

Accelerated Monotonic Algorithms for Transmission Tomography

Hakan Erdođan and Jeffrey A. Fessler
 4415 EECS Dept., Univeristy of Michigan, Ann Arbor, MI 48109
 email: erdogan@umich.edu

Abstract

We present a framework for designing fast and monotonic algorithms for transmission tomography penalized-likelihood image reconstruction. The new algorithms are based on paraboloidal surrogate functions for the log-likelihood. Due to the form of the log-likelihood function, it is possible to find low curvature surrogate functions that guarantee monotonicity. Unlike previous methods, the proposed surrogate functions lead to monotonic algorithms even for the nonconvex log-likelihood that arises due to background events such as scatter and random coincidences. The gradient and the curvature of the likelihood terms are evaluated only once per iteration. Since the problem is simplified, the CPU time per iteration is less than that of current algorithms which directly minimize the objective, yet the convergence rate is comparable. The simplicity, monotonicity and speed of the new algorithms are quite attractive. The convergence rates of the algorithms are demonstrated using real PET transmission scans.

1 Introduction

Emission tomography systems require attenuation correction for quantitatively accurate image reconstruction. Transmission scans are performed to estimate the attenuation maps for correction. The accuracy of this correction is very important in emission tomography [1]. Statistical methods provide a valuable tool to reconstruct attenuation maps in photon limited tomography applications. Penalized likelihood is an appealing reconstruction method since it provides an easy means to regularize the problem. Using the Poisson log-likelihood eliminates the negative bias which occurs in the weighted least squares and conventional methods [2].

However, up to now, no practically realizable monotonic (or convergent) algorithm has been found which would optimize the penalized likelihood problem when the objective is not convex. The objective is not convex when there are “background” counts in the data. This is unescapable in PET and SPECT, due to the accidental co-

incidences in PET and emission crosstalk in SPECT. The assumption of no background counts can only be valid in X-ray CT. In this paper, we present an algorithm which is guaranteed to be monotonic even in the non-convex objective function case. This algorithm depends on paraboloidal surrogate functions to the log-likelihood which transform the problem into a simpler optimization problem at each iteration. To obtain the paraboloid, we find 1-D parabolic functions that are tangent to and lie above each of the terms in the log-likelihood, similar to Huber’s method for robust linear regression [3].

2 The Problem

The measurements in a photon limited application such as PET and SPECT are well-modeled as Poisson random variables. It is realistic to assume the following model:

$$y_i \sim \text{Poisson}\{b_i e^{-[\mathbf{A}\mu]_i} + r_i\}, \quad i = 1, \dots, N$$

where N is the number of measurements. μ_j for $j = 1 \dots p$, is the average linear attenuation coefficient in voxel j and p denotes the number of voxels. $[\mathbf{A}\mu]_i = \sum_{j=1}^p a_{ij} \mu_j$ represents the line integral of the attenuation map μ and $\mathbf{A} = \{a_{ij}\}$ is the $N \times p$ system matrix. r_i is the mean number of background events, b_i is the blank scan count and y_i represents the number of coincident transmission events in the i^{th} detector pair. We assume in this paper that $\{b_i\}$, $\{r_i\}$ and $\{a_{ij}\}$ are known nonnegative constants.

The log-likelihood can be expressed in the following form:

$$-L(\mu) = \sum_{i=1}^N h_i([\mathbf{A}\mu]_i), \quad (1)$$

where:

$$h_i(l) = (b_i e^{-l} + r_i) - y_i \log(b_i e^{-l} + r_i). \quad (2)$$

Penalized likelihood image reconstruction formulation is given below:

$$\hat{\mu} = \arg \min_{\mu \geq 0} \Phi(\mu), \quad \Phi(\mu) = -L(\mu) + \beta R(\mu). \quad (3)$$

The roughness penalty included in the objective function (in a very general form) can be given by [4]:

$$R(\mu) = \sum_{k=1}^K \psi_k([\mathbf{C}\mu]_k),$$

where ψ_k are potential functions penalizing the deviations from the set of equations $\mathbf{C}\mu \approx 0$ for $k = 1 \dots K$ where K is number of such equations. The β in equation (3) is a parameter which controls the level of smoothness in the final reconstructed image. For more explanation of the penalty function, see [4].

The objective function defined in (3) is not convex when there are nonzero background counts ($r_i \neq 0$) in the data. In this realistic case, there is no guarantee that there is a single global minimum. In this paper we introduce an algorithm that is monotonic even when the objective is not convex. The new approach is based on successive paraboloidal surrogate functions.

3 Paraboloidal Surrogates Algorithms

The penalized-likelihood objective function $\Phi(\mu)$ has a complex form that precludes analytical minimization. Therefore, we apply an iterative approach, where at each iteration we would like to find a ‘‘surrogate’’ function $\phi(\mu; \mu^n)$ which is easier to minimize or to monotonically decrease than $\Phi(\mu)$. To ensure monotonicity, a decrease in the surrogate should lead to a decrease in the original objective function. Consider surrogate functions which satisfy the following properties:

1. $\phi(\mu^n; \mu^n) = \Phi(\mu^n)$
2. $\nabla \phi(\mu; \mu^n)|_{\mu=\mu^n} = \nabla \Phi(\mu)|_{\mu=\mu^n}$,
3. $\phi(\mu; \mu^n) \geq \Phi(\mu)$ for $\mu \geq 0$.

A function which satisfies these criteria can be easily shown to satisfy the following monotonicity condition:

$$\Phi(\mu) - \Phi(\mu^n) \leq \phi(\mu; \mu^n) - \phi(\mu^n; \mu^n), \quad \forall \mu \geq 0. \quad (5)$$

The EM algorithm [5] provides a statistical method for constructing surrogate functions $\phi(\mu; \mu^n)$ satisfying the above conditions. However, in the transmission tomography problem, the EM surrogate is difficult to minimize and leads to slow convergence. In this paper, we construct a simpler surrogate using ordinary calculus rather than statistical techniques.

Let μ^n denote the estimate of the attenuation map at the n th iteration and let $l_i^n = [\mathbf{A}\mu^n]_i$ denote the corresponding estimate of the i th line integral of the attenuation coefficient. We choose the following quadratic form for the surrogate functions q_i :

$$q_i(l; l_i^n) \triangleq h_i(l_i^n) + \dot{h}_i(l_i^n)(l - l_i^n) + \frac{1}{2}c_i(l_i^n)(l - l_i^n)^2, \quad (6)$$

where $c_i(l_i^n)$ is the curvature of the parabola q_i . This construction ensures that $q_i(l_i^n; l_i^n) = h_i(l_i^n)$ and $\dot{q}_i(l_i^n; l_i^n) = \dot{h}_i(l_i^n)$ similar to equation (4). To ensure monotonicity, we must choose the curvatures to satisfy the following inequality at each iteration:

$$h_i(l) \leq q_i(l; l_i^n), \quad \text{for } l \geq 0. \quad (7)$$

After determining the parabolas, the following function can be easily verified to be a global surrogate function for the objective $\Phi(\mu)$ which satisfies the properties in (4):

$$\phi(\mu; \mu^n) = Q(\mu; \mu^n) + \beta R(\mu), \quad (8)$$

where

$$\begin{aligned} Q(\mu; \mu^n) &\triangleq \sum_{i=1}^N q_i([\mathbf{A}\mu]_i; l_i^n) \\ &\equiv \mathbf{d}_h(l^n)' \mathbf{A}(\mu - \mu^n) + \\ &\quad \frac{1}{2}(\mu - \mu^n)' \mathbf{A}' \mathbf{D}(c_i) \mathbf{A}(\mu - \mu^n), \end{aligned} \quad (9)$$

where the column vector $\mathbf{d}_h(l^n) \triangleq [\dot{h}_i(l_i^n)]_{i=1}^N$, \mathbf{x}' denotes the transpose of \mathbf{x} , and $\mathbf{D}(c_i)$ is the $N \times N$ diagonal matrix with diagonal entries $c_i(l_i^n)$ for $i = 1, \dots, N$.

The surrogate function $\phi(\mu; \mu^n)$ in equation (8) consists of the sum of a paraboloid (or a quadratic form) and the convex penalty term. An algorithm that decreases the function ϕ will also monotonically decrease the objective function if the inequality in (7) holds. In the following sections, we will elaborate on different choices of curvatures and monotonic coordinate descent algorithm applied to the surrogate.

3.1 Curvatures

The conventional ‘‘Newton’’ choice for the curvatures $c_i(l_i^n)$ is simply $c_i(l_i^n) = \ddot{h}_i(l_i^n)$. However, this choice does not ensure monotone convergence. The following two choices for curvatures do ensure monotonicity.

3.1.1 Maximum Curvature

A simple choice for $c_i(l_i^n)$ that ensures monotonicity (7) is the maximum second derivative in the feasible region (from mean value theorem, see for example [6], page 228). The feasible region for the projections is $[0, \infty)$ due to the nonnegativity constraint. Hence,

$$c_i(l_i^n) = \max_{l \in [0, \infty)} \{\ddot{h}_i(l)\}$$

is guaranteed to satisfy (7). We show in [7] that the closed form expression for $c_i(l_i^n)$ is:

$$c_i(l_i^n) = [\ddot{h}_i(0)]_+ = \left[\left(1 - \frac{y_i r_i}{(b_i + r_i)^2} \right) b_i \right]_+ \quad (11)$$

where $[x]_+ = x$ for $x > 0$ and zero otherwise.

However, this ‘‘maximum curvature’’ choice for $c_i(l_i^n)$ is very conservative and results in slow convergence. It is intuitive that smaller $c_i(l_i^n)$ values will lead to faster convergence, as analyzed in [7]. This is due to the fact that smaller curvatures mean wider parabolooids and longer step sizes. Next, we determine the smallest curvature that satisfies the monotonicity condition (7).

3.1.2 Optimum Curvature

The optimal choice for the curvatures is the solution to the following constrained optimization problem for each i : $c_i(l_i^n) = \min \{c \geq 0 : h_i(l) \leq h_i(l_i^n) + \dot{h}_i(l_i^n)(l - l_i^n) + 1/2c(l - l_i^n)^2 \forall l \geq 0\}$. This curvature yields the fastest convergence rate while still guaranteeing monotonicity.

By exploiting the properties of the marginal negative log-likelihood functions for each projection (h_i) in transmission tomography, we show in [7] that the paraboloid with the following curvature satisfies the optimality condition:

$$c_i(l_i^n) = \begin{cases} \left[\frac{2 \frac{h_i(0) - h_i(l_i^n) + \dot{h}_i(l_i^n)(l_i^n)}{(l_i^n)^2} \right]_+, & l_i^n > 0, \\ \left[\ddot{h}_i(0) \right]_+, & l_i^n = 0. \end{cases} \quad (12)$$

There are some numerical issues with the computation of these curvatures which are addressed in [7]. Next, we consider a set of fixed curvatures which have some computational advantages.

3.1.3 Nonmonotonic Precomputed Curvature

By relaxing the monotonicity requirement, we can develop faster yet ‘‘almost always’’ monotonic algorithms. We can do this by choosing curvatures $c_i(l_i^n)$ in equation (6) such that

$$h_i(l) \approx q_i(l; l_i^n),$$

rather than requiring the inequality (7). In this case, the parabolooids are quadratic ‘‘approximations’’ to the log-likelihood function at each iteration. A reasonable choice for the curvatures in that case is:

$$c_i = \ddot{h}_i \left(\log \frac{b_i}{y_i - r_i} \right) = (y_i - r_i)^2 / y_i. \quad (13)$$

These curvatures c_i in (13) are close approximations to the second derivative of h_i functions at the projection values $\mathbf{A}\hat{\mu}$ where $\hat{\mu}$ is the solution to the penalized-likelihood problem [8]. Since these curvatures do not depend on the iterations (like ‘‘maximum curvature’’), some computation can be saved by precomputing the denominator terms in the coordinate descent update.

3.2 Monotonic Coordinate Descent

Once the curvatures are determined, the penalized likelihood problem is transformed into a problem with a quadratic likelihood plus the convex penalty function. We use one iteration of coordinate descent algorithm to decrease the surrogate function, and this results in a simple monotonic algorithm. For the penalty part, we use Huber’s iterative method [3, 4] (see also [7]). Since the likelihood part is quadratic, this coordinate descent method is guaranteed to decrease the surrogate function. The update equation is as follows:

$$\hat{\mu}_j^{\text{new}} = \left[\hat{\mu}_j - \frac{\dot{Q}_j^n(\hat{\mu}) + \beta \dot{R}_j(\hat{\mu})}{d_j^n + \beta \hat{p}_j} \right]_+ \quad (14)$$

where:

$$\dot{Q}_j^n(\hat{\mu}) = \sum_{i=1}^N a_{ij} \dot{q}_i(\hat{l}_i), \quad d_j^n \triangleq \sum_{i=1}^N a_{ij}^2 c_i(l_i^n),$$

$$\dot{R}_j(\hat{\mu}) = \sum_{k=1}^K c_{kj} \dot{\psi}_k([\mathbf{C}\hat{\mu}]_k), \quad \hat{p}_j \triangleq \sum_{k=1}^K c_{kj}^2 \omega_{\psi_k}([\mathbf{C}\hat{\mu}]_k),$$

where $\omega_{\psi}(t) \triangleq \dot{\psi}(t)/t$. To prove monotonicity, $\psi_k(\cdot)$ have to be symmetric, differentiable, convex and $\omega_{\psi}(0)$ should be finite and nonzero and $\omega_{\psi}(t)$ should be non-increasing for $t \geq 0$ [3].

This is an update that monotonically decreases the value of $\phi(\cdot; \mu^n)$ and consequently the value of $\Phi(\cdot)$. One iteration is finished when all pixels are updated via (14) in a sequential order. We call this method the Paraboloidal Surrogates Coordinate Descent (PSCD) method. A coarse outline of the algorithm is given in Table 1. For computational considerations and the detailed algorithm flow table, see [7].

4 Results

To assess the effectiveness and speed of the new PS algorithms, we present results using real PET data. We acquired a 15-hour blank scan (b_i ’s) and a 12-min transmission scan data (y_i ’s) using a Siemens/CTI ECAT EXACT 921 PET scanner with rotating rod sources for transmission. The phantom used was an anthropomorphic thorax phantom (Data Spectrum, Chapel Hill, NC). The projection space was 160 radial bins and 192 angles, and the reconstructed images were 128×128 with 4.5 mm. pixels. The system matrix $\{a_{ij}\}$ was computed by using 6 mm. wide strip integrals with 3 mm. spacing, which roughly approximates the system geometry [2].

In PL reconstructions, we used the following penalty function:

$$R(\mu) = \frac{1}{2} \sum_{j=1}^p \sum_{k \in \mathcal{N}_j} w_{jk} \psi(\mu_j - \mu_k)$$

Initialize $\hat{\mu}, \hat{l}_i$.
for each iteration
 Compute $\hat{q}_i = \hat{h}_i$ and c_i 's at current iterate
 for each pixel j
 Compute $\hat{Q}_j^n(\hat{\mu}), d_j^n$, set $\hat{\mu}^{\text{old}} = \hat{\mu}$.
 for a couple subiterations
 Compute $R_j(\hat{\mu}), \hat{p}_j$,
 $\hat{Q}_j^n(\hat{\mu}) = \hat{Q}_j^n(\hat{\mu}^{\text{old}}) + d_j^n(\hat{\mu}_j - \hat{\mu}_j^{\text{old}})$.
 Update $\hat{\mu}_j$ using (14)
 end
 Update $\hat{q}_i := \hat{q}_i + a_{ij}c_i(\hat{\mu}_j - \hat{\mu}_j^{\text{old}})$
end
Update $\hat{l}_i := \hat{l}_i + \frac{\hat{q}_i - \hat{h}_i}{c_i}$
end

Table 1: Coarse outline of PSCD algorithm.

with the edge-preserving nonquadratic potential function that was introduced in [9]

$$\psi(x) = \delta^2 [|x/\delta| - \log(1 + |x/\delta|)].$$

Fig. 1 shows that the proposed PSCD algorithms increased the penalized-likelihood almost as fast as the coordinate descent algorithm (CD) [10] applied to the original objective per iteration. It shows that the surrogate paraboloids closely approximate the original log-likelihood. More importantly, in Fig. 2 the PSCD algorithms are seen to be much faster than coordinate descent in terms of the actual CPU time. This reduction is due to the lack of exponentiations in updating the gradient terms in PSCD algorithms. The ‘‘maximum curvature’’ method introduced in Section 3.1 precomputes the denominator terms (d_j^n) for the likelihood part since $c_i(l_i^n)$'s do not depend on the iterations. However, these $c_i(l_i^n)$'s are much larger than the optimal curvatures, so more iterations are required for PS,M,CD than PS,O,CD to converge. We also compared the PSCD algorithms to the general purpose constrained Quasi-Newton algorithm (LBFGS) [11], and the functional substitution coordinate descent (FSCD) [12, 13] algorithm in Figures 1 and 2. Although the LBFGS algorithm takes about 25% less CPU time (0.88 seconds) per iteration than PSCD algorithms, it did not converge as fast as the proposed algorithms. FSCD algorithm (which is guaranteed to be monotonic when $r_i = 0$) had the largest CPU time per iteration due to extra exponentiations needed. Table 2 compares the number of iterations and CPU seconds required (in a DEC 600 5-333 MHz workstation) to minimize the objective function for different methods. The CPU times per iteration are also tabulated. For comparison purposes, a single forward and backprojection requires about 0.78 CPU seconds. If a monotonic algorithm is re-

Real data, $r_i \neq 0$	monotonic		nonmonotonic			
	PS,M,CD	PS,O,CD	FSCD	CD	PS,P,CD	GD _{3x3}
iterations for convergence	18	12	11	11	11	14
CPU s for convergence	23	17	56	44	15	18
CPU s per iteration	1.2	1.3	4.9	3.8	1.2	1.1

Table 2: Comparison of CPU times and iterations for the PS algorithm versus FS, CD and GCD methods. Convergence in this table means $\Phi(\mu^0) - \Phi(\mu^n) > 0.999 [\Phi(\mu^0) - \Phi(\mu^*)]$ where $\Phi(\mu^*)$ is the smallest objective value obtained.

quired, the PSCD algorithm with the optimal curvature (PS,O,CD) is the fastest algorithm. Among the nonmonotonic algorithms, another PS method, PSCD with precomputed curvatures (PS,P,CD) is the fastest. It converged in about 15 seconds with the real data used. The CPU time per iteration is the same as PSCD with maximum curvature (PS,M,CD) since they both precompute the denominator (d_j^n) terms, but this method decreases the objective faster since the curvatures are smaller. Nevertheless PS,P,CD is not guaranteed to be monotonic.

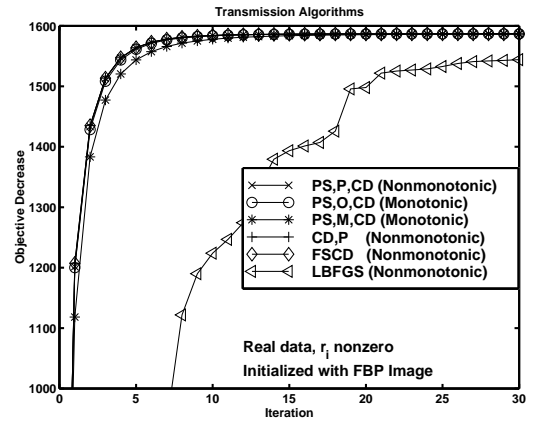


Figure 1: Comparison of objective function decrease $\Phi(\mu^0) - \Phi(\mu^n)$ versus iteration number n of PSCD methods with coordinate descent, FSCD and LBFGS methods for real phantom data.

5 Conclusion

We have introduced a new class of algorithms for minimizing penalized-likelihood objective functions for trans-

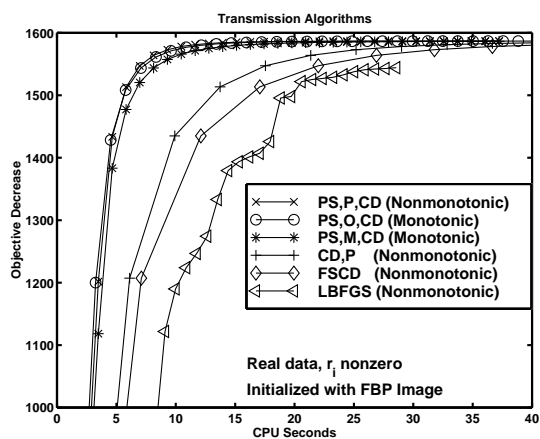


Figure 2: Same as Figure 1, but x-axis is CPU seconds on a DEC AlphaStation 600 5-333 MHz.

mission tomography. The algorithms are monotonic even with the nonconvex objective function. In the strictly convex case, the proposed algorithms are guaranteed to converge to the global minimum by a proof similar to that in [14].

The algorithms we introduced are simple, easy to understand, fast and monotonic. The simplicity in part is due to the additive form of (1), which is a direct consequence of independent measurements. Since this algorithm is simple, it might replace the use of FBP in the clinic.

In our opinion, the PS,O,CD algorithm supersedes all of our previous methods [2, 8], and is our recommended algorithm for penalized-likelihood transmission tomography. The PS,P,CD algorithm is a faster but nonmonotonic alternative which can be used for noncritical applications. As a possible compromise, one can run PS,P,CD algorithm and check the objective function $\Phi(\mu)$ after each iteration to verify that it has decreased. If the objective does not decrease (happens very rarely), then PS,O,CD algorithm can be applied to the previous iterate to ensure monotonicity.

References

- [1] S. C. Huang, E. J. Hoffman, M. E. Phelps, and D. E. Kuhl, "Quantitation in positron emission computed tomography: 2 Effects of inaccurate attenuation correction," *J. Comp. Assisted Tomo.*, vol. 3, no. 6, pp. 804–814, December 1979.
- [2] J. A. Fessler, "Hybrid Poisson/polynomial objective functions for tomographic image reconstruction from transmission scans," *IEEE Tr. Im. Proc.*, vol. 4, no. 10, pp. 1439–50, October 1995.
- [3] P. J. Huber, *Robust statistics*, Wiley, New York, 1981.
- [4] J. A. Fessler, "Grouped coordinate descent algorithms for robust edge-preserving image restoration," in *Proc. SPIE 3071, Im. Recon. and Restor. II*, pp. 184–94, 1997.
- [5] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comp. Assisted Tomo.*, vol. 8, no. 2, pp. 306–316, April 1984.
- [6] P. G. Ciarlet, *Introduction to numerical linear algebra and optimisation*, Cambridge, Cambridge, 1982.
- [7] H. Erdoğan and J. A. Fessler, "Fast monotonic algorithms for transmission tomography," *IEEE Tr. Med. Im.*, 1998. Submitted.
- [8] 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.
- [9] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Tr. Med. Im.*, vol. 9, no. 4, pp. 439–446, December 1990. Corrections, June 1991.
- [10] C. A. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Tr. Im. Proc.*, vol. 5, no. 3, pp. 480–92, March 1996.
- [11] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "L-BFGS-B: Fortran subroutine for large-scale bound constrained optimization," Technical Report ?, Dept. EECS, Northwestern Univ., July 1996.
- [12] J. Zheng, S. Saquib, K. Sauer, and C. Bouman, "Functional substitution methods in optimization for Bayesian tomography," *IEEE Tr. Im. Proc.*, March 1997. Submitted to IEEE Tr. Image Proc.
- [13] S. Saquib, J. Zheng, C. A. Bouman, and K. D. Sauer, "Provably convergent coordinate descent in statistical tomographic reconstruction," in *Proc. IEEE Intl. Conf. on Image Processing*, 1996.
- [14] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Tr. Sig. Proc.*, vol. 42, no. 10, pp. 2664–77, October 1994.