Convergence of MAP-EM Algorithms with Nonquadratic Smoothing Priors

Soo-Jin Lee

Department of Electronic Engineering, Paichai University

Abstract

Bayesian MAP-EM approaches have been quite useful for tomographic reconstruction in that they can stabilize the instability of well-known ML-EM approaches, and can incorporate *a priori* information on the underlying emission object. However, MAP reconstruction algorithms with expressive priors often suffer from the optimization problem when their objective functions are *non*quadratic. In our previous work [1], we showed that the use of deterministic annealing method greatly reduces computational burden for optimization and provides a good solution for nonquadratic objective functions. Here, we further investigate the convergence of the deterministic annealing algorithm; our experimental results show that, while the solutions obtained by a simple quenching algorithm depend on the initial conditions, the estimates converged via deterministic annealing algorithm are consistent under various initial conditions.

I. INTRODUCTION

In emission computed tomography (ECT), a dominant source of image degradation is the noise due to variation in the number of photons emitted from a radionuclide. In addition, the three most significant physical factors that contribute to the image degradation are attenuation, scatter, and detector response. Well-known maximum likelihood (ML) approaches using the expectation maximization (EM) algorithm are attractive in that it can naturally express accurate system models of physical effects, and can accurately model the statistical character of the data. Unfortunately, however, these approaches are known to be unstable for the noise levels and numbers of measurements that characterize ECT. In contrast, maximum *a posteriori* (MAP) approaches in the context of a Bayesian framework overcome this instability by incorporating prior information while retaining the above advantages of ML-EM approaches. In this case, prior information typically takes the form of constraints on the local spatial structure of the underlying emission object. Over the last decade, a host of different formulations for priors have been proposed in the literature [2, 3, 4, 5, 6, 7]; some of these implicitly model the underlying radionuclide density as globally smooth, and others extend the smoothness model by allowing for spatial discontinuities. Discontinuity preservation is associated with a smoothing penalty that is a *non*quadratic function [8, 9, 1] of nearby pixel differences, whereas conventional (e.g. membrane) smoothing priors use quadratic penalties. The nonquadratic priors can exhibit good performance, but suffer difficulties in optimization.

Our previous work [1] showed that the use of deterministic annealing algorithm can greatly reduce the computational burden for optimization and provide a good solution for nonquadratic objective functions. In this work we further investigate the deterministic annealing algorithm applied to our tomographic reconstruction problem, and test the reliability of the algorithm in terms of its convergence to one solution under various conditions.

II. BAYESIAN RECONSTRUCTION MODELS WITH NONQUADRATIC PRIORS

Geman and Geman [8] introduced the powerful idea of including unobservable "line processes" in the image model in order to preserve discontinuities in the image. Versions of the line processes have been proposed for medical imaging [4, 10, 5]. The line processes are binary variables (0 or 1) that act to suspend smoothness constraints at sites where they are turned on ($l = 1$). A horizontal line process at location (i, j), $l_{i,j}^h = 1$, indicates a horizontal edge (discontinuity between pixels along the vertical direction) and a vertical line process, $l_{i,j}^v = 1$, indicates a vertical edge (discontinuity along the horizontal direction).

The Bayesian approach combines the likelihood and the prior to yield *a posteriori* probability via Bayes' Theorem. Since the priors include the unobservable binary variables l, which represent the information regarding discontinuities, we start by formulating our reconstruction problem from Bayes' Theorem with the aid of the line processes:

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

where f, l, and g are the source intensities, line processes, and projection data, respectively, and F, L, G are the associated 2-D random fields. The term $\Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})$ is the prior probability for the fields F and L, $\Pr(\mathbf{G} = \mathbf{g}|\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})$ the likelihood, and $\Pr(\mathbf{G} = \mathbf{g})$ a constant. One possible

approach is then to estimate both the source field and the discontinuity field simultaneously by maximizing the *a posteriori* probability distribution, the MAP estimate $(\hat{\mathbf{f}}, \hat{\mathbf{l}})$:

$$
\begin{aligned}
(\hat{\mathbf{f}}, \hat{\mathbf{l}}) &= \arg\max_{(\mathbf{f},\mathbf{l})} \Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l} | \mathbf{G} = \mathbf{g}) \\
&= \arg\min_{(\mathbf{f},\mathbf{l})} [- \log \Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}) \\
&\quad - \log \Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})],
\end{aligned}
$$

where the penultimate equality derives from the fact that $\mathbf{G}$ is independent of $\mathbf{L}$.

We model both the likelihood and prior as Gibbs distributions with energy function E and the partition function Z (a suitable normalization):

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

where E is the associated Gibbs prior energy function, and Z is a normalization of no concern here. 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 $\mathbf{x}$.

Since the number of detected counts is independently Poisson distributed, the likelihood is modeled as independent Poisson processes:

$$
\Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}) = \prod_{t,\theta} \frac{\bar{g}_{t,\theta}^{g_{t,\theta}} \exp(-\bar{g}_{t,\theta})}{g_{t,\theta}!}, \quad (2)
$$

where $\bar{g}_{t,\theta} \stackrel{\text{def}}{=} \sum_{i,j} \mathcal{H}_{t,\theta;i,j} f_{i,j}$. In Equation (2), $g_{t,\theta}$ is the number of detected counts in the detector bin indexed by t at angle θ, $\bar{g}_{t,\theta}$ is the expected number of counts for a particular source $\mathbf{f}$, and $\mathcal{H}_{t,\theta;i,j}$ is the probability that a photon emitted from source location (i,j) hits detector bin t at angle θ. In this case, the likelihood energy E_D is given by

$$
E_D(\mathbf{f}) = \sum_{t,\theta} [-g_{t,\theta} \log(\bar{g}_{t,\theta}) + \log(g_{t,\theta}!)] + \sum_{t,\theta} \bar{g}_{t,\theta}. \quad (3)
$$

The nonquadratic prior energies derived from mechanical analogs [11, 1] are given by

$$
E_P^M(\mathbf{f},\mathbf{l}) = \lambda \sum_{i,j} \left[f_v^2(i,j)(1 - l_{i,j}^h) + f_h^2(i,j)(1 - l_{i,j}^v) \right]
$$

$$
+ \alpha \sum_{i,j} \left(l_{i,j}^h + l_{i,j}^v \right) \quad (4)
$$

for weak membrane (WM), and

$$
E_P^P(\mathbf{f},\mathbf{l}) =
$$

$$
\lambda \sum_{i,j} \left\{ \left[f_{vv}^2(i,j) + 2 f_{hv}^2(i,j) + f_{hh}^2(i,j) \right] (1 - l_{i,j}) \right\}
$$

$$
+ \alpha \sum_{i,j} l_{i,j} \quad (5)
$$

for weak plate (WP). In (4) and (5), the positive parameter λ weights the stabilizer respect to $E_D(f)$. It

is shown [1, 9] that the line processes can be expressed implicitly in the "broken parabola" function, and the problem in this case is reduced to a minimization over the source intensities $\mathbf{f}$ only. However, the reduced energy functions are still nonqudratic.

From (3), (4), and (5), the MAP estimates become the minimization of the following overall energy functions:

$$
\begin{aligned}
E^M(\mathbf{f},\mathbf{l}) &= E_D(\mathbf{f}) + E_P^M(\mathbf{f},\mathbf{l}) \\
E^P(\mathbf{f},\mathbf{l}) &= E_D(\mathbf{f}) + E_P^P(\mathbf{f},\mathbf{l})
\end{aligned}
$$

where E^P and E^M are overall energy functions for WP and WM, respectively. The above overall energy functions that involve the nonquadratic prior terms (E_M^P and E_P^P) are clearly non-convex, and may have numerous stable states.

III. MINIMIZING ENERGY FUNCTIONS

In general, energy functions with weak continuity constraints may have many local minima, many of which are not global; any intermediate state between two successive stable states is higher than either of the stable states. Since gradient-based descent methods search for the minimum of a function by successively adjusting the arguments of the function until it can not be reduced any further, they may stick at a local minimum with the higher energy state than the global minimum.

To overcome the problem of local minima, an optimization method that has an ability to "jump out" of a local minimum is necessary. Kirkpatrick, *et al.* [12] proposed a method, called simulated annealing, in order to optimize non-convex functionals. The simulated annealing method is based on an algorithm invented by Metropolis [13] that simulates thermodynamics, specifically the way that liquids freeze and crystallize, or metals cool and anneal. At high temperatures, the molecules of a liquid metal move freely with respect to one another. If the liquid metal is cooled quickly or "quenched", it does not reach the minimum energy state but rather ends up in amorphous state having somewhat higher energy. On the other hand, if it is cooled slowly so that the atoms can be redistributed as they lose mobility, nature is able to find the minimum energy state. One difficulty in the simulated annealing method is that it might take very long to attain thermal equilibrium at low temperatures.

An alternative approach is to use a "continuation method" [1, 9, 14]. In this case the energy function is approached by a *sequence* of energy functions indexed by a parameter β. The sequence is obtained by transformation of the probability distributions in Equation (1) to

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

where $Z(\mathbf{x}; \beta)$ is the partition function parametrized by β. As β approaches infinity, this new distribution becomes a highly peaked version of the distribution in Equation (1). As far as MAP estimation is concerned, the location of the maximum is unchanged. A sequence of energy functions is constructed which more closely approximates the original energy function as a control parameter β is increased. This technique is called deterministic annealing. The parameter β in deterministic annealing may be identified as the inverse of a computational temperature used in conventional simulated annealing approaches. 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 to the energy functions at low temperatures. The extensive derivations of deterministic annealing for both WM and WP can be found in [1].

IV. SIMULATIONS AND RESULTS

It has been shown [9] that there exist local minima in the broken parabola objective function by comparing the images reconstructed by different annealing schedules (i.e., "quenching" versus DA). In [9], quenching was performed by executing the DA algorithm at a single, very high value of β. The results showed that reconstructions obtained with quenching yielded inferior quality. However, we note here that, in order to validate the convergence of the solution to a minimum, it is necessary to show that reconstructions are independent of initial estimates. In other words, different initial estimates should result in the same final reconstruction.

To test the above idea, we used a 64×64 piecewise constant phantom shown in Figure 1(a). For projection data, we used 64 projection angles over $360°$ with 96 detector bins at each projection. The total number of detector counts was approximately 800K with attenuation. To simplify the problem, we considered attenuation only. For initial estimates, we used three widely different estimates shown in Figure 2: a constant image (denoted A) with 100 for the pixel intensity, a checkerboard image (denoted B) with 1 and 100 for the pixel intensities, and a noise image (denoted C) with the pixel intensities uniformly distributed in [1,255]. Note that the range of grey levels used for the initial estimates is relatively large compared to that for the phantom whose highest intensity is 15.

We tested two different annealing schedules, DA and quenching, and used the same hyperparameter settings ($\lambda = 1.0$ and $\alpha = 0.5$) for both algorithms. In DA algorithm, the initial value for the parameter β was chosen to be small ($\beta = 0.001$) enough to yield very smooth initial reconstructions [9] for various initial estimates. Iterations at a given β was terminated when

the relative energy change was negligible, and the entire simulation was terminated when the sequence of energy functions reach the broken parabola. With these criteria, there were approximately 5 to 25 iterations per each β, and the annealing schedule ran through 16 values with a doubling at each new value of β. The total number of iterations in this case was around 220. For the quenching algorithm, we set $\beta = 16.3840$, which was the same as the final value of β for the DA algorithm. The average number of iterations for quenching was around 250.

Figure 1(b) shows root mean squared error (RMSE) curves for both DA and quenching with three different initial estimates. The RMSE here was measured for the entire image. It is clearly shown that the RMSE's for the DA algorithm started from different initial estimates converge to the same value, but those for the quenching algorithm converge to the different values depending on the initial estimates. In addition, the average RMSE for quenching is much larger than that for DA. Notice that, after 50 iterations, while the DA algorithm yields very similar quality of reconstructions in terms of RMSE, the quenching algorithm produces irregular quality through the entire simulation depending on the initial estimates. Figure 2 shows reconstructions at various iterations. In general, while the DA algorithm produces consistent reconstructions after a certain number of iterations (60 to 70 iterations in this simulation) regardless of the initial estimates, the quenching algorithm yields different reconstructions at each iteration depending on the initial estimates.

V. CONCLUSION

We have tested the convergence of MAP-EM reconstruction algorithms with nonquadratic smoothing priors. Our simulation results show that, while the solutions obtained by the simple quenching algorithm depend on the initial conditions, the estimates converged via deterministic annealing algorithm are consistent under the various initial conditions considered in the experiments. In addition, the results from quenching show anecdotally that there exist local minima in the broken parabola objective function. It is a difficult problem to prove mathematically whether the DA minimum is global. This could be tested experimentally by comparing the DA results with those obtained by simulated annealing.

VI. REFERENCES

[1] S. J. Lee, A. Rangarajan, and G. Gindi, "Bayesian Image Reconstruction in SPECT Using Higher Order Mechanical Models as Priors", *IEEE Trans. on Medical Imaging*, MI-14(4), pp. 669–680, Dec. 1995.

[2] S. Geman and D. E. McClure, "Statistical Methods for Tomographic Image Reconstruction", *Bulletin of the*
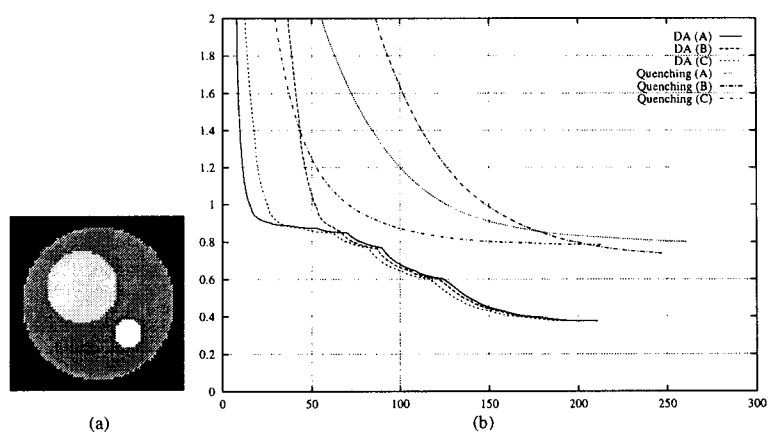
(a)   (b)

Fig. 1 (a) Phantom (b) RMSE plots for the DA and quenching algorithms with three different initial estimates. A, B, and C denote the constant, checkerboard, and noise initial estimates, respectively.
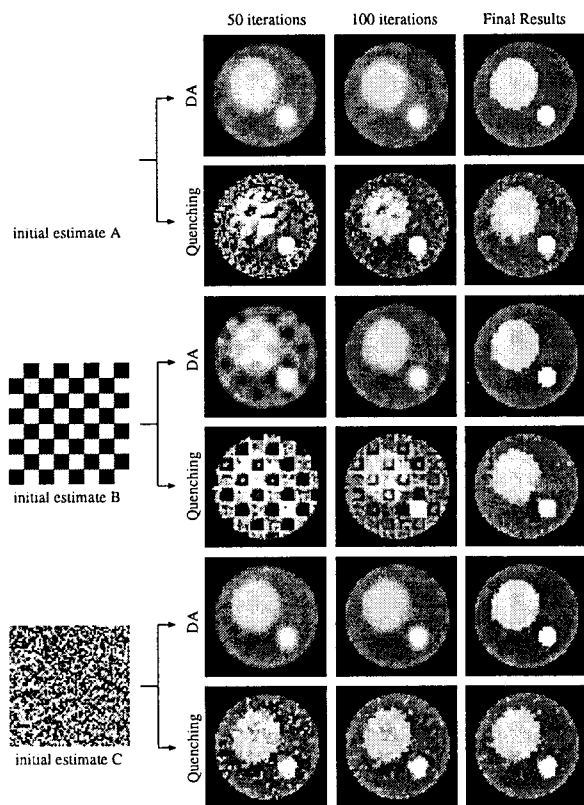

Fig. 2 Quenching versus deterministic annealing.

International Statistical Institute, LII-4, pp. 5–21, 1987.

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

[4] 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 Trans. on Pattern Analysis and Machine Intelligence*, PAMI-13(5), pp. 413–425, May 1991.

[5] G. Gindi, M. Lee, A. Rangarajan, and G. Zubal, "Bayesian Reconstruction of Functional Images Using Anatomical Information as Priors", *IEEE Trans. on Medical Imaging*, MI-12, pp. 670–680, Dec. 1993.

[6] S. J. Lee, G. R. Gindi, I. G. Zubal, and A. Rangarajan, "Using Ground-Truth Data to Design Priors in Bayesian SPECT Reconstruction", In Y. Bizais, C. Barillot, and R. D. Paola, editors, *Information Processing in Medical Imaging*, pp. 27–38, Kluwer Academic Publishers, 1995.

[7] S.J. Lee, I.T. Hsiao, and G.R. Gindi, "The Thin Plate as a Regularizer in Bayesian SPECT Reconstruction", *IEEE Trans. on Nuclear Science*, NS-44(3), pp. 1381–1387, June 1997.

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

[9] G. Gindi, A. Rangarajan, M. Lee, P. J. Hong, and G. Zubal, "Bayesian Reconstruction for Emission Tomography via Deterministic Annealing", In H. Barrett and A. Gmitro, editors, *Information Processing in Medical Imaging*, pp. 322–338, Springer-Verlag, 1993.

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

[12] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing", *Science*, 220, pp. 671–680, 1983.

[13] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and M. Teller, "Equation of State Calculations by Fast Computing Machines", *Journal of Chemical Physics*, 21, pp. 1087–1091, 1953.

[14] A. Rangarajan, M. Lee, G. Zubal, and G. Gindi, "A Continuation Method for Emission Tomography", In *Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference*, volume II, pp. 1204–1207, October 1992.