcal{H}_{t\theta;kl}\hat{f}^n_{kl}}$$

$$X_2 \stackrel{\text{def}}{=} (1-z^h_{ij}) + (1-z^h_{i-1,j}) + (1-z^v_{ij}) + (1-z^v_{i,j-1})$$

$$X_3 \stackrel{\text{def}}{=} f_{i+1,j}(1-z^h_{ij}) + f_{i-1,j}(1-z^h_{i-1,j}) + f_{i,j+1}(1-z^v_{ij}) + f_{i,j-1}(1-z^v_{i,j-1}).$$

The M-step thus consists of two kinds of ICM updates. We first update $\mathbf{f}$ according to update equation (12) over the entire image and then update $\mathbf{z}$ according to (11) over the entire image as well. Once the new estimate $\hat{\mathbf{f}}^{n+1}$ has been obtained, a new M-step objective function is generated by substituting $\hat{\mathbf{f}}^{n+1}$ for $\hat{\mathbf{f}}^n$ to get $M(\mathbf{f},\mathbf{z}|\hat{\mathbf{f}}^{n+1};\beta)$. Then the two-step ICM procedure is repeated. In this manner, the entire GEM procedure is carried out until convergence. A special case occurs in the update of f_{ij} when $X_2 = 0$ which can occur when all neighboring z^v and z^h values are unity. Then the update equation for f_{ij} is merely the standard EM update equation since the smoothing action of the prior is suspended at this location.

The strategy above was predicated on setting the partial derivative of $M(\mathbf{f},\mathbf{z}|\hat{\mathbf{f}}^n;\beta)$ with respect to f_{ij} to zero and solving for f_{ij}. This does not guarantee that the obtained solution is a minimum of $M(\mathbf{f},\mathbf{z}|\hat{\mathbf{f}}^n;\beta)$. In the Appendix, we sketch a proof showing that the solution obtained is indeed a minimum.

4 Relationship to Other Work

The approach taken in this paper is related to other efforts in which a Bayesian prior was utilized to obtain piecewise smooth reconstructions. In those efforts, the prior was expressed via Gibbs potential functions.

Geman and McClure [7] used a ϕ function of the form $\phi(x) = \frac{x^2}{1+x^2}$ but they also mentioned the possible use of a more general ϕ function: $\phi(x) = \frac{|x|^k}{1+|x|^k}, k = 1, 2, \ldots$ There was no associated mixed variable objective function in this case.

Hebert and Leahy [5] suggested three ϕ functions that had slightly different properties: (i) $\phi(x) = \frac{x^2}{1+x^2}$, (ii) $\phi(x) = \log(1+x^2)$, and (iii) $\phi(x) = x^2$. Their main concern was to integrate a ϕ function with the likelihood energy into a GEM algorithm. There was no associated mixed variable objective function.

Green [8] and Lange [9] suggested a variety of *convex* ϕ functions (Green suggested $\phi(x) = \log\cosh(x)$). While having several mathematically desirable properties, these ϕ functions do not promote the formation of discontinuities. There were no associated mixed variable objective functions.

The approach of Leahy and Yan [10] is perhaps closest in spirit to our work. Their energy function consists of the likelihood energy and a prior that includes clique potentials on the line processes. The search method is a combination of ICM on the source intensities and Iterated Conditional Averages (ICA) on the line processes. In addition, they vary a temperature parameter in ICA to make the line processes converge to zero or one. However, they do not have a ϕ function and convergence of their ICA algorithm is not guaranteed.

Gindi *et al.* [11] used a "cusp" like potential, $\phi(x) = \frac{|x|}{1+|x|}$ and justified this potential through a series of modifications of the weak membrane energy function. The cusp could then be derived from a new mixed variable objective function in the same manner that the broken parabola was derived from the weak membrane energy function.

Lalush and Tsui [12] proposed a parametrized ϕ function flexible enough through parameter modifications to subsume a large family of ϕ functions. The design of the ϕ function was heuristic and the motivation was for empirical testing on images under a variety of conditions. There was no theoretical justification for the form of the ϕ function.

Many other researchers employed Bayesian approaches to impose smoothness and piecewise smoothness, notably Johnson *et al.* [13], and Liang *et al.* [14] on reconstructions, but their methods differ from the ones discussed here.

5 Results

We simulated the algorithm using the math phantom shown in Figure 2(a). The 40x40 phantom contains three "hot" regions of intensity 110, three cold regions of intensity 80, and a background of intensity 100. We used 40 projection angles over 360° with 40 detector bins at each projection. In our algorithm, $\mathcal{H}$ is the Radon transform. Figure 2(b) shows an EM–ML result obtained after 45 iterations which corresponds to the minimum RMS error over the entire image. Note that the picture is noisy and the low contrast hot regions are barely discernible. Figures 2(e) – (h) show the source and line process reconstructions for the deterministic annealing algorithm at high and low temperatures. The line process images depict the expected value of the line processes, $\mathbf{z}$ (see Appendix). Note that all three hot regions are recovered and the corresponding line processes are approximately correct. The initial condition for $\mathbf{f}$ was a constant intensity of 50, and $\mathbf{z}$ was initialized to 0.5. We used 13 values of β with an initial β of 0.03125 and a β doubling annealing schedule. At a given temperature, we optimize until the energy difference at successive iterations satisfies $|\Delta E| \leq \tau_k$ where k is a temperature iteration number and $\tau_{k+1} = \tau_k/2$ with $\tau_1 = 0.3$.

The annealing is terminated when $z_{ij} \leq 0.1$ or $z_{ij} \geq 0.9$, for all (i, j). The parameters were $\lambda = 0.1$ and $\alpha = 2.7$.

We substituted for the annealing strategy a "quenching" algorithm that is simply the deterministic annealing algorithm run at a single, very high value of β ($\beta = 256$) with the same parameters. The results in Figures 2(c) and (d) show an inferior reconstruction with profuse and erroneous line processes compared to the annealing results in Figures 2(e) – (h).

Table 1 show the quantitative results in the form of RMS errors in the seven regions of interest (ROI) corresponding to the six squares and background of the phantom. The RMS error for EM is taken at the 45^{th} iteration. As seen, the best total RMS error was achieved by the deterministic annealing algorithm. Qualitative and quantitative comparison to quenching shows that temperature annealing is beneficial.

The above results demonstrate an improvement due to annealing. However, it is interesting to note that a comparably good result can be obtained by merely minimizing the energy function at a fixed, intermediate value of β much lower than the final β of an annealing sequence. In essence, this corresponds to the choice of a "smooth" ϕ function.

This effect is observed in the simulations depicted in Figure 3. The 64x64 phantom shown in Figure 3(a) consists of 16 small squares (sizes 3x3, 6x6, 10x10, and 15x15), each with an intensity of 16, over a background level of 4. The projection data was collected for 64 angles.

Figures 3(b) and (c) show the reconstruction and line processes, respectively, for the deterministic annealing algorithm. We started at a low value of β and used a doubling schedule, eventually using 7 values of β. At each β value, we perform a descent operation, stopping when the percentage change in energy is below the threshold $\tau = 0.01/(\text{temperature iteration index})$. The index is incremented at each new value of β. The actual values used were: initial $\beta = 0.0001$, final $\beta = 0.0064$, $\alpha = 4$, and $\lambda = 5000$.

Figures 3(d) and (e) show the result of running the algorithm at a single value of β corresponding to the *initial* β used above. The results look quite comparable to the deterministic annealing results. Figures 3(f) and (g) show the result of running the algorithm at a single β corresponding to the *final* value used in the annealing schedule. The simulations are much worse than the annealing results as expected. All simulations used the same value of the parameters α and λ.

The above results seem to indicate that a single, intermediate value of β may be sufficient to obtain a good reconstruction and that annealing is not necessary. However, it is difficult to pick the "right" intermediate β in advance (see for instance Figures 2(e) and (f)). Deterministic annealing through a broad range of β should be sufficient to overcome this problem.

6 Extension to Higher Order Models

The prior model in this paper favors source distributions that are piecewise flat. A tacit assumption underlying this model is that the radionuclide density tends to be uniformly spatially distributed in anatomical regions, and that the

Encapsulated Postscript figure unavailable

Figure 2: Top row left (a): 64x64 phantom. Top row right (b): EM–ML Reconstruction. Second row left (c) Quenching Reconstruction. Second row right (d) Quenching line processes. Third row left (e) Reconstruction at second β. Third row right (f) Line processes at second β. Bottom row left (g) Reconstruction at final β. Bottom row right (h) Line processes at final β.

Encapsulated Postscript figure unavailable

Figure 3: Top row (a): 64x64 phantom. Second row (b) and (c): Reconstruction and line processes for deterministic annealing. Third row (d) and (e): Single low value of β reconstruction and line processes. Bottom row (f) and (g): Single high value of β reconstruction and line processes.

Table 1: RMS error in ROI for various algorithms

	EM-ML	Quenching	Deterministic Annealing
Top left	4.935	4.074	4.984
Top middle	5.385	5.271	3.394
Top right	4.831	1.120	0.382
Bottom left	7.046	6.826	6.400
Bottom middle	5.039	3.168	3.449
Bottom right	5.210	1.816	2.314
Base region	5.656	3.183	2.004
Total image	4.293	2.633	2.264

average concentration differs across anatomical regions. To the extent that this is not true, different prior models may be needed. It is still desirable to assume regions with a piecewise *homogeneous* spatial distribution (we still have discontinuities) but the piecewise flat assumption is too restrictive.

The piecewise flat mechanical model (the weak membrane) can be extended to a more expressive piecewise *linear* model – the weak plate [4]. The weak plate prior allows for ramp like regions in the reconstruction. Here, discontinuities correspond to "creases", i.e. discontinuities in the source gradient rather than in the source itself. The weak membrane prior has the unfortunate effect of turning sufficiently steep ramps into stepped terraces. A prior energy function corresponding to the weak plate is

$$E_P(\mathbf{f},\mathbf{l}) = \sum_{ij} \left\{ \left[f_{vv}(i,j) + f_{hh}(i,j)\right]^2 (1 - l_{ij}) + \alpha l_{ij} \right\}$$

where $f_{vv}(i,j)$ and $f_{hh}(i,j)$ are the second derivatives in the vertical and horizontal directions respectively. The corresponding M-step objective function is

$$\begin{aligned} M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta) &\stackrel{\text{def}}{=} \sum_{ij}\sum_{t\theta}\left[H_{t\theta ij}f_{ij} - g_{t\theta}\frac{H_{t\theta ij}\hat{f}^n_{ij}}{\sum_{kl}H_{t\theta kl}\hat{f}^n_{kl}}\log(f_{ij})\right] \\ &+ \sum_{ij}\left[\frac{-1}{\beta}\log\left(\exp(-\beta\lambda[f_{vv}(i,j)+f_{hh}(i,j)]^2) + \exp(-\beta\lambda\alpha)\right)\right]. \end{aligned}$$

A deterministic annealing algorithm can be derived from the above weak plate M-step energy function. In general, weak plate energy functions take longer to converge [4]. It is possible to improve the quality and efficiency by initializing the algorithm with the results of first running a weak membrane algorithm.

7 Conclusions

The impetus for the use of continuation methods is twofold. First, continuation allows us to approach the "correct" broken parabola potential function via the control parameter β. Second, the method may support the efficient search for good minima in the reconstruction problem though this issue has not been explored here.

The first point is important insofar as the broken parabola is, in fact, the correct potential that is derived from the original weak membrane energy function. Our continuation method provides one example of a methodology in which one first states the desiderata regarding spatial smoothness constraints (and their suspension using line processes) via an objective function. An associated ϕ function is derived from this objective function. This ϕ function by itself may be difficult to apply directly. For example, as shown in Section 2, the broken parabola had such problems. The continuation method may be used to approach the ϕ function and hence retain the original intent of the objective function.

Appendix

In this section, we start with (9) and systematically derive the objective function corresponding to the change of variables as in (11). The M-step objective function is

$$\begin{aligned} M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta) &= \sum_{ij}\sum_{t\theta}\left[\mathcal{H}_{t\theta;ij}f_{ij} - g_{t\theta}\frac{\mathcal{H}_{t\theta,ij}\hat{f}^n_{ij}}{\sum_{kl}\mathcal{H}_{t\theta;kl}\hat{f}^n_{kl}}\log(f_{ij})\right] \\ &+ \sum_{ij}\Big[-\frac{1}{\beta}\log\big(\exp(-\beta\lambda f_v^2(i,j)) + \exp(-\beta\lambda\alpha)\big)\Big] \\ &+ \sum_{ij}\Big[-\frac{1}{\beta}\log\big(\exp(-\beta\lambda f_h^2(i,j)) + \exp(-\beta\lambda\alpha)\big)\Big]. \end{aligned}$$

A change of variables in the objective function can be performed in the following manner. Consider the following transformation to be applied at every location (i,j):

$$\psi(f_v, s^h) \stackrel{\text{def}}{=} \xi(s^h) + \mu(f_v^2 - s^h) \tag{13}$$

where a similar relationship obtains for f_h and s^v. In (13),

$$\xi(s^h) \stackrel{\text{def}}{=} -\frac{1}{\beta}\log\big(\exp(-\beta\lambda s^h) + \exp(-\beta\lambda\alpha)\big)$$

and μ is a Lagrange parameter to be determined shortly. Note that the definition of $\xi(s^h)$ is written in terms of a dummy variable s^h. The variable s^h is forced towards the value f_v^2 by the action of the Lagrange constraint term in (13). Differentiating with respect to s^h and setting the result to zero, we get

$$\mu = \xi^{'}(s^h).$$

The transformation $\psi(f_v, s^h)$, after Lagrange parameter elimination, becomes

$$\psi(f_v, s^h) = \xi(s^h) + \xi^{'}(s^h)(f_v^2 - s^h). \tag{14}$$

In (14), when the derivative with respect to s^h is set to zero, we get

$$\frac{\partial\psi(f_v, s^h)}{\partial s^h} = 0 \;\Rightarrow\; \xi'(s^h) + \xi''(s^h)(f_v^2 - s^h) - \xi'(s^h) = 0$$
$$\Rightarrow\; \xi''(s^h)(f_v^2 - s^h) = 0$$

The fixed point at $s^h = f_v^2$ is a minimum provided the second derivative of $\psi(f_v, s^h)$ with respect to s^h is greater than zero. The second derivative condition is

$$\frac{\partial^2\psi(f_v, s^h)}{\partial(s^h)^2} = \xi'''(s^h)(f_v^2 - s^h) - \xi''(s^h) \;>\; 0.$$

At the fixed point $f_v^2 = s^h$, the above condition reduces to $\xi''(s^h) \;< 0$. Now

$$\xi''(s^h) = -\beta\lambda^2 \frac{\exp(-\beta\lambda(s^h - \alpha))}{[1 + \exp(-\beta\lambda(s^h - \alpha))]^2} \;\leq\; 0.$$

$\xi''(s^h)$ is less than or equal to zero everywhere and is equal to zero only when $s^h \to \infty$. Thus the second derivative condition, $\xi''(s^h) \;<\; 0$, is verified.

We now rewrite (14) as

$$\psi(f_v, s^h) = f_v^2\xi'(s^h) + (\xi(s^h) - s^h\xi'(s^h)).$$

We define the transformation $z^h \stackrel{\text{def}}{=} 1 - \frac{\xi'(s^h)}{\lambda}$ (note that this transformation is a restatement of (11)). We may invert the above transformation to get

$$s^h \stackrel{\text{def}}{=} s(z^h) = \alpha + \frac{1}{\beta\lambda}\log(\frac{z}{1-z}).$$

Using the above inverse transformation, we get

$$\psi(f_v, s(z^h)) = \lambda\left[f_v^2(1 - z^h) + \alpha z^h\right] + \frac{1}{\beta}\left[z^h\log(z^h) + (1 - z^h)\log(1 - z^h)\right]. \tag{15}$$

When (15) is minimized with respect to z^h, we obtain

$$z^h = \frac{1}{1 + \exp(-\beta\lambda(f_v^2 - \alpha))}$$

which can be shown to correspond to the expected value of the line process in the β exponentiated posterior distribution. It is, in fact these quantities z^h (and z^v) that are displayed as the evolving line processes in Section 5.

Using (15), the transformed M-step objective function is

$$\begin{aligned} M(\mathbf{f}, \mathbf{z}|\hat{\mathbf{f}}^n; \beta) \;&=\; \sum_{ij}\sum_{t\theta}\left[\mathcal{H}_{t\theta;ij}f_{ij} - g_{t\theta}\frac{\mathcal{H}_{t\theta;ij}\hat{f}^n_{ij}}{\sum_{kl}\mathcal{H}_{t\theta;kl}\hat{f}^n_{kl}}\log(f_{ij})\right] \qquad (16)\\ &+\; \lambda\sum_{ij}\left[f_v^2(i,j)(1 - z^h_{ij}) + \alpha z^h_{ij}\right] \end{aligned}$$

$$
\begin{aligned}
&+ \quad \lambda \sum_{ij} \left[f_h^2(i,j)(1 - z_{ij}^v) + \alpha z_{ij}^v \right] \\
&+ \quad \frac{1}{\beta} \sum_{ij} \left[z_{ij}^h \log(z_{ij}^h) + (1 - z_{ij}^h) \log(1 - z_{ij}^h) \right] \\
&+ \quad \frac{1}{\beta} \sum_{ij} \left[z_{ij}^v \log(z_{ij}^v) + (1 - z_{ij}^v) \log(1 - z_{ij}^v) \right] .
\end{aligned}
$$

The ICM update equations for f_{ij} given in (12) can now be obtained by differentiating (16) with respect to f_{ij} and setting it to zero. This update equation is the solution to a quadratic equation with only one solution making physical sense. Also, the solution guarantees non-negativity of f_{ij}. Similarly, the update equation for z_{ij} given in (11) can be obtained by differentiating (16) and setting it to zero.

To ensure that the solution obtained by differentiating (16) with respect to f_{ij} (or z_{ij}) and setting it to zero is indeed a minimum, it is sufficient to show that $M(\mathbf{f}, \mathbf{z}|\hat{\mathbf{f}}^n; \beta)$ is separately convex with respect to f_{ij} and z_{ij}. This can be shown by examining the second derivatives:

$$
\begin{aligned}
\frac{\partial^2 M(\mathbf{f}, \mathbf{z}|\hat{\mathbf{f}}^n; \beta)}{\partial f_{ij}^2} &= \frac{1}{f_{ij}^2} \left(\sum_{t\theta} g_{t\theta} \frac{\mathcal{H}_{t\theta;ij} \hat{f}_{ij}^n}{\sum_{kl} \mathcal{H}_{t\theta;kl} \hat{f}_{kl}^n} \right) \\
&+ \quad 2\lambda(4 - z_{ij}^h - z_{i-1,j}^h - z_{ij}^v - z_{i,j-1}^v) \ \geq \ 0, \\
\frac{\partial^2 M(\mathbf{f}, \mathbf{z}|\hat{\mathbf{f}}^n; \beta)}{\partial (z_{ij}^h)^2} &= \frac{1}{\beta z_{ij}^h (1 - z_{ij}^h)} \ \geq \ 0
\end{aligned}
$$

(The case where the second derivatives go to zero are pathological). We have guaranteed that a single ICM step taken in any direction is a descent step since the fixed point in that direction corresponds to a minimum. This minimum is achieved in a single step. Hence, convergence of the ICM algorithm to a local minimum of $M(\mathbf{f}, \mathbf{z}|\hat{\mathbf{f}}^n; \beta)$ is guaranteed.

References

[1] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6(6), pp. 721–741, November 1984.

[2] D. Geiger and F. Girosi, "Parallel and Deterministic Algorithms from MRFs: Surface Reconstruction", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-13(5), pp. 401–412, May 1991.

[3] A. Rangarajan and R. Chellappa, "A continuation method for image estimation and segmentation", Technical Report CAR-TR-586, Center for Automation Research, University of Maryland, October 1991.

[4] A. Blake and A. Zisserman, *"Visual Reconstruction"*, Artificial Intelligence, MIT Press, Cambridge, MA, 1987.

[5] T. Hebert and R. Leahy, "A Generalized EM Algorithm for 3-D Bayesian Reconstruction for Poisson Data using Gibbs Priors", *IEEE Transactions on Medical Imaging*, vol. MI-8(2), pp. 194–202, June 1989.

[6] J. Besag, "On the statistical analysis of dirty pictures", *Journal of the Royal Statistical Society, B*, vol. 48(3), pp. 259–302, 1986.

[7] S. Geman and D. E. McClure, "Statistical Methods for Tomographic Image Reconstruction", In *Proceedings of the 46th Session of the ISI, Bulletin of the ISI*, 1987.

[8] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm", *IEEE Transactions on Medical Imaging*, vol. MI-9(1), pp. 84–93, March 1990.

[9] K. Lange, "Convergence of EM Image Reconstruction Algorithms with Gibbs Smoothing", *IEEE Transactions on Medical Imaging*, vol. MI-9(4), pp. 439–446, December 1990.

[10] R. Leahy and X. Yan, "Incorporation of Anatomical MR data for improved Functional Imaging with PET", In A. C. F. Colchester and D. J. Hawkes, editors, *Information Processing in Medical Imaging*, pp. 105–120, Springer–Verlag, 1991.

[11] G. Gindi, M. Lee, A. Rangarajan, and I. G. Zubal, "Bayesian Reconstruction of Functional Images Using Registered Anatomical Images as Priors", In A. C. F. Colchester and D. J. Hawkes, editors, *Information Processing in Medical Imaging*, pp. 121–131, Springer–Verlag, 1991.

[12] D. S. Lalush and B. M. W. Tsui, "Simulation evaluation of Gibbs prior distributions for use in maximum *a posteriori* SPECT reconstructions", *IEEE Transactions on Medical Imaging*, vol. MI-11, pp. 267–275, 1992.

[13] V. E. Johnson, W. H. Wong, X. Hu, and C.-T. Chen, "Image Restoration Using Gibbs Priors: Boundary Modeling, Treatment of Blurring, and Selection of Hyperparameter", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-13(5), pp. 413–425, May 1991.

[14] Z. Liang, R. Jaszczak, C. Floyd, and K. Greer, "A Spatial Interaction Model for Statistical Image Processing", In D. A. Ortendahl and J. Llacer, editors, *Information Processing in Medical Imaging*, pp. 1–14, Wiley-Liss, 1989.