EM and Gradient Algorithms for Transmission Tomography with Background Contamination

Jeffrey A. Fessler
Old address: 3480 Kresge III, Box 0552, University of Michigan, Ann Arbor, MI 48109-0552
email: fessler@umich.edu
Technical Report UM-PET-JF-94-1

October 2, 2003

The original version of this was in 1994. The algorithms described herein are obsolete and not recommended. See the book chapter [3] for more contemporary methods and comprehensive references.

Abstract

This report describes slight extensions of the expectation-maximization (EM) algorithm and the gradient algorithm [1] for penalized-likelihood transmission reconstruction but that accommodates nonzero additive background contamination in the Poisson model. For definitions of the notation, etc., see [1, 2].

I. Gradient Algorithm

Lange [1, 4, 5] has proposed an iterative gradient algorithm that has the desirable property that it automatically enforces nonnegativity. In this paper we present a slightly modified version of this algorithm that accommodates nonzero r_n factors. First, observe that we can rewrite the partial derivatives of Φ as follows:

$$\frac{\partial}{\partial \mu_j}\Phi(\mu) = \dot{L}_j^{(+)}(\mu) - \beta \dot{P}_j^{(+)}(\mu) - \dot{L}_j^{(-)} - \beta \dot{P}_j^{(-)}(\mu),$$

where

$$\begin{aligned} \dot{L}_j^{(+)}(\mu) &= \sum_n a_{nj}\left(\bar{y}_n(\mu) - r_n + \frac{y_n r_n}{\bar{y}_n(\mu)}\right) \qquad (1) \\ \dot{L}_j^{(-)} &= \sum_n a_{nj} y_n \\ \dot{P}_j^{(+)}(\mu) &= \dot{P}_j(\mu) - \mu_j \ddot{P}_j(\mu) \\ \dot{P}_j^{(-)}(\mu) &= \mu_j \ddot{P}_j(\mu). \end{aligned}$$

This work was supported in part by NIH grant CA-60711 and DOE grant DE-FG02-87ER60561.

Provided ϕ is a strictly convex function, one sees that $\dot{L}_j^{(-)} + \beta \dot{P}_j^{(-)}(\mu) > 0$. This suggests using the following iteration:

$$\begin{aligned} \mu_j^{i+1} &= \mu_j^i + \omega^i \frac{\mu_j^i}{\dot{L}_j^{(-)} + \beta \dot{P}_j^{(-)}(\mu^i)} \frac{\partial}{\partial \mu_j}\Phi(\mu)\bigg|_{\mu=\mu^i} \\ &= \mu_j^i + \omega^i \mu_j^i \frac{\dot{L}_j^{(+)}(\mu) - \beta \dot{P}_j^{(+)}(\mu) - \dot{L}_j^{(-)} - \beta \dot{P}_j^{(-)}(\mu)}{\dot{L}_j^{(-)} + \beta \dot{P}_j^{(-)}(\mu^i)} \\ &= \mu_j^i \left(1 - \omega^i + \omega^i \frac{\dot{L}_j^{(+)}(\mu^i) - \beta \dot{P}_j^{(+)}(\mu^i)}{\dot{L}_j^{(-)} + \beta \dot{P}_j^{(-)}(\mu^i)}\right). \qquad (2) \end{aligned}$$

Since $\dot{L}_j^{(-)} + \beta \dot{P}_j^{(-)}(\mu^i)$ is positive, if $\mu_j^i > 0$ then $\mu_j^{i+1} > 0$ provided $\omega^i \leq 1$. Since this recursion does not guarantee monotone increases in the objective $\Phi(\mu^i)$, we begin each iteration with $\omega^i = 1$, and then if necessary halve it until $\Phi(\mu^{i+1}) > \Phi(\mu^i)$. This has never been necessary in our experiments with this algorithm to date. We refer to (2) as the *gradient algorithm*.

II. EM Algorithm

Lange and Carson [6] proposed an EM algorithm for transmission tomography using a Taylor series approximation for the M-step. Ollinger [7] reported that the EM algorithm did not completely converge with this approximation, and proposed a 1-D Newton's method for the M-step in the pure maximum likelihood case (no smoothness penalty). When one includes a smoothness penalty, the M-step of Ollinger's (or Lange and Carson's) method would require simultaneous solution of p coupled equations. Lange [1] has adapted a clever convexity method due to De Pierro [8, 9] to the M-step of [6]. We have adapted this same convexity method to the M-step of Ollinger [7]. For completeness we summarize the approach here; see [6, 7] for additional details.

Define the following function:

$$Q_{\mathrm{EM}}(\mu;\mu^i) = \sum_n \sum_{j=1}^{p} \bar{X}_{nj} \log e^{-a_{nj}\mu_j} + (\bar{X}_{n,j-1} - \bar{X}_{nj}) \log(1 - e^{-a_{nj}\mu_j}),$$

where

$$\bar{X}_{nj} = \gamma^i_{n,j-1} - \gamma^i_{nj} + y_n \left(\frac{\gamma^i_{np}}{\gamma^i_{np} + r_n} \right)$$

and

$$\gamma^i_{nj} = b_n \prod_{k=1}^{j} e^{-a_{nk}\mu^i_k}.$$

The function Q_{EM} corresponds to the conditional log-likelihood of a complete-data space for transmission tomography, and as such one can show [6, 10] that

$$L(\mu) - L(\mu^i) \geq Q_{\mathrm{EM}}(\mu;\mu^i) - Q_{\mathrm{EM}}(\mu^i;\mu^i).$$

Therefore, if we define

$$\Phi_{\mathrm{EM}}(\mu;\mu^i) = Q_{\mathrm{EM}}(\mu;\mu^i) - \beta P^\star(\mu;\mu^i), \qquad (3)$$

then one can show that

$$\Phi(\mu) - \Phi(\mu^i) \geq \Phi_{\mathrm{EM}}(\mu;\mu^i) - \Phi_{\mathrm{EM}}(\mu^i;\mu^i),$$

so by choosing μ^{i+1} to maximize $\Phi_{\mathrm{EM}}(\cdot;\mu^i)$ we ensure monotonic increases in Φ. The convenient aspect of $\Phi_{\mathrm{EM}}(\cdot;\mu^i)$ is that it is a separable function of $\mu_1, \ldots, \mu_p$, so maximizing $\Phi_{\mathrm{EM}}(\cdot;\mu^i)$ requires p separate 1-D maximizations.

Unfortunately, those maximizations do not have a closed form, so following Ollinger [7] we apply Newton's method to each parameter. One can show

$$\begin{aligned}
\left.\frac{\partial}{\partial \mu_j} Q_{\mathrm{EM}}(\mu;\mu^i)\right|_{\mu=\mu^i} &= \dot{L}_j(\mu^i) \\
-\left.\frac{\partial^2}{\partial \mu_j^2} Q_{\mathrm{EM}}(\mu;\mu^i)\right|_{\mu=\mu^i} &= \sum_n a^2_{nj} \frac{\gamma^i_{nj}}{1 - \exp(-a_{nj}\mu^i_j)} \\
\left.\frac{\partial}{\partial \mu_j} P^\star(\mu;\mu^i)\right|_{\mu=\mu^i} &= \dot{P}_j(\mu^i) \\
\left.\frac{\partial^2}{\partial \mu_j^2} P^\star(\mu;\mu^i)\right|_{\mu=\mu^i} &= 2\ddot{P}_j(\mu^i).
\end{aligned}$$

Combine the above with (3) to yield the iteration:

$$\mu^{i+1}_j = \left[\mu^i_j + \omega^i \frac{\frac{\partial}{\partial \mu_j} \Phi_{\mathrm{EM}}(\mu^i;\mu^i)}{-\frac{\partial^2}{\partial \mu_j^2} \Phi_{\mathrm{EM}}(\mu^i;\mu^i)} \right]_+, \qquad (4)$$

where ω^i is chosen using the halving search to assure monotonicity, starting with $\omega^i = 1$. The key difference between the coordinate-ascent update in [2] and (4) is that the latter uses a *simultaneous* update, and as such it is more amenable to parallel implementations.

REFERENCES

[1] K. Lange and J. A. Fessler. Globally convergent algorithms for maximum a posteriori transmission tomography. *IEEE Tr. Im. Proc.*, 4(10):1430–8, October 1995.

[2] J. A. Fessler. Hybrid Poisson/polynomial objective functions for tomographic image reconstruction from transmission scans. *IEEE Tr. Im. Proc.*, 4(10):1439–50, October 1995.

[3] J. A. Fessler. Statistical image reconstruction methods for transmission tomography. In M. Sonka and J. Michael Fitzpatrick, editors, *Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis*, pages 1–70. SPIE, Bellingham, 2000.

[4] K. Lange, M. Bahn, and R. Little. A theoretical study of some maximum likelihood algorithms for emission and transmission tomography. *IEEE Tr. Med. Im.*, 6(2):106–14, June 1987.

[5] K. Lange. Convergence of EM image reconstruction algorithms with Gibbs smoothing. *IEEE Tr. Med. Im.*, 9(4):439–46, December 1990. Corrections, T-MI, 10:2(288), June 1991.

[6] K. Lange and R. Carson. EM reconstruction algorithms for emission and transmission tomography. *J. Comp. Assisted Tomo.*, 8(2):306–16, April 1984.

[7] J. M. Ollinger. Maximum likelihood reconstruction of transmission images in emission computed tomography via the EM algorithm. *IEEE Tr. Med. Im.*, 13(1):89–101, March 1994.

[8] A R De Pierro. On the relation between the ISRA and the EM algorithm for positron emission tomography. *IEEE Tr. Med. Im.*, 12(2):328–33, June 1993.

[9] A R De Pierro. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *IEEE Tr. Med. Im.*, 14(1):132–137, March 1995.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc. Ser. B*, 39(1):1–38, 1977.