Fast Parallelizable Algorithms for Transmission Image Reconstruction

Jeffrey A. Fessler, Edward P. Ficaro, Neal H. Clinthorne, and Kenneth Lange

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

Abstract— This paper presents a new class of algorithms for penalized-likelihood reconstruction of attenuation maps from low-count transmission scans. We derive the algorithms by applying to the transmission log-likelihood a variation of the convexity technique developed by De Pierro for the emission case. The new algorithms overcome several limitations associated with previous algorithms. (1) Fewer exponentiations are required than in the transmission EM algorithm or in coordinate-ascent algorithms. (2) The algorithms intrinsically accommodate nonnegativity constraints, unlike many gradient-based methods. (3) The algorithms are easily parallelizable, unlike coordinate-ascent algorithms and perhaps line-search algorithms. We show that the algorithms converge faster than several alternatives, even on conventional workstations. We give examples from low-count PET transmission scans and from truncated fan-beam SPECT transmission scans.

I. INTRODUCTION

The importance of statistical methods for reconstructing attenuation maps has increased recently due to the widening availability of SPECT systems equipped with transmission sources [1], the necessity of reconstructing 2D attenuation maps for reprojection to form 3D attenuation correction factors in septaless PET [2,3], and the potential for reducing transmission noise in whole body PET images and in other protocols requiring short transmission scans [4]. The filtered backprojection (FBP) method and the data-weighted least-squares method for transmission image reconstruction lead to systematic biases at low counts [5], due to the nonlinearity of the logarithm. To eliminate these biases, one can use statistical methods which require no logarithms [5].

Several reconstruction algorithms based on the Poisson statistical model for transmission scans have appeared recently [6–12,5], all of which converge faster than the original transmission ML-EM algorithm [13]. Nevertheless, these methods are still less than ideal due to one or more of the following reasons.

- The EM algorithms [13,9] and coordinate-ascent algorithms [14,12,5] require at least one exponentiation per nonzero element in the system matrix per iteration, which is a large computational expense.
- Enforcing nonnegativity in gradient-based algorithms [10,11] is possible but somewhat awkward.
- Many algorithms are poorly suited to parallel processors, such as the i860 arrays that pervade septaless PET sites. This is true of coordinate-ascent methods and of algorithms that use line searches, since a line-search step may not parallelize easily.

This paper describes a new class of algorithms for reconstructing attenuation maps from low-count transmission scans. These algorithms are parallelizable, easily accommodate nonnegativity constraints and nonquadratic convex penalties, and require a moderate number of exponentiations. The derivation of these transmission algorithms exploits two ideas underlying recent developments in algorithms for emission tomography: updating the parameters in groups [15,16], and the convexity technique of De Pierro [17,18]. Integrating these two ideas leads to new algorithms that converge quickly with less computation than previous methods.

II. PROBLEM

For brevity we consider the transmission measurement model without additive background events (random coincidences, scatter, emission crosstalk, etc.), although the method can be extended to include those effects. We assume

$$y_i \sim \text{Poisson}\{b_i \exp(-\langle a_{i\cdot}, \theta_{\text{true}} \rangle)\}, \qquad (1)$$

where $\langle a_{i\cdot}, \theta \rangle = \sum_j a_{ij}\theta_j$, represents the ith "line integral," y_i denotes the transmission measurement of the ith detector, b_i denotes the ith blank scan measurement, θ_j denotes the unknown attenuation coefficient in the jth voxel, and the a_{ij}'s are the transmission system model. We assume $\{b_i\}$ and $\{a_{ij}\}$ are known.

The transmission log-likelihood is [13]:

$$L(\theta; y) = \sum_i h_i(\langle a_{i\cdot}, \theta \rangle) \qquad (2)$$

where

$$h_i(l) = y_i \log(b_i e^{-l}) - b_i e^{-l}. \qquad (3)$$

Note that each h_i is a concave function over all of $\mathbb{R}$. The algorithms developed below apply to any problem of the form (2) with concave h_i, including weighted least squares.

The goal is to compute a penalized-likelihood estimate $\hat{\theta}(y)$ of θ, defined by

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

where the objective includes a roughness penalty

$$R(\theta) = \sum_j \frac{1}{2} \sum_k w_{jk} \psi(\theta_j - \theta_k), \qquad (5)$$

where $w_{jk} = 1$ for horizontal and vertical neighboring pixels and is zero otherwise. For concreteness, in this paper we have used one of the penalties in [7]:

$$\psi(x) = \delta^2 \left[|x/\delta| - \log(1 + |x/\delta|) \right], \qquad (6)$$

which approaches $\psi(x) = x^2/2$ as $\delta \to \infty$, but provides a degree of edge preservation for finite δ. Since

$$\dot{\psi}(x) = \frac{d}{dx}\psi(x) = \frac{x}{1 + |x/\delta|}$$

implies $|\dot{\psi}(x)| < \delta$, this potential function has bounded influence. The derivative of $\psi(\cdot)$ requires no transcendental functions, which speeds computation. Since ψ is strictly convex and $L(\cdot, y)$ is concave, the objective Φ is strictly concave under mild conditions on $\mathbf{A}$. This concavity is central to the development of the algorithms below.

Direct maximization of (4) is intractable, so one must use iterative algorithms. Generic numerical methods such as steepest ascent do not exploit the specific structure of Φ, nor do they easily accommodate nonnegativity constraints. Thus for fastest convergence, one must seek algorithms tailored to this problem. Relevant properties of L include:

- $L(\theta; y)$ is a sum of concave functions $h_i(\cdot)$.
- The arguments of the functions $h_i(\cdot)$ are inner products.
- The inner product coefficients are all nonnegative.

These properties suggest the use of Jensen's inequality.

III. Algorithms

As shown by frequency domain analysis in [19], sequential updates such as coordinate ascent converge very rapidly for tomographic reconstruction. Unfortunately, coordinate ascent requires a large number of exponentiations for transmission tomography. Consider the partial derivative of the log-likelihood with respect to the jth pixel value:

$$\dot{L}_j(\theta) = \frac{\partial}{\partial \theta_j} L(\theta; y) = \sum_i a_{ij} \left[1 - \frac{y_i}{\bar{y}_i(\theta)} \right] b_i e^{-\langle a_i, \theta^n \rangle}, \quad (7)$$

where $\bar{y}_i(\theta) = b_i \exp(-\langle a_i, \theta \rangle)$ (see Eqn. (8) of [5]). In a coordinate ascent algorithm, one must evaluate $\dot{L}_j(\theta^n)$ at the current image estimate θ^n. Since $\langle a_i, \theta \rangle$ changes immediately after each pixel is updated, from (7) each iteration requires M exponentiations, where M is the number of nonzero a_{ij}'s. At the other extreme, the transmission scaled-gradient algorithm [11] updates all pixels simultaneously. Thus the terms in (7) can be computed simultaneously before updating the pixels, so only N exponentiations are required, where N is the number of rays. Typically N is two orders of magnitude smaller than M. In other words, there is an "economy of scale" in terms of computation by updating all pixels simultaneously[2]. However, simultaneous updates lead to slow convergence [5].

Rather than updating *all* pixels simultaneously, we propose to update only certain *groups of pixels* simultaneously. If there are G groups of pixels, then only NG exponentiations are needed. On the other hand, if the pixels in each group are well-separated spatially, then we anticipate that

they will be fairly decoupled, so the simultaneous update will not suffer from slow convergence. The results below confirm this intuition.

Let $\mathcal{S}$ be a subset of the pixels $\{1, \ldots, p\}$, let $\tilde{\mathcal{S}}$ be its complement[3], and let $p_{\mathcal{S}}$ be the cardinality of $\mathcal{S}$. Then at the nth iteration we update $\theta_{\mathcal{S}}$ while holding $\theta_{\tilde{\mathcal{S}}}^n$ fixed [15]. Unfortunately it is even too difficult to maximize $\Phi(\theta_{\mathcal{S}}, \theta_{\tilde{\mathcal{S}}}^n)$ over $\theta_{\mathcal{S}}$ directly, so we will settle for finding an approach that chooses a $\theta_{\mathcal{S}}^{n+1}$ that at least provides *monotonic* increases in the objective function:

$$\Phi(\theta_{\mathcal{S}}^{n+1}, \theta_{\tilde{\mathcal{S}}}^n) \geq \Phi(\theta_{\mathcal{S}}^n, \theta_{\tilde{\mathcal{S}}}^n) = \Phi(\theta^n).$$

To assure monotonicity, we use a generalization of De Pierro's *optimization transfer* idea [17, 18], and substitute a *surrogate function* $\phi(\theta_{\mathcal{S}}; \theta^n)$ with a corresponding *region of monotonicity* $\mathcal{R}_{\mathcal{S}} \subseteq \mathbb{R}^{p_s}$ that must satisfy:

$$\Phi(\theta_{\mathcal{S}}, \theta_{\tilde{\mathcal{S}}}^n) - \Phi(\theta^n) \geq \phi(\theta_{\mathcal{S}}; \theta^n) - \phi(\theta_{\mathcal{S}}^n; \theta^n) \quad \forall \theta_{\mathcal{S}} \in \mathcal{R}_{\mathcal{S}}. \quad (8)$$

The SAGE-like update [15, 16] then looks like:

$$\theta_{\mathcal{S}}^{n+1} = \arg \max_{\theta_{\mathcal{S}} \in \mathcal{R}_{\mathcal{S}}} \phi(\theta_{\mathcal{S}}; \theta^n), \quad (9)$$
$$\theta_j^{n+1} = \theta_j^n, \quad j \in \tilde{\mathcal{S}}.$$

The condition (8) ensures immediately that the iterates produced by the above generic algorithm monotonically increase the objective: $\Phi(\theta^{n+1}) \geq \Phi(\theta^n)$.

We restrict attention here to additively separable surrogate functions $\phi(\cdot; \theta^n)$ satisfying

$$\phi(\theta_{\mathcal{S}}; \theta^n) = \sum_{j \in \mathcal{S}} \phi_j(\theta_j; \theta^n).$$

We use a modification of De Pierro's method [17, 18] to choose the ϕ_j's, rather than the EM approach of [15, 16]. Note that

$$\langle a_i, [\theta_{\mathcal{S}}, \theta_{\tilde{\mathcal{S}}}^n] \rangle = \sum_{j \in \mathcal{S}} \alpha_{ij} \frac{a_{ij}}{\alpha_{ij}} (\theta_j - \theta_j^n) + \langle a_i, \theta^n \rangle \quad (10)$$

for any choice[4] of α_{ij} that satisfies $\sum_{j \in \mathcal{S}} \alpha_{ij} = 1$. In particular, in this paper we define

$$\alpha_{ij} = a_{ij} / \sum_{k \in \mathcal{S}} a_{ik}. \quad (11)$$

Then by concavity of $h_i(\cdot)$ and since $\alpha_{ij} \geq 0$:

$$L([\theta_{\mathcal{S}}, \theta_{\tilde{\mathcal{S}}}^n]; y) = \sum_i h_i(\langle a_i, [\theta_{\mathcal{S}}, \theta_{\tilde{\mathcal{S}}}^n] \rangle) \geq \sum_{j \in \mathcal{S}} Q_j(\theta_j; \theta^n),$$

where using (10):

$$Q_j(\theta_j; \theta^n) = \sum_i \alpha_{ij} h_i \left(\frac{a_{ij}}{\alpha_{ij}} (\theta_j - \theta_j^n) + \langle a_i, \theta^n \rangle \right). \quad (12)$$

---

[2]Even if the exponentiations are computed approximately, using table lookups for example, the ratio between N and M remains unchanged.

[3]In a grouped coordinate ascent method, $\mathcal{S}$ varies with n. To simplify notation, we leave this dependence implicit.
[4]We assume $\alpha_{ij} = 0$ if and only if $a_{ij} = 0$ so that (10) is well defined.

Assuming the groups are chosen so that no two neighboring pixels are in the same group[5], then we can define[6]

$$\phi_j(\theta_j; \theta^n) = Q_j(\theta_j; \theta^n) - \beta \sum_k w_{jk} \psi(\theta_j - \theta_k^n). \quad (13)$$

Each ϕ_j only depends on one θ_j, so since $\mathcal{R}_\mathcal{S} = \mathbb{R}^p$ in this problem due to the concavity of $h_i(\cdot)$, the maximization step in (9) reduces to separate 1D maximizations. Thus (9) becomes the parallelizable operations:

$$\theta_j^{n+1} = \arg\max_{\theta_j \geq 0} \phi_j(\theta_j; \theta^n), \quad j \in \mathcal{S}. \quad (14)$$

A. Convergence

It is fairly straightforward to apply the general convergence proof in [15] to prove that the sequence of estimates $\{\theta^n\}$ produced by the above algorithm monotonically increases $\Phi(\cdot)$ and converges from any starting image to the unique global maximizer of Φ subject to $\theta \geq 0$. There are a few caveats that must be considered however. When using finite precision arithmetic, monotonicity often does not hold exactly once the sequence gets nears the maximum. Also, usually one will not perform exact 1D maximizations as implied by (14), but rather partial or approximate maximizations (see below). Finally, when one includes additive background effects in the statistical model, the log-likelihood is no longer globally concave [5]. Nevertheless, it is comforting to know that at least under ideal circumstances the convergence is well understood.

B. The Maximization Step

One simple approach to implementing the maximization (14) would be to apply a few iterations of the 1D Newton Raphson method:

$$\theta_j^{\text{work}} = \theta_j^n$$

$$\theta_j^{\text{work}} := \left[\theta_j^{\text{work}} + \frac{\frac{d}{d\theta_j}\phi_j(\theta_j; \theta^n)\big|_{\theta_j = \theta_j^{\text{work}}}}{-\frac{d^2}{d\theta_j^2}\phi_j(\theta_j; \theta^n)\big|_{\theta_j = \theta_j^{\text{work}}}} \right]_+ \quad (15)$$

$$\theta_j^{n+1} = \theta_j^{\text{work}}.$$

Typically the middle step would be repeated a few times. Unfortunately, the partial derivatives of $\phi_j(\cdot; \theta^n)$ are fairly expensive to compute.

To reduce computation, we apply methods from [12] and [5]. For the numerator, we approximate the Q_j function (but not the penalty!) by its second order Taylor series, in a spirit similar to [12]. For the denominator, we use a trick similar to [5] for precomputing an approximation to the second derivative of the Q_j function, and a new trick for the penalty term that exploits its bounded curvature.

[5]If a group contains neighboring pixels, then one can also apply De Pierro's device [17,18] to the penalty function to ensure (8).

[6]Note that the $\frac{1}{2}$ in (5) disappears in (13) since each pair of pixels is counted twice in (5).

The second-order Taylor's approximation about θ_j^n for the $Q_j(\cdot; \theta^n)$ component of the numerator is:

$$Q_j(\theta_j; \theta^n) \approx Q_j(\theta_j^n; \theta^n) + \dot{L}_j(\theta^n)(\theta_j - \theta_j^n) - \frac{d_j(\theta^n)}{2}(\theta_j - \theta_j^n)^2,$$

because from (12) it follows that

$$\frac{d}{d\theta_j}Q_j(\theta_j; \theta^n)\bigg|_{\theta_j = \theta_j^n} = \frac{\partial}{\partial\theta_j}L(\theta)\bigg|_{\theta = \theta^n} = \dot{L}_j(\theta^n),$$

and where (also from (12)):

$$d_j(\theta^n) = -\frac{d^2}{d\theta_j^2}Q_j(\theta_j; \theta^n)\bigg|_{\theta_j = \theta_j^n} = -\sum_i \frac{a_{ij}^2}{\alpha_{ij}}\ddot{h}_i(\langle a_i, \theta^n\rangle).$$

Note that θ^n only enters $d_j(\theta^n)$ through its projections $\langle a_i, \theta^n\rangle$. Thus $d_j(\theta^n)$ is fairly insensitive to θ^n, so we replace $\langle a_i, \theta^n\rangle$ with a precomputed approximation to the ith line integral, such as $\log(b_i/y_i)$. Therefore we replace $d_j(\theta^n)$ with

$$\hat{d}_j = \sum_i \frac{a_{ij}^2}{\alpha_{ij}}\ddot{h}_i(\log(b_i/y_i)), \quad (16)$$

which one can precompute *prior to iterating*. Thus, we replace the numerator of (15) with this approximation:

$$\frac{d}{d\theta_j}\phi_j(\theta_j; \theta^n)\bigg|_{\theta_j = \theta_j^{\text{work}}} \approx$$

$$\dot{L}_j(\theta^n) - \hat{d}_j \cdot (\theta_j^{\text{work}} - \theta_j^n) - \beta\sum_k w_{jk}\dot{\psi}(\theta_j^{\text{work}} - \theta_k^n). \quad (17)$$

For the denominator of (15), note that

$$-\frac{d^2}{d\theta_j^2}\phi_j(\theta_j; \theta^n)\bigg|_{\theta_j = \theta_j^n} = d_j(\theta^n) + \beta\sum_k w_{jk}\ddot{\psi}(\theta_j^n - \theta_k).$$

Since ψ has bounded curvature:

$$\ddot{\psi}(x) = \frac{1}{1 + |x/\delta|} \leq 1, \quad (18)$$

we replace the denominator of (15) with

$$-\frac{d^2}{d\theta_j^2}\phi_j(\theta_j; \theta^n)\bigg|_{\theta_j = \theta_j^{\text{work}}} \approx \hat{d}_j + \beta\sum_k w_{jk}, \quad (19)$$

which can be precomputed as described in [5]. This approach provides a form of built in under-relaxation because of the bounded curvature (18) of ψ.

To summarize, in practice we replace (15) with (17) and (19), and apply 2 or 3 iterations of (15). No forward or backprojections are computed during these subiterations, so they compute quickly. As in [19, 14, 12, 5], we keep a updated "forward projection" $\langle a_i, \theta^n\rangle$ to further save computation when evaluating the "backprojection" step (7).

Note that if one were to use only one subiteration of Newton Raphson[7] then the "maximization step" would have the following form:

$$\theta_{\mathcal{S}}^{n+1} = \theta_{\mathcal{S}}^n + \mathbf{D}^{-1}\nabla_{\theta_{\mathcal{S}}}^T\Phi(\theta^n), \qquad (20)$$

where $\mathbf{D}$ is a $p_{\mathcal{S}} \times p_{\mathcal{S}}$ diagonal matrix with entries $\hat{d}_j$ for $j \in \mathcal{S}$. Thus this algorithm is related to the scaled gradient algorithm of [11], but with a very different scaling matrix, and with groups of pixels rather than all pixels.

C. Special Cases

In the special case where the subset $\mathcal{S}$ contains only one pixel ($\mathcal{S} = \{j\}$), the above algorithm is equivalent to coordinate ascent [14, 12, 5], i.e., it turns out that $\phi_j(\theta_j; \theta^n) = \Phi(\theta_1^n, \ldots, \theta_{j-1}^n, \theta_j, \theta_{j+1}^n, \ldots, \theta_p^n)$. At the other extreme, when $\mathcal{S} = \{1, \ldots, p\}$, the above algorithm is similar to the scaled gradient algorithm [11], with a different diagonal scaling (and one that leads to faster convergence). However, the algorithms that are between those two extreme choices of $\mathcal{S}$ are the most useful, as discussed next.

D. Grouped Ascent

Optimization algorithms seem to involve the following tradeoff. The more parameters one updates simultaneously, the smaller the step sizes must be to ensure monotonicity, since the parameters are coupled. Therefore updating the parameters in small groups typically yields the fastest convergence rates, with coordinate ascent (one parameter at a time) being the extreme case. On the other hand, as mentioned above there are often "economies of scale" that can be used when updating several parameters simultaneously. So the actual computation per iteration is often reduced by updating larger groups. Thus for fast convergence but moderate computation, we would thus like to update the parameters using a few large groups, but chosen such that the parameters within each group are relatively uncoupled.

We have investigated the following grouped ascent method. We divide the image into blocks of size $m \times m$, for small m, and then update only 1 pixel out of each $m \times m$ block on a given subiteration. The number of groups is thus m^2, with p/m^2 pixels per group. Thus the required number of exponentiations is then only $m^2 N$, which is considerably smaller than the number of nonzero a_{ij} for small m. Note that $m = 1$ is the scaled gradient algorithm, and $m = \sqrt{p}$ is the coordinate ascent algorithm. As one increases m, the pixels within each group become more separated and therefore less coupled, which increases the convergence rate, but the computation also increases. Thus there is a basic tradeoff that can be adapted to the characteristics of the particular architecture.

IV. Results

To examine the convergence rates, we performed simulations using the thorax phantom shown in Fig. 1. The object

[7] One subiteration is adequate when ψ is quadratic, for example.

Fig. 1. Top: thorax phantom attenuation map (θ_{true}). Middle: FBP reconstruction. Bottom: penalized-likelihood reconstruction $\hat{\theta}$.

is 128×64 4.5mm pixels, and the system had 192 radial bins and 256 angular samples over $180°$, with 6mm wide strip integrals on 3mm spacing [5]. We used $\delta = 0.004\text{cm}^{-1}$ in (6), and generated noisy data with 3M counts. (Most of these counts correspond to rays that do not intersect the object.) Fig. 1 also shows the FBP reconstructed attenuation map and the penalized-likelihood reconstructed image using 10 iterations of the grouped-ascent algorithm. The statistical method appears to produce better image quality. (See [20] for quantitative comparisons.)

Fig. 2 shows that with $m = 3$ (9 groups), the proposed grouped-ascent algorithm increased the penalized log-likelihood almost as fast as the coordinate ascent algorithm per iteration. More important is the actual CPU time, which is shown in Fig. 3 (for DEC 3000/800). Because of the fewer exponentiations, the grouped-ascent algorithms use far less CPU time per iteration than the coordinate ascent algorithm, so with $m = 3$ the penalized log-likelihood flattens out in about 25 CPU seconds, whereas coordinate ascent takes over 100 CPU seconds. Furthermore, the grouped ascent algorithm is parallelizable, so with appropriate hardware could be significantly accelerated. Note that "$1 \star 1$ grouped ascent" is similar to the scaled gradient algorithm of [11].

Fig. 4 shows similar results for truncated transmission scans, such as would be obtained in fan-beam SPECT.

V. Discussion

Based on the results in this paper and recent work [11,5], we consider the transmission EM algorithm [13,9] to be ob-
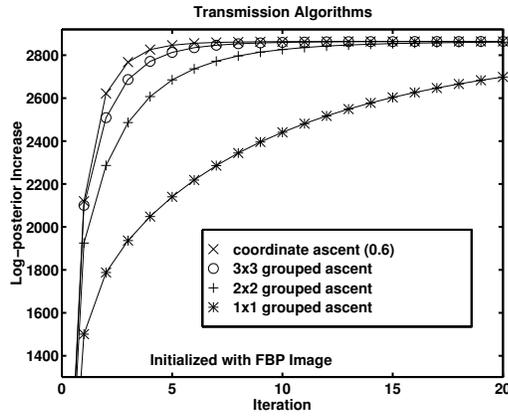
Fig. 2. Objective function increase $\Phi(\theta^n) - \Phi(\theta^0)$ versus iteration n.
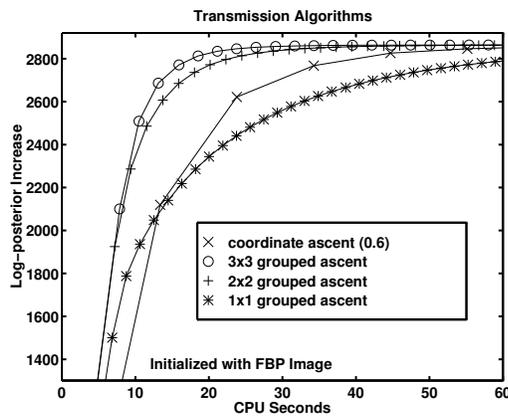
Fig. 3. Objective function increase $\Phi(\theta^n) - \Phi(\theta^0)$ versus CPU time.
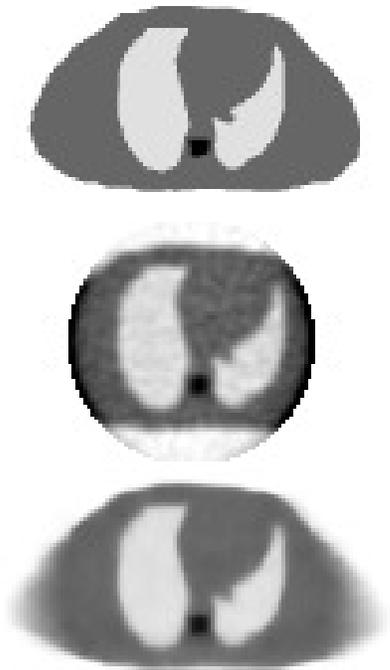
Fig. 4. As in Fig. 1, but for truncated transmission data.

solete. For penalized likelihood transmission image reconstruction, our proposed grouped ascent algorithm has fast convergence, reduced exponentiations per iteration, easily accommodates nonnegativity, and is flexibly parallelizable.

After this abstract was submitted, we became aware of a very similar algorithm developed independently by Sauer et al. [21].

REFERENCES

[1] EP Ficaro, JA Fessler, WL Rogers, and M Schwaiger. Comparison of Americium-241 and Technicium-99m as transmission sources for attenuation correction of Thallium-201 SPECT imaging of the heart. J. Nuc. Med., 35(4):652–663, Apr. 1994.

[2] K Wienhard, L Eriksson, S Grootoonk, M Casey, U Pietrzyk, and WD Heiss. Performance evaluation of a new generation positron scanner ECAT EXACT. J. Comp. Assisted Tomo., 16(5):804–813, Sept. 1992.

[3] SR Cherry, M Dahlbom, and EJ Hoffman. High sensitivity, total body PET scanning using 3D data acquisition and reconstruction. IEEE Tr. Nuc. Sci., 39(4):1088–1092, Aug. 1992.

[4] SR Meikle, M Dahlbom, and SR Cherry. Attenuation correction using count-limited transmission data in positron emission tomography. J. Nuc. Med., 34(1):143–150, Jan. 1993.

[5] JA Fessler. Hybrid Poisson/polynomial objective functions for tomographic image reconstruction from transmission scans. IEEE Tr. Im. Proc., 4(10):1439–50, Oct. 1995.

[6] 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):106–114, June 1987.

[7] K Lange. Convergence of EM image reconstruction algorithms with Gibbs smoothing. IEEE Tr. Med. Im., 9(4):439–446, Dec. 1990. Corrections, June 1991.

[8] E Mumcuoglu, R Leahy, and S Cherry. A statistical approach to transmission image reconstruction from ring source calibration measurements in PET. In Conf. Rec. of the IEEE Nuc. Sci. Symp. Med. Im. Conf., vol. 2, pp. 910–912, 1992.

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

[10] EU Mumcuoglu, R Leahy, SR Cherry, and Z Zhou. Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images. IEEE Tr. Med. Im., 13(3):687–701, Dec. 1994.

[11] K Lange and JA Fessler. Globally convergent algorithms for maximum a posteriori transmission tomography. IEEE Tr. Im. Proc., 4(10):1430–8, Oct. 1995.

[12] C Bouman and K Sauer. A unified approach to statistical tomography using coordinate descent optimization. IEEE Tr. Im. Proc., 1993. To appear.

[13] K Lange and R Carson. EM reconstruction algorithms for emission and transmission tomography. J. Comp. Assisted Tomo., 8(2):306–316, Apr. 1984.

[14] C Bouman and K Sauer. Fast numerical methods for emission and transmission tomographic reconstruction. In Proc. 27th Conf. Info. Sci. Sys., Johns Hopkins, pp. 611–616, 1993.

[15] JA Fessler and AO Hero. Space-alternating generalized expectation-maximization algorithm. IEEE Tr. Sig. Proc., 42(10):2664–2677, Oct. 1994.

[16] JA Fessler and AO Hero. Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms. IEEE Tr. Im. Proc., 4(10):1417–29, Oct. 1995.

[17] AR De Pierro. On the relation between the ISRA and the EM algorithm for positron emission tomography. IEEE Tr. Med. Im., 12(2):328–333, June 1993.

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

[19] K Sauer and C Bouman. A local update strategy for iterative reconstruction from projections. IEEE Tr. Sig. Proc., 41(2):534–548, Feb. 1993.

[20] JA Fessler. Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography. IEEE Tr. Im. Proc., 5(3), Mar. 1996. To appear.

[21] KD Sauer, S Borman, and CA Bouman. Parallel computation of sequential pixel updates in statistical tomographic reconstruction. In Proc. IEEE Intl. Conf. on Image Processing, 1995.