A Paraboloidal Surrogates Algorithm for Convergent Penalized-Likelihood Emission Image Reconstruction

Jeffrey A. Fessler and Hakan Erdoğan

4240 EECS Bldg., University of Michigan, Ann Arbor, MI 48109-2122

Abstract

We present a new algorithm for penalized-likelihood emission image reconstruction. The algorithm monotonically increases the objective function, converges globally to the unique maximizer, and easily accommodates the nonnegativity constraint and nonquadratic but convex penalty functions. The algorithm is based on finding paraboloidal surrogate functions for the log-likelihood at each iteration: quadratic functions that are tangent to the log-likelihood at the current image estimate, and lie below the log-likelihood over the entire nonnegative orthant. These conditions ensure monotonicity. The paraboloidal surrogates are maximized easily using existing algorithms such as coordinate ascent. Simulation results show that the proposed algorithm converges faster than the SAGE algorithm, yet the new algorithm is somewhat easier to implement.

I. INTRODUCTION

Statistical methods for tomographic image reconstruction are increasingly being used clinically. For widespread acceptance of statistical methods, fast converging algorithms are essential. This paper summarizes a new algorithm for reconstructing emission images from Poisson sinogram measurements. The algorithm
- monotonically increases the objective function,
- is globally convergent,
- naturally accommodates the nonnegativity constraint,
- requires less CPU time per iteration than the space-alternating generalize EM (SAGE) algorithm [1].

For reconstruction problems (such as 2D PET and SPECT) where the system matrix G can be precomputed and stored, we recommend this new algorithm over our previously published algorithms for penalized-likelihood reconstruction. For reconstruction problems where the system matrix is represented in factored form [2], methods that can better exploit this representation, such as the conjugate-gradient algorithm, appear to remain preferable.

II. THE PROBLEM

We assume the usual Poisson statistical model for the emission measurements Y_i, $i = 1, \ldots, n_d$:

$$Y_i \sim \text{Poisson}\left\{ \bar{y}_i(\underline{\lambda}^{\text{true}}) \right\} \qquad (1)$$

$$\bar{y}_i(\underline{\lambda}) = \sum_{j=1}^{n_p} a_{ij}\lambda_j + r_i, \qquad (2)$$

where λ_j is the unknown activity in the jth pixel, $j = 1, \ldots, n_p$, the r_i's denote background events such as random coincidences and scatter, and $A = \{a_{ij}\}$ represents the system model, including ray-dependent factors such as attenuation and detector efficiency. We assume the r_i's and a_{ij}'s are known nonnegative constants. We assume $r_i > 0$, which is reasonable and realistic since any real PET scan will have nonzero randoms. (It also turns out to simplify our algorithm.) We wish to estimate $\underline{\lambda} = [\lambda_1, \ldots, \lambda_{n_p}]$ from a measurement realization $\{Y_i = y_i\}_{i=1}^{n_d}$. The log-likelihood is:

$$L(\underline{\lambda}) \equiv \sum_{i=1}^{n_d} y_i \log \bar{y}_i(\underline{\lambda}) - \bar{y}_i(\underline{\lambda}), \qquad (3)$$

neglecting constants independent of $\underline{\lambda}$. For penalized-likelihood image reconstruction, one seeks the image that maximizes an objective function as follows:

$$\hat{\underline{\lambda}} = \arg \max_{\underline{\lambda} \geq \underline{0}} \Phi(\underline{\lambda}), \quad \Phi(\underline{\lambda}) = L(\underline{\lambda}) - \beta R(\underline{\lambda}), \qquad (4)$$

where $R(\underline{\lambda})$ is a roughness penalty included for regularization. The resolution/noise tradeoff is controlled by the regularization parameter β [3]. The problem is to find the maximizer $\hat{\underline{\lambda}}$ using an algorithm that converges rapidly and that requires as little CPU time per iteration as possible. Usually these are conflicting requirements, since often one can modify an algorithm to converge in fewer iterations, but at the expense of more work per iteration.
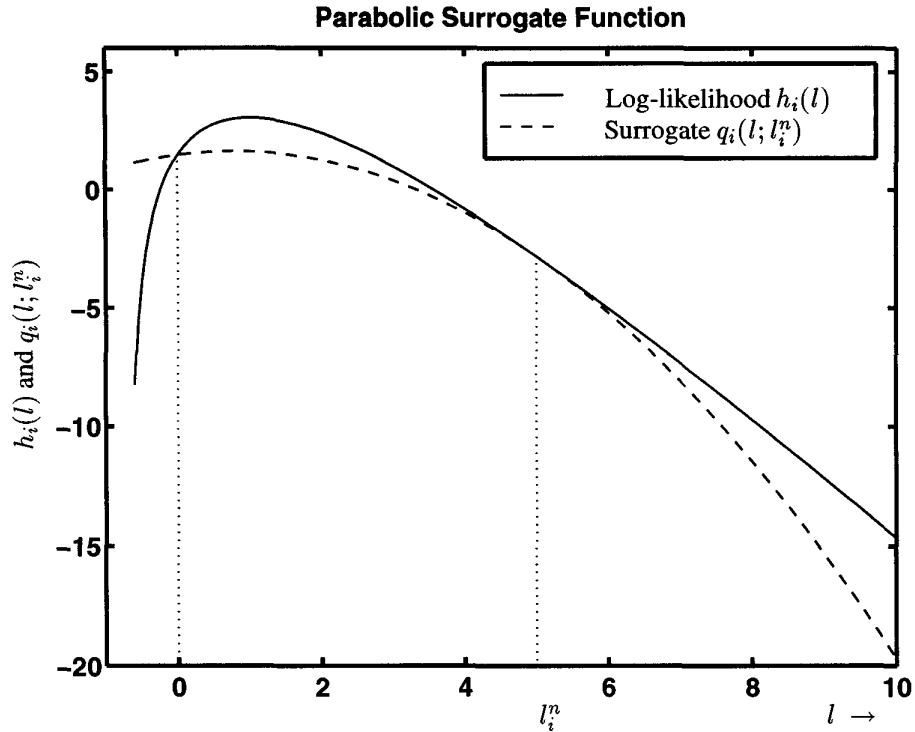
Parabolic Surrogate Function

Fig. 1 Illustration of 1D parabolic surrogate function. Note that $q_i(l; l_i^n) \leq h_i(l)$ for $l \geq 0$.

III. THE NEW ALGORITHM

We had previously recommended the SAGE algorithm [1] as a fast globally-convergent algorithm for this problem. The construction of the SAGE algorithm requires certain minimizations that are somewhat unusual in the tomographic literature, and must be implemented carefully to achieve reasonable CPU time per iteration.

The new algorithm we propose is based on the following simple idea, illustrated in one dimension in Fig. 1. Since the log-likelihood $L(\underline{\lambda})$ is difficult to maximize directly, we endeavor to find a *surrogate function* $Q(\underline{\lambda}; \underline{\lambda}^n)$ that is easier to maximize, and maximize that function at the nth iteration, *i.e.*:

$$\underline{\lambda}^{n+1} = \arg\max_{\underline{\lambda} \geq \underline{0}} Q(\underline{\lambda}; \underline{\lambda}^n) - \beta R(\underline{\lambda}) \qquad (5)$$

If we choose the sequence of surrogate functions $Q(\underline{\lambda}; \underline{\lambda}^n)$ properly, then the sequence of iterates $\{\underline{\lambda}^n\}$ will converge to the maximizer $\hat{\underline{\lambda}}$. The SAGE algorithm is indirectly based on this idea; the expected conditional log-likelihood of the "complete" data space given the observed data is a type of surrogate function that indeed satisfies the conditions sufficient to ensure convergence [1]. However, the statistical construction of the surrogate functions for the SAGE and other EM algorithms can seem somewhat mysterious.

In this paper, we propose a new approach to constructing surrogate functions that uses only basic calculus principles. We can rewrite the log-likelihood $L(\underline{\lambda})$ in (3) as follows:

$$L(\underline{\lambda}) = \sum_{i=1}^{n_d} h_i([A\underline{\lambda}]_i)$$

where

$$[A\underline{\lambda}]_i = \sum_{j=1}^{n_p} a_{ij} \lambda_j$$

is the projection of the object along the ith ray, and

$$h_i(l) = y_i \log(l + r_i) - (l + r_i)$$

is the *marginal log-likelihood* of the ith measurement. The h_i functions are concave, and strictly concave if $y_i > 0$.

Our new proposed strategy for choosing the surrogate function Q is to first find one-dimensional parabolic surrogate functions $q_i(l; l_i^n)$, as illustrated in Fig. 1, where

$$l_i^n \triangleq [A\underline{\lambda}^n]_i.$$

We then combine these 1D surrogate functions to form an overall surrogate function as follows:

$$Q(\underline{\lambda}; \underline{\lambda}^n) \triangleq \sum_{i=1}^{n_d} q_i([A\underline{\lambda}]_i; l_i^n). \qquad (6)$$

Provided that the 1D surrogate functions satisfy the following three properties:

- $q_i(l_i^n; l_i^n) = h_i(l_i^n)$,
- $\dot{q}_i(l_i^n; l_i^n) = \dot{h}_i(l_i^n)$
- $q_i(l; l_i^n) \leq h_i(l), \ \forall l \geq 0$,

then it can be shown that the recursive algorithm given by (5) will monotonically increase $L(\underline{\lambda})$, and in fact can be shown to converge globally by a proof similar to that in [1].

Using the fact that the h_i's are concave and that the first derivatives of the h_i's are convex, one can show [4] that the above three conditions will be satisfied if we choose parabolic surrogate functions as follows:

$$q_i(l; l_i^n) = h_i(l_i^n) + \dot{h}_i(l_i^n)(l - l_i^n) - \frac{1}{2} n_i(l_i^n)(l - l_i^n)^2, \quad (7)$$

where

$$n_i(l) \triangleq \begin{cases} \frac{2}{l^2}\left[h_i(l) - h(0) - l\dot{h}_i(l) \right], & l > 0 \\ -\ddot{h}_i(l), & l = 0. \end{cases} \quad (8)$$

The basic idea is illustrated in Fig. 1. The parabolic surrogate function $q_i(l; l_i^n)$ has the same value as the marginal log-likelihood $h_i(l)$ at the current projection value $l = l_i^n$, and has the same slope at that point. This is evident from (7). In addition, the parabolic function lies *below* $h_i(l)$ for all nonnegative l. This is the key to having a monotonic algorithm[1]. The proof that this q_i choice satisfies the third of the above three conditions is somewhat detailed, and is described in [4]. When the parabolic q_i functions are "assembled" as in (6) to form an overall surrogate function, the final form of Q is a paraboloid, *i.e.* a quadratic form. One can maximize this quadratic form by any number of different methods. (Exact maximization is not necessary, due to the iteration (5).) We have chosen to use successive over-relaxation or coordinate ascent [5] for this maximization, since it easily accommodates both the nonnegativity constraint as well as nonquadratic penalty functions.

In words, the algorithm alternates between two steps. The first step is to find the coefficients of the surrogate parabolas in (7) using (8). This involves a single-pass over the sinogram with trivial computation. Then the paraboloid (6) is partially maximized by a single cycle of coordinate ascent. This step requires roughly the equivalent of one forward and one backprojection, similar to most iterative algorithms.

---

[1] One could also find *approximating* parabolas using Newton's method, but these parabolas can cross $h_i(l)$, and the result is a nonmonotonic algorithm that is not guaranteed to converge. Our construction using (8) avoids this problem.

IV. RESULTS

We used the same brain emission simulation reported in [1] to evaluate the proposed algorithm. In [1], the SAGE algorithm was compared to many alternatives, including Green's one-step late (OSL) method [6], Kaufman's bounded line search [7], and the generalized EM (GEM) algorithm of Hebert and Leahy [8]. We found that the SAGE algorithm converged faster than all of the above methods, so here we focus on comparing the proposed algorithm to just the SAGE algorithm.

Figure 2 plots the increase in the log-likelihood $\Phi(\underline{\lambda}^n) - \Phi(\underline{\lambda}^0)$ versus CPU time on a DEC AlphaStation 600 5/333 workstation. The proposed algorithm (paraboloidal surrogates coordinate ascent (PSCA) algorithm) converges a little bit faster than SAGE, in part because it requires about 10% less CPU time per iteration, and in part because it increases Φ more each iteration. The differences are modest since both algorithms converge quite quickly, so there is limited room remaining for improvement. The new algorithm is simpler to implement than SAGE as well.

The resulting images are indistinguishable from those shown in [1] since we used the same objective function and both PSCA and SAGE algorithms are globally convergent.

V. DISCUSSION

The proposed algorithm is based on the "optimization transfer" principle. Since the original objective function Φ is cannot be maximized directly, we instead maximize a sequence of surrogate functions $\phi(\underline{\lambda}; \underline{\lambda}^n)$. The key is to choose surrogate functions that are easier to maximize than Φ, but have low curvature (high curvature surrogate functions lead to slow convergence rate [4,9]). Many algorithms in the literature are based (implicitly or explicitly) on optimization transfer ideas, include EM, SAGE, grouped coordinate ascent [10], the convex algorithm [11], and ISRA [12–14]. In almost all cases, the surrogate functions are *separable*, which makes them trivial to maximize, but also means that they have very high curvature and hence poor convergence rate. The paraboloidal surrogate functions that we have proposed here and in [4] are the first that we are aware of that are nonseparable. Generally nonseparable functions are harder to maximize. Fortunately, a notable exception is quadratic surrogate functions, which is the choice we have made. This choice may be suboptimal; it is possible that there exist other nonseparable surrogate functions that are easily maximized but have even lower curvature and hence yield faster convergence rates.
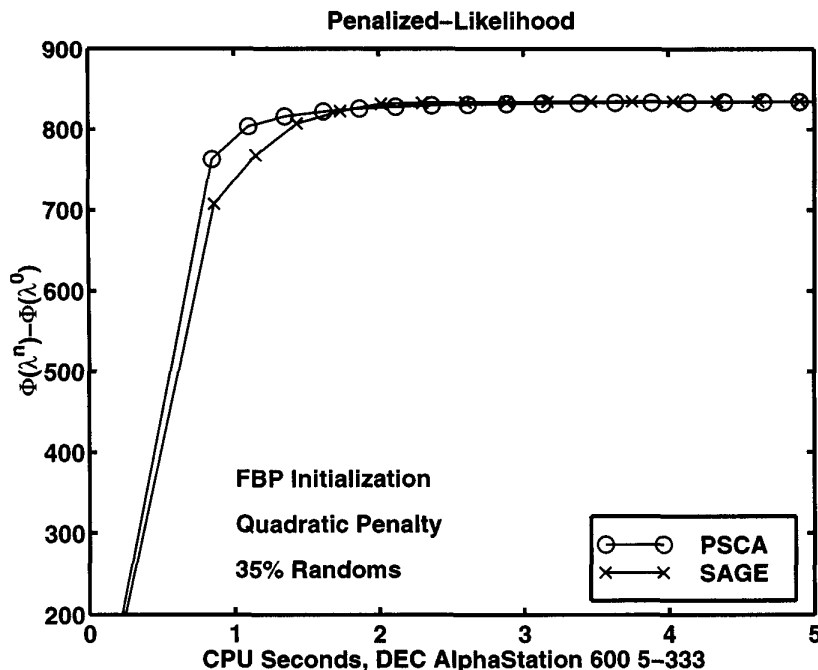
Penalized–Likelihood

Fig. 2 Comparison of convergence rate of proposed paraboloidal surrogate coordinate ascent (PSCA) algorithm versus SAGE algorithm.

VI. REFERENCES

[1] J. A. Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Tr. Im. Proc.*, vol. 4, no. 10, pp. 1417–29, October 1995.

[2] J. Qi, R. M. Leahy, S. R. Cherry, A. Chatziioannou, and T. H. Farquhar, "High resolution 3D Bayesian image reconstruction using the microPET small-animal scanner," *Phys. Med. Biol.*, vol. 43, no. 4, pp. 1001–14, April 1998.

[3] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Tr. Im. Proc.*, vol. 5, no. 9, pp. 1346–58, September 1996.

[4] H. Erdoğan and J. A. Fessler, "Fast monotonic algorithms for transmission tomography," *IEEE Tr. Med. Im.*, 1998. Submitted.

[5] J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Tr. Med. Im.*, vol. 13, no. 2, pp. 290–300, June 1994.

[6] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Tr. Med. Im.*, vol. 9, no. 1, pp. 84–93, March 1990.

[7] L. Kaufman, "Implementing and accelerating the EM algorithm for positron emission tomography," *IEEE Tr. Med. Im.*, vol. 6, no. 1, pp. 37–51, March 1987.

[8] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Tr. Med. Im.*, vol. 8, no. 2, pp. 194–202, June 1989.

[9] H. Erdoğan and J. A. Fessler, "Accelerated monotonic algorithms for transmission tomography," in *Proc. IEEE Intl. Conf. on Image Processing*, vol. 2, pp. 680–4, 1998.

[10] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange, "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction," *IEEE Tr. Med. Im.*, vol. 16, no. 2, pp. 166–75, April 1997.

[11] K. Lange and J. A. Fessler, "Globally convergent algorithms for maximum a posteriori transmission tomography," *IEEE Tr. Im. Proc.*, vol. 4, no. 10, pp. 1430–8, October 1995.

[12] M. E. Daube-Witherspoon and G. Muehllehner, "An iterative image space reconstruction algorithm suitable for volume ECT," *IEEE Tr. Med. Im.*, vol. 5, no. 2, pp. 61–66, June 1986.

[13] J. M. Ollinger, "Iterative reconstruction-reprojection and the expectation-maximization algorithm," *IEEE Tr. Med. Im.*, vol. 9, no. 1, pp. 94–98, March 1990.

[14] A. R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," *IEEE Tr. Med. Im.*, vol. 12, no. 2, pp. 328–333, June 1993.

See http://www.eecs.umich.edu/~fessler.

1135