Globally Convergent Ordered Subsets Algorithms: Application to Tomography

Sangtae Ahn and Jeffrey A. Fessler

Abstract— **We present new algorithms for penalized-likelihood image reconstruction: modified BSREM (block sequential regularized expectation maximization) and relaxed OS-SPS (ordered subsets separable paraboloidal surrogates). Both of them are globally convergent to the unique solution, easily incorporate convex penalty functions, and are parallelizable—updating all voxels (or pixels) simultaneously. They belong to a class of relaxed ordered subsets algorithms. We modify the scaling function of the existing BSREM (De Pierro and Yamagishi, 01) so that we can prove global convergence without previously imposed assumptions. We also introduce a diminishing relaxation parameter into the existing OS-SPS (Erdoğan and Fessler, 99) to achieve global convergence. We also modify the penalized-likelihood function to enable the algorithms to cover a zero-background-event case. Simulation results show that the algorithms are both globally convergent and fast.**

I. Introduction

STATISTICAL image reconstruction methods have shown improved image quality over conventional filtered backprojection (FBP) methods (*e.g.*, [1]). Statistical methods can use accurate physical models and take the stochastic nature of noise into account; in addition, they can easily enforce object constraints like nonnegativity. For ML estimation, the expectation maximization (EM) algorithm [2] was introduced into emission and transmission tomography by Shepp and Vardi [3], and Lange and Carson [4]. The EM algorithm, despite its nice properties such as monotonicity and positivity, suffers slow convergence; there has been considerable efforts to develop accelerated algorithms.

Ordered subsets (OS) algorithms, also known as block-iterative or incremental gradient methods, have shown significantly accelerated "convergence" compared to EM (*e.g.*, [5]). The ordered subsets idea is to use only one subset of the measurement data in place of the total data for each update. Hudson and Larkin [5] presented the OS-EM algorithm and showed its order-of-magnitude acceleration over EM. However, OS-EM generally oscillates rather than converges to a ML solution, which is the typical behavior of OS algorithms without relaxation. We would like an algorithm to be not only fast but also *globally convergent*. An algorithm is said to be *globally convergent* if for any starting point the algorithm is guaranteed to generate a sequence of points converging to a solution [6, p.182]. When implemented appropriately, a diminishing relaxation parameter (or stepsize) can help OS algorithms converge by suppressing the limit cycle (*e.g.*, [7]).

Due to the ill-posedness of the image reconstruction problem ML estimation yields unsatisfactory noisy images. Regularization or penalization, equivalently a Bayesian approach, can treat the problem. For regularized OS algorithms, De Pierro and Yamagishi [8] have recently presented the block sequential regularized EM (BSREM) algorithm, an extension of the row-action maximum likelihood algorithm (RAMLA) [7]. They have provided a convergence proof for BSREM with a few "a posteriori" assumptions: i) the convergence of the objective sequence and ii) the positivity and boundedness of each iterate. We relax such assumptions by making some modifications to BSREM: i) a modified scaling function and ii) a modified log-likelihood if needed (zero-background-event case). Neither of these modifications changes the value of the final solution. They just ensure that the iterates converge to that solution.

The paraboloidal surrogates (PS) method [9] [10] is another attractive family of reconstruction algorithms. While the EM algorithm maximizes a surrogate function obtained by conditional expectation (E-step) for each update, the PS methods employ paraboloic (quadratic) surrogate functions. Separable surrogates are natural for OS algorithms; the use of separable paraboloidal surrogates and OS principles leads to the OS-SPS algorithm [11] (originally suggested as OSTR in the context of transmission tomography), which is fast but not globally convergent. We introduce relaxation into the algorithm to obtain relaxed OS-SPS, which we show to be globally convergent. We also modify the log-likelihood if needed (zero backgrounds). The relaxed OS-SPS can be seen as a diagonally scaled version of incremental subgradient methods [12] [13]. Of related work, Kudo, Nakazawa, and Saito [14] applied to emission tomography reconstruction their block-gradient optimization similar to the incremental subgradient methods; Bertsekas [15] proposed a hybrid algorithm from the least mean squares (OS) and the steepest descent methods (non-OS).

II. The Problem

A. Penalized-Likelihood Image Reconstruction

We focus on emission computed tomography: positron emission tomography (PET) or single photon emission computed tomography (SPECT). Assuming usual Poisson distributions, the measurement model for emission scans is as follows:

$$y_i \sim \text{Poisson}\left\{\sum_{j=1}^{p} a_{ij}\lambda_j + r_i\right\}, \;\; i = 1, 2, \ldots, N \qquad (1)$$

where $y_i \geq 0$ is the number of photons counted in the ith bin, $\lambda_j \geq 0$ is the activity at the jth voxel (or pixel), $\boldsymbol{A} = \{a_{ij}\}$ is the system matrix (incorporating scanning time, detector efficiencies, attenuation, scan geometry, etc.), and $r_i \geq 0$ is the mean number of background events such as scatters and random coincidences. The goal is to estimate the activity $\boldsymbol{\lambda} =$

This work was supported in part by the NSF under Grant BES-9982349 and by the NIH under Grant CA-60711.

The authors are with the EECS Department, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: sangtaea@eecs.umich.edu and fessler@eecs.umich.edu).

$[\lambda_1, \lambda_2, \ldots, \lambda_p]'$ based on the measurement y_i's with known $\boldsymbol{A}$ and r_i's where $'$ denotes the transpose.

The log-likelihood of the data y_i's is, ignoring constants independent of $\boldsymbol{\lambda}$, as follows:

$$L(\boldsymbol{\lambda}) = \sum_{i=1}^{N} h_i(l_i(\boldsymbol{\lambda})) \qquad (2)$$

where $h_i(l) = y_i \log l - l$ and $l_i(\boldsymbol{\lambda}) = \sum_{j=1}^{p} a_{ij}\lambda_j + r_i$. The penalized-likelihood (PL) estimation is to maximize the following penalized-likelihood objective function over $\boldsymbol{\lambda} \geq \mathbf{0}$:

$$\Phi(\boldsymbol{\lambda}) = L(\boldsymbol{\lambda}) - \beta R(\boldsymbol{\lambda}) \qquad (3)$$

where $\beta \geq 0$ is a regularization parameter which controls the level of smoothness in the reconstructed image, and R is, for example, the following type of roughness penalty function:

$$R(\boldsymbol{\lambda}) = \frac{1}{2}\sum_{j=1}^{p}\sum_{k\in\mathcal{N}_j} \omega_{jk}\psi(\lambda_j - \lambda_k) \qquad (4)$$

where $\mathcal{N}_j$ denotes the neighborhood of the jth voxel (or pixel), ω_{jk} is a weighting factor and ψ is a potential function. We assume that ψ is convex, continuously differentiable and bounded below; so is R. In addition, we assume $\min_{\boldsymbol{\lambda}} R(\boldsymbol{\lambda}) = 0$ without loss of generality. If $\beta = 0$, the problem becomes maximum likelihood (ML) estimation.

We assume that the objective function Φ is strictly concave, so there exists the unique PL solution $\hat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda}\geq\mathbf{0}} \Phi(\boldsymbol{\lambda})$. This can be ensured by choosing R appropriately. Under this assumption one can verify that there exists an upper bound $U \in (0, \infty)$ for $\hat{\boldsymbol{\lambda}}$, given any $\boldsymbol{\nu} \in \mathbb{R}^p$ such that $\Phi(\boldsymbol{\nu}) \in \mathbb{R}$, as follows [16]:

$$U \triangleq \max_{1\leq j\leq p}\left\{\frac{y_{i_j}}{a_{i_j j}}, \frac{\tilde{y}_{i_j}}{a_{i_j j}}\right\} + 1 \qquad (5)$$

where $i_j \triangleq \arg\max_{0\leq i\leq N} a_{ij}$, $\forall j$, and

$$\tilde{y}_i \triangleq \exp\left(h_i(y_i+1) + \sum_{k\neq i} h_k(y_k) - \Phi(\boldsymbol{\nu})\right), \ \forall i.$$

As a result, we have

$$\hat{\boldsymbol{\lambda}} \in B \triangleq \{\boldsymbol{\lambda} \in \mathbb{R}^p : 0 \leq \lambda_j \leq U, \ \forall j\}.$$

Next, we modify the penalized-likelihood for a zero-background-event case ($r_i = 0$) to make the objective function well-behaved over $\boldsymbol{\lambda} \in B$ (especially on boundaries). Although non-zero background events are realistic, we would like our algorithms to also be applied to the $r_i = 0$ case which has been used more frequently despite its oversimplicity. Decide $I = \{1 \leq i \leq N : r_i = 0\}$. Given any $\boldsymbol{\nu} \in B$ such that $\Phi(\boldsymbol{\nu}) \in \mathbb{R}$, consider the following modified objective function:

$$\tilde{\Phi}(\boldsymbol{\lambda}) \triangleq \sum_{\substack{i=1\\ i\in I}}^{N} \tilde{h}_i(l_i(\boldsymbol{\lambda})) + \sum_{\substack{i=1\\ i\notin I}}^{N} h_i(l_i(\boldsymbol{\lambda})) - \beta R(\boldsymbol{\lambda}) \qquad (6)$$

where

$$\tilde{h}_i(l) \triangleq \begin{cases} \frac{\ddot{h}_i(\epsilon)}{2}(l-\epsilon)^2 + \dot{h}_i(\epsilon)(l-\epsilon) + h_i(\epsilon) & \text{for } l < \epsilon \\ h_i(l) \quad \text{for } l \geq \epsilon \end{cases}$$

with

$$\epsilon \triangleq \frac{1}{2}\min_{\substack{i\in I\\ y_i\neq 0}}\left\{\exp\left(\frac{\Phi(\boldsymbol{\nu}) - \sum_{k\neq i} h_k(y_k)}{y_i}\right), y_i, \sum_{j=1}^{p} a_{ij}\nu_j\right\}.$$

Then we can show [16] that

$$\hat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda}\in B} \Phi(\boldsymbol{\lambda}) = \arg\max_{\boldsymbol{\lambda}\in B} \tilde{\Phi}(\boldsymbol{\lambda}),$$

meaning that this modified objective function has the same maximizer as the original. The modified objective function $\tilde{\Phi}$ has the nice properties that $\nabla\tilde{\Phi}$ is Lipschitz continuous and bounded over B contrast to Φ (when $r_i = 0$), while preserving the maximizer $\hat{\boldsymbol{\lambda}}$. We will henceforth take $\tilde{\Phi}$ as our objective function but keep the notation Φ for simplicity; likewise, h_i will denote $\tilde{h}_i$ when $r_i = 0$.

B. *Ordered Subsets (OS) Algorithms*

Most iterative algorithms for finding $\hat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda}\geq\mathbf{0}} \Phi(\boldsymbol{\lambda})$ use the gradient $\nabla\Phi(\boldsymbol{\lambda})$ of the objective function which involves a sum over sinogram indices, *i.e.*, backprojection. Among them, many "parallelizable" algorithms—able to update all the voxels (or pixels) simultaneously—can be written in the following form:

$$\lambda_j^{n+1} = \lambda_j^n + \alpha_n d_j(\boldsymbol{\lambda}^n)\nabla_j\Phi(\boldsymbol{\lambda}^n), \ \ j = 1, 2, \ldots, p \qquad (7)$$

where $\alpha_n > 0$ is a stepsize (or relaxation parameter), $d_j(\boldsymbol{\lambda})$'s are nonnegative scaling functions and $\nabla_j\Phi(\boldsymbol{\lambda})$ is, ignoring here the regularization term for simplicity, as follows:

$$\nabla_j\Phi(\boldsymbol{\lambda}) = \sum_{i=1}^{N} a_{ij}\dot{h}_i(l_i(\boldsymbol{\lambda})). \qquad (8)$$

Ordered subsets (OS) algorithms are obtained by replacing the sum $\sum_{i=1}^{N}$ in (8) with a sum $\sum_{i\in S_m}$ over a subset S_m of $\{1, 2, \ldots, N\}$. Let $\{S_m\}_{m=1}^{M}$ be a partition of $\{1, 2, \ldots, N\}$, and let

$$f_m(\boldsymbol{\lambda}) \triangleq \sum_{i\in S_m} h_i(l_i(\boldsymbol{\lambda})) \qquad (9)$$

be a sub-objective function, $\forall m$, resulting in $\Phi = \sum_{m=1}^{M} f_m$, where the regularization term is included in one or more of the f_m's. Suppose the following "subset balance"-like conditions hold:

$$\nabla f_1(\boldsymbol{\lambda}) \cong \nabla f_2(\boldsymbol{\lambda}) \cong \cdots \cong \nabla f_M(\boldsymbol{\lambda}) \qquad (10)$$

or, equivalently,

$$\nabla\Phi(\boldsymbol{\lambda}) \cong M\nabla f_m(\boldsymbol{\lambda}), \ \ \forall m. \qquad (11)$$

Now an ordered subsets version of (7) can be obtained, by substituting $M\nabla_j f_m(\boldsymbol{\lambda})$ for $\nabla_j\Phi(\boldsymbol{\lambda})$, as follows:

$$\lambda_j^{n,0} = \lambda_j^n$$

$$\lambda_j^{n,m} = \lambda_j^{n,m-1} + \alpha_n d_j(\boldsymbol{\lambda}^{n,m-1}) \nabla_j f_m(\boldsymbol{\lambda}^{n,m-1}), \forall m$$
$$\lambda_j^{n+1} = \lambda_j^{n,M} \tag{12}$$

where the factor M is absorbed into α_n. We refer to each update as the mth subiteration of the nth iteration. In the tomography context, the partition $\{S_m\}_{m=1}^M$ are naturally chosen so that projections within one subset correspond to projections with downsampled projection angles. It is desirable to order the subsets in a way that projections corresponding to one subset are as perpendicular as possible to previously used angles at each subiteration [17].

The OS algorithms have been successful in speeding up convergence. However, they generally get into a limit cycle when using a constant stepsize $\alpha_n = \alpha$ and do not converge to the solution $\hat{\boldsymbol{\lambda}}$. We may need to use a diminishing stepsize such that $\alpha_n \to 0$ to suppress the limit cycle. Even if an algorithm with relaxation converges to some $\boldsymbol{\lambda}^*$, we still must ensure that the limit is the solution $\hat{\boldsymbol{\lambda}}$.

III. Globally Convergent Ordered Subsets Algorithms

We present relaxed OS algorithms which are globally convergent: modified BSREM and relaxed OS-SPS. The goal is to maximize the penalized-likelihood $\Phi(\boldsymbol{\lambda})$ over $\boldsymbol{\lambda} \geq \mathbf{0}$ (equivalently, $\boldsymbol{\lambda} \in B$). Given a partition $\{S_m\}_{m=1}^M$ of $\{1, 2, \ldots, N\}$, we take the following sub-objective functions:

$$f_m(\boldsymbol{\lambda}) = \sum_{i \in S_m} h_i(l_i(\boldsymbol{\lambda})) - \beta_m R(\boldsymbol{\lambda}), \ \forall m \tag{13}$$

where $\beta_m (\geq 0)$'s satisfy $\sum_{m=1}^M \beta_m = \beta$. (Note $\Phi = \sum_{m=1}^M f_m$.)

A. Modified BSREM

De Pierro and Yamagishi [8] have recently presented the BSREM algorithm and proved its global convergence on the following assumptions: i) the sequence $\{\boldsymbol{\lambda}^n\}$ generated by the algorithm is positive and bounded and ii) the objective sequence $\{\Phi(\boldsymbol{\lambda}^n)\}$ converges. These conditions are not automatically ensured by the form of the original BSREM. We eliminate those assumptions by modifying the algorithm.

The basic idea of the modification is to ensure all iterates lying in the interior of the constraint set B by suitably choosing a scaling function $d_j(\boldsymbol{\lambda})$ and a relaxation parameter $\{\alpha_n\}$ so that implementation and convergence analysis become convenient. For EM-like algorithms, we observe that the form of $d_j(\boldsymbol{\lambda}) =$ (some term) $\times \lambda_j$ helps each iterate to keep positivity, *i.e.*, to avoid crossing the lower boundary $\lambda_j = 0$. We can also do a similar thing for the upper boundary $\lambda_j = U$. Consider the following modified BSREM algorithm:

$$\boldsymbol{\lambda}^{n,0} = \boldsymbol{\lambda}^n$$
$$\boldsymbol{\lambda}^{n,m} = \boldsymbol{\lambda}^{n,m-1} + \alpha_n \boldsymbol{D}(\boldsymbol{\lambda}^{n,m-1}) \nabla f_m(\boldsymbol{\lambda}^{n,m-1}), \ \forall m$$
$$\boldsymbol{\lambda}^{n+1} = \boldsymbol{\lambda}^{n,M} \tag{14}$$

where $\boldsymbol{D}(\boldsymbol{\lambda}) = \mathrm{diag}\{d_1(\boldsymbol{\lambda}), \ldots, d_p(\boldsymbol{\lambda})\}$ with

$$d_j(\boldsymbol{\lambda}) = \begin{cases} \lambda_j & \text{for } 0 \leq \lambda_j < \frac{U}{2} \\ U - \lambda_j & \text{for } \frac{U}{2} \leq \lambda_j \leq U \end{cases}, \ \forall j. \tag{15}$$

Then it can be shown [16] that each $\boldsymbol{\lambda}^{n,m}$ generated by (14) belongs to the interior of B when i) $\boldsymbol{\lambda}^0 \in \mathrm{Int}\, B$ and ii) $0 < \alpha_n < \alpha_0, \ \forall n$, where

$$\alpha_0 = \min_{m,\, j} \left\{ \frac{U/2}{\sum_{i \in S_m} y_i + \gamma}, \ \frac{1}{\sum_{i \in S_m} a_{ij} + \gamma} \right\} \tag{16}$$

with $\gamma = \max_{j,\, m,\, \boldsymbol{\lambda} \in B} \{\beta_m \nabla_j R(\boldsymbol{\lambda}), \ \beta_m \lambda_j \nabla_j R(\boldsymbol{\lambda}), \ 0\}$. Therefore, each iterate satisfies positivity and boundedness. The following Theorem [16] guarantees the global convergence of the modified BSREM:

Theorem 1: Suppose that $0 < \alpha_n < \alpha_0$, $\sum_{n=1}^\infty \alpha_n = \infty$ and $\sum_{n=1}^\infty \alpha_n^2 < \infty$. Then the sequence $\{\boldsymbol{\lambda}^n\}$ generated by (14) with an initial point $\boldsymbol{\lambda}^0 \in \mathrm{Int}\, B$ converges to $\hat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda} \in B} \Phi(\boldsymbol{\lambda})$.

For any strictly concave function Φ and sub-objective functions f_m's such that $\Phi = \sum_{m=1}^M f_m$, if $\boldsymbol{D}(\boldsymbol{\lambda}) \nabla f_m(\boldsymbol{\lambda})$ and $\nabla \Phi(\boldsymbol{\lambda})$ is Lipschitz continuous, then we can obtain similar results.

B. Relaxed OS-SPS

We consider another family of OS algorithms with a constant matrix $\boldsymbol{D}$ as follows:

$$\boldsymbol{\lambda}^{n,0} = \boldsymbol{\lambda}^n$$
$$\boldsymbol{\lambda}^{n,m} = \mathcal{P}_B\left(\boldsymbol{\lambda}^{n,m-1} + \alpha_n \boldsymbol{D} \nabla f_m(\boldsymbol{\lambda}^{n,m-1})\right), \ \forall m$$
$$\boldsymbol{\lambda}^{n+1} = \boldsymbol{\lambda}^{n,M} \tag{17}$$

where $\boldsymbol{D} = \mathrm{diag}\{d_1, \ldots, d_p\}$ with $d_j > 0, \ \forall j$, and $\mathcal{P}_B(\boldsymbol{\lambda})$ is the projection of $\boldsymbol{\lambda} \in \mathbb{R}^p$ onto B. If $\boldsymbol{D} = \boldsymbol{I}$, the algorithm (17) becomes one of incremental subgradient methods [13]; so it can be viewed as a diagonally scaled incremental subgradient method.

The special case of (17) is OS-SPS [11] (if $\alpha_n = 1$ and d_j is the inverse of a corresponding "precomputed curvature"). For example, for OS-SPS with a quadratic penalty of $\psi(t) = t^2/2$, the diagonal elements are as follows:

$$d_j = M \left(-\sum_{i=1}^N a_{ij} \ddot{h}_i(y_i) \sum_{k=1}^p a_{ik} + 2\beta \sum_{k \in \mathcal{N}_j} \omega_{jk} \right)^{-1}, \ \forall j.$$

If we allow a diminishing stepsize, we obtain a relaxed OS-SPS. The relaxed OS-SPS, contrast to ordinary OS-SPS, is globally convergent by the following Theorem [16].

Theorem 2: Suppose that $\alpha_n > 0, \ \forall n$, and $\sum_{n=1}^\infty \alpha_n = \infty$. Then the sequence $\{\boldsymbol{\lambda}^n\}$ generated by (17) converges to $\hat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda} \in B} \Phi(\boldsymbol{\lambda})$.

For any strictly concave function Φ and sub-objective functions f_m's such that $\Phi = \sum_{m=1}^M f_m$, if $f_m(\boldsymbol{\lambda})$ is concave and $\nabla f_m(\boldsymbol{\lambda})$'s are bounded, then we can obtain similar results.

IV. Results

We performed PL image reconstruction for two-dimensional PET simulation using the Shepp-Logan digital phantom to test the performance of the modified BSREM and the relaxed OS-SPS. We chose 16 subsets by angular downsampling in sinogram. For comparison, we also performed SPS (1 subset with "optimum curvature"), ordinary OS-SPS (16 subsets with "precomputed curvature") and De Pierro's modified EM (DPEM) [18]. The sinogram had 128 radial bins and 160 angles uniformly sampled over 180 degrees. The total photon counts amounted to 5×10^6 and, r_i corresponded to a uniform field of 10% of random coincidences. The reconstructed images were 128 by 128 pixels and attenuation was not considered. We used a first-order quadratic penalty function $\psi(t) = t^2/2$ with a regularization parameter $\beta = 8$. The FBP reconstruction was used as a starting image. As a relaxation parameter, we chose $\alpha_n = 11/(10+n)$ for relaxed OS-SPS, and $\alpha_n = 4 \times 10^{-4}/n$ for modified BSREM through a few trials.

The upper bound U for $\hat{\boldsymbol{\lambda}}$ computed by (5) in this example was too large for computers capability ($\sim \exp 10^5$); so U is virtually infinity. This implies that in practice $d_j(\boldsymbol{\lambda})$ of (15) becomes λ_j as the original BSREM, and $\mathcal{P}_B(\cdot)$ of (17) becomes just $[\cdot]_+$ where $[\boldsymbol{\lambda}]_{+,j} = \max\{0, \lambda_j\}$. A real "practical" problem lies in determining the upper bound α_0 for a relaxation parameter to guarantee that the modified (or original) BSREM is positive. Equation (16) gives an extremely small relaxation parameter in this example (since the gradient of the penalty function can be very large on the constraint set B); the convergence rate will become very slow with the small relaxation. Thresholding (like $\mathcal{T}\{\lambda_j\} = \max\{\lambda_j, \delta\}$ for some small $\delta > 0$) may ensure positivity without the concern for relaxation parameters. But the algorithm including thresholding is already a "new" algorithm and we will have to analyze the global convergence for this "new" algorithm. Even though the "new" algorithm seems to be also globally convergent, it is not easy to prove it. In this simulation for modified BSREM, we chose a relaxation parameter reasonably small to keep iterates positivity and reasonably large for the algorithm to be fast, through a few trials. In contrast, relaxed OS-SPS does not suffer from this difficulty in choosing α_0, so it may be more convenient in practice.

Fig. 1 shows objective increases in the penalized log-likelihood $\Phi(\boldsymbol{\lambda}^n) - \Phi(\boldsymbol{\lambda}^0)$. As can be seen in the figure, OS algorithms (OS-SPS, modified BSREM and relaxed OS-SPS) are faster than non-OS algorithms (SPS and DPEM). Fig. 2 shows the same figure zoomed in for the OS algorithms. We can observe that modified BSREM and relaxed OS-SPS keep increasing—actually, converge to the optimal point— whereas OS-SPS stagnated at a suboptimal point. Although the relaxed OS-SPS seems to be superior to the modified BSREM in the figure, we do not jump to such a conclusion from this preliminary data since the performances depend on relaxation parameters. Fig. 3 displays the reconstructed images by 20 iterations of relaxed OS-SPS and modified BSREM (also the digital phantom and FBP reconstruction).

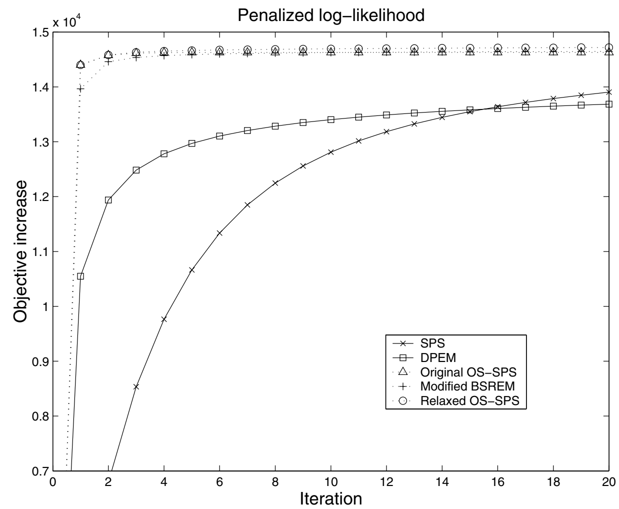

Fig. 1. Comparison of objective increase rates of non-OS algorithms—SPS (with "optimum curvature") and De Pierro's modified EM (DPEM)—and OS algorithms—original OS-SPS, modified BSREM, and relaxed OS-SPS—with 16 subsets: the OS algorithms are shown to be faster than the non-OS algorithms.

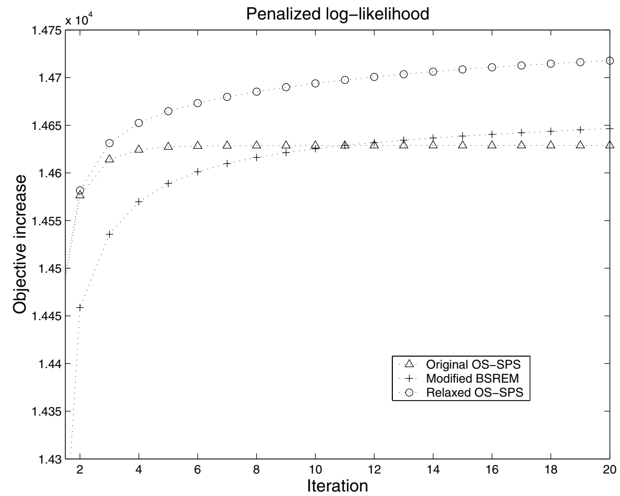

Fig. 2. Comparison of objective increase rates of original OS-SPS, modified BSREM, and relaxed OS-SPS (same as Fig. 1 except zoomed in): modified BSREM and relaxed OS-SPS are globally convergent whereas original OS-SPS gets stuck at a suboptimal point.

V. Discussion

We presented relaxed OS algorithms (modified BSREM and relaxed OS-SPS) which are globally convergent. Simulation results showed that both algorithms can be faster than non-OS and unrelaxed OS algorithms, as well as globally convergent. However, for modified BSREM, as mentioned in Section IV, it was hard to determine the relaxation parameter (without thresholding) for ensuring fast and global convergence. By contrary, it is relatively easy to determine the relaxation parameter for relaxed OS-SPS. Since the original OS-SPS has already been correctly scaled, it is effective to use a relaxation parameter starting with 1 and decreasing as $\sim O(\frac{1}{n})$. Nevertheless, there is still much room for improvement in choosing the relaxation parameters. Training of relaxation parameters for a specific type of tasks may be performed [7] [17]. But we believe that convergence rate analysis will give us more insights related to relaxation and to the choice of scaling functions. We will direct future research

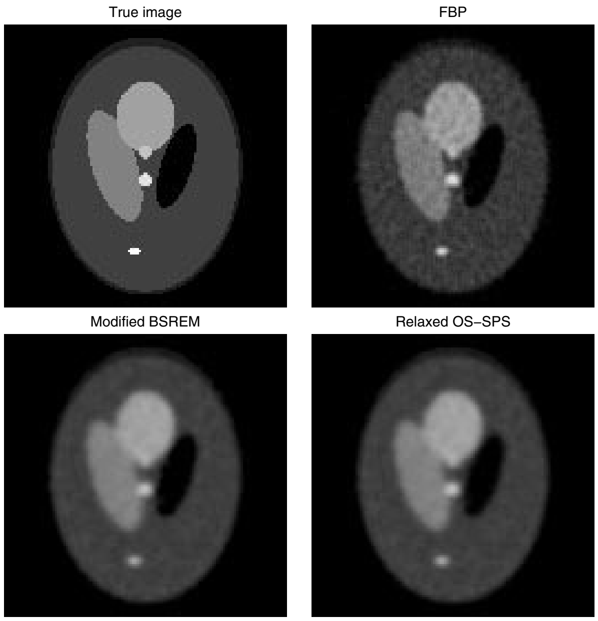

Fig. 3. Reconstructed images (16 subsets and 20 iterations): true image; Shepp-Logan phantom (top-left), starting image; FBP reconstruction (top-right), modified BSREM reconstruction (bottom-left), and relaxed OS-SPS reconstruction (bottom-right)

to convergence rate analysis.

An extension to transmission tomography is straightforward for zero background events. However, for a nonzero-background-event case, the penalized-likelihood becomes non-concave; it is an open problem to find a global maximizer. At least we hope to find a stationary point.

References

[1] L. A. Shepp, Y. Vardi, J. B. Ra, S. K. Hilal, and Z. H. Cho, "Maximum likelihood PET with real data," *IEEE Tr. Nuc. Sci.*, vol. 31, no. 2, pp. 910–3, Apr. 1984.

[2] 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*, vol. 39, no. 1, pp. 1–38, 1977.

[3] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Tr. Med. Im.*, vol. 1, no. 2, pp. 113–22, Oct. 1982.

[4] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comp. Assisted Tomo.*, vol. 8, no. 2, pp. 306–316, Apr. 1984.

[5] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Tr. Med. Im.*, vol. 13, no. 4, pp. 601–9, Dec. 1994.

[6] D. G. Luenberger, *Linear and nonlinear programming*, Addison-Wesley, Massachusetts, 2 edition, 1984.

[7] J. A. Browne and A. R. De Pierro, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography," *IEEE Tr. Med. Im.*, vol. 15, no. 5, pp. 687–99, Oct. 1996.

[8] A. R. De Pierro and M. E. B. Yamagishi, "Fast EM-like methods for maximum 'a posteriori' estimates in emission tomography," *IEEE Tr. Med. Im.*, vol. 20, no. 4, pp. 280–8, Apr. 2001.

[9] J. A. Fessler and H. Erdoğan, "A paraboloidal surrogates algorithm for convergent penalized-likelihood emission image reconstruction," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 1998, vol. 2, pp. 1132–5.

[10] H. Erdoğan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Tr. Med. Im.*, vol. 18, no. 9, pp. 801–14, Sept. 1999.

[11] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–51, Nov. 1999.

[12] A. Nedic and D. Bertsekas, "Convergence rate of incremental subgradient algorithms," in *Stochastic Optimization: Algorithms and Applications*, S. Uryasev and P. M. Pardalos, Eds., pp. 263–304. Kluwer, New York, 2000.

[13] A. Nedic and D. P. Bertsekas, "Incremental subgradient methods for non-differentiable optimization," *SIAM J. Optim.*, vol. 12, no. 1, pp. 109–38, 2001.

[14] H. Kudo, H. Nakazawa, and T. Saito, "Convergent block-iterative method for general convex cost functions," in *Proc. of the 1999 Intl. Mtg. on Fully 3D Im. Recon. in Rad. Nuc. Med.*, 1999, pp. 247–250.

[15] D. P. Bertsekas, "A new class of incremental gradient methods for least squares problems," *SIAM J. Optim.*, vol. 7, no. 4, pp. 913–26, Nov. 1997.

[16] S. Ahn and J. A. Fessler, "Globally convergent ordered subsets algorithms for emission tomography," in preparation.

[17] G. T. Herman and L. B. Meyer, "Algebraic reconstruction techniques can be made computationally efficient," *IEEE Tr. Med. Im.*, vol. 12, no. 3, pp. 600–609, Sept. 1993.

[18] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Tr. Med. Im.*, vol. 14, no. 1, pp. 132–137, Mar. 1995.