

# Incremental Optimization Transfer Algorithms: Application to Transmission Tomography

Sangtae Ahn, Jeffrey A. Fessler, Doron Blatt, and Alfred O. Hero

**Abstract**—No convergent ordered subsets (OS) type image reconstruction algorithms for transmission tomography have been proposed to date. In contrast, in emission tomography, there are two known families of convergent OS algorithms: methods that use relaxation parameters (Ahn and Fessler, 2003), and methods based on the incremental expectation maximization (EM) approach (Hsiao *et al.*, 2002). This paper generalizes the incremental EM approach by introducing a general framework that we call “incremental optimization transfer.” Like incremental EM methods, the proposed algorithms accelerate convergence speeds and ensure global convergence (to a stationary point) under mild regularity conditions without requiring inconvenient relaxation parameters. The general optimization transfer framework enables the use of a very broad family of non-EM surrogate functions. In particular, this paper provides the first convergent OS-type algorithm for transmission tomography. The general approach is applicable to both monoenergetic and polyenergetic transmission scans as well as to other image reconstruction problems. We propose a particular incremental optimization transfer method for (nonconcave) penalized-likelihood (PL) transmission image reconstruction by using separable paraboloidal surrogates (SPS). Results show that the new “transmission incremental optimization transfer (TRIOT)” algorithm is faster than nonincremental ordinary SPS and even OS-SPS yet is convergent.

## I. INTRODUCTION

Ordered subsets (OS) algorithms, also known as block iterative or incremental gradient methods, have been very popular in the medical imaging community for statistical image reconstruction due to their fast initial convergence. However, ordinary (unrelaxed) OS algorithms generally do not converge to an optimal solution but rather approach a suboptimal limit cycle.

There have been three known families of convergent incremental (or OS type) algorithms: methods that use relaxation parameters, methods based on the incremental EM approach, and incremental aggregated gradient (IAG) methods.

Relaxation parameters are used widely to render OS algorithms convergent (see [1] and the references therein). However, it is inconvenient to determine optimal relaxation parameters to achieve fast convergence rates and global convergence as well. In contrast, incremental EM algorithms do not require user-specified relaxation parameters [2]. They are convergent yet

This work was supported in part by the National Institutes of Health (NIH) under Grants CA-60711 and CA-87634, by the Department of Energy (DOE) under Grant DE-FG02-87ER60561, and by a Rackham Predoctoral Fellowship.

The authors are with the Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI 48109-2122 USA (e-mail: sangtaea@umich.edu, fessler@umich.edu, dblatt@eecs.umich.edu, hero@eecs.umich.edu).

faster than ordinary EM algorithms although slower initially than nonconvergent OS-EM type algorithms. Such incremental EM algorithms have been applied to emission tomography [3]–[6]. Recently, Blatt *et al.* proposed a convergent incremental gradient method, called incremental aggregated gradient (IAG), that does not require relaxation parameters [7].

This paper generalizes the incremental EM algorithms by introducing an approach called “incremental optimization transfer.” Incremental optimization transfer is a general framework in which one can develop many different algorithms by using a very broad family of application-dependent surrogate functions. In this paper, we focus on PL image reconstruction for transmission tomography, which is a challenging nonconcave maximization problem. We propose a particular incremental optimization transfer algorithm that uses separable paraboloidal surrogates (SPS) [8]. The proposed “transmission incremental optimization transfer (TRIOT)” algorithm is convergent yet converges faster than ordinary SPS [8]; it can be further accelerated by initializing through a few iterations of OS-SPS. It is parallelizable and accommodates most types of system model.

## II. INCREMENTAL OPTIMIZATION TRANSFER

Most objective functions of interest in image reconstruction can be expressed as a sum of several subobjective functions:

$$\Phi(\mathbf{x}) = \sum_{m=1}^M \Phi_m(\mathbf{x}), \quad (1)$$

where  $\Phi_m : \mathcal{X} \subset \mathbb{R}^p \rightarrow \mathbb{R}$  is a continuously differentiable function whose domain  $\mathcal{X}$  is a nonempty, convex and closed set. We consider the following optimization problem:

$$\text{maximize } \Phi(\mathbf{x}) \text{ subject to } \mathbf{x} \in \mathcal{X}. \quad (2)$$

Since usually there exists no closed-form solution to the above problem, one must apply iterative algorithms. Assume that for each subobjective function  $\Phi_m$ , we find a surrogate function,  $\phi_m : \mathcal{X}^2 \subset \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ , that is easier to maximize than  $\Phi_m$  and that satisfies the following usual minorization conditions [9], [10]:

$$\begin{aligned} \phi_m(\mathbf{x}; \mathbf{x}) &= \Phi_m(\mathbf{x}), & \forall \mathbf{x} \in \mathcal{X} \\ \phi_m(\mathbf{x}; \bar{\mathbf{x}}) &\leq \Phi_m(\mathbf{x}), & \forall \mathbf{x}, \bar{\mathbf{x}} \in \mathcal{X}, \end{aligned} \quad (3)$$

where  $\mathcal{X}^n$  denotes the  $n$ -ary Cartesian product over the set  $\mathcal{X}$ . It follows from the above conditions that

$$\Phi_m(\mathbf{x}) - \Phi_m(\bar{\mathbf{x}}) \geq \phi_m(\mathbf{x}; \bar{\mathbf{x}}) - \phi_m(\bar{\mathbf{x}}; \bar{\mathbf{x}}), \quad \forall \mathbf{x}, \bar{\mathbf{x}} \in \mathcal{X}.$$

In other words, choosing  $\mathbf{x}$  such that  $\phi_m(\mathbf{x}; \bar{\mathbf{x}}) \geq \phi_m(\bar{\mathbf{x}}; \bar{\mathbf{x}})$  ensures that  $\Phi_m(\mathbf{x}) \geq \Phi_m(\bar{\mathbf{x}})$ .

We define the following “augmented” objective function:

$$F(\mathbf{x}; \bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_M) = \sum_{m=1}^M \phi_m(\mathbf{x}; \bar{\mathbf{x}}_m). \quad (4)$$

Since

$$\max_{\bar{\mathbf{x}}_m \in \mathcal{X}} \phi_m(\mathbf{x}; \bar{\mathbf{x}}_m) = \phi_m(\mathbf{x}; \mathbf{x}) = \Phi_m(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{X}, \forall m$$

that is,

$$\max_{(\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_M) \in \mathcal{X}^M} F(\mathbf{x}; \bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_M) = \Phi(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{X},$$

one can rewrite the optimization problem (2) equivalently as follows:

$$\begin{aligned} & \text{maximize } F(\mathbf{x}; \bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_M) \\ & \text{subject to } (\mathbf{x}; \bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_M) \in \mathcal{X}^{M+1}, \end{aligned} \quad (5)$$

in a sense that  $\mathbf{x}^* \in \mathcal{X}$  is an optimal solution of (2) if and only if  $(\mathbf{x}^*; \bar{\mathbf{x}}_1^*, \dots, \bar{\mathbf{x}}_M^*) \in \mathcal{X}^{M+1}$  is an optimal solution of (5) for some  $(\bar{\mathbf{x}}_1^*, \dots, \bar{\mathbf{x}}_M^*) \in \mathcal{X}^M$ . Therefore we can find a solution to problem (2) by maximizing  $F$  with respect to  $(\mathbf{x}; \bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_M)$ . By alternating between updating  $\mathbf{x}$  and one of the  $\bar{\mathbf{x}}_m$ 's, we obtain an “incremental optimization transfer algorithm” outlined in Table I.

The incremental optimization transfer algorithm monotonically increases the augmented objective function  $F$ , but not necessarily the original objective function  $\Phi$  [11]. If one has only one subobjective function in (1), that is,  $M = 1$ , then the incremental optimization transfer algorithm reduces to an ordinary optimization transfer algorithm (e.g., EM and SPS) [8], [12], [13]. The incremental approach ( $M > 1$ ) usually leads to faster convergence rates than nonincremental methods ( $M = 1$ ). The incremental EM algorithms [2], [3] including COSEM [4], [5] are a special case where the surrogates  $\phi_m$  are constructed by EM principles. If one were to maximize just one of the  $\phi_m$ 's instead of the sum shown in (4), then one would have ordinary OS type algorithms (e.g., OS-EM and OS-SPS) [8], [14].

One can show [15] that every limit point of the sequence generated by an incremental optimization transfer algorithm is an element of the set of stationary points regardless of initial points when the following general sufficient conditions hold: (i) each  $\Phi_m$  and  $\phi_m(\cdot; \bar{\mathbf{x}})$  is continuously differentiable, (ii) the iterates are bounded (e.g.,  $\mathcal{X}$  is a bounded set), (iii) the surrogates  $\phi_m$  satisfy the minorization conditions in (3), (iv) the gradients of  $\Phi_m$  and  $\phi_m(\cdot; \bar{\mathbf{x}})$  match at  $\bar{\mathbf{x}}$ , and (v) the maximizer in (T-1) is defined uniquely (e.g.,  $\phi_m(\cdot; \bar{\mathbf{x}}_m)$  is strictly concave).

### III. APPLICATION TO TRANSMISSION TOMOGRAPHY

We develop a particular incremental optimization transfer algorithm for transmission tomographic reconstruction. We use quadratic surrogates [8], [10] rather than EM surrogates because the standard complete-data proposed in [16] for transmission tomography does not yield a closed-form M-step [17].

#### A. Problem

We assume the following Poisson statistical model for (monoenergetic) transmission measurements:

$$y_i \sim \text{Poisson}\{b_i e^{-[\mathbf{A}\mathbf{x}]_i} + r_i\}, \quad i = 1, \dots, N \quad (4)$$

where  $y_i$  denotes the transmission measurement of the  $i$ th detector,  $b_i$  denotes the blank scan counts of the  $i$ th detector,  $r_i$  denotes the mean number of background counts, and  $[\mathbf{A}\mathbf{x}]_i = \sum_{j=1}^p a_{ij} x_j$  represents the  $i$ th line integral of the attenuation map in which  $x_j$  is the unknown attenuation coefficient in the  $j$ th pixel,  $\mathbf{A} = \{a_{ij}\}$  is the system matrix, and  $N$  and  $p$  are the number of detectors and pixels, respectively. We assume that  $\{b_i\}$ ,  $\{a_{ij}\}$ , and  $\{r_i\}$  are known nonnegative constants. We focus on penalized-likelihood (PL), also known as maximum *a posteriori* (MAP), estimation for the attenuation map reconstruction. Our goal is to compute a PL estimate  $\hat{\mathbf{x}}^{\text{PL}}$  which is defined by

$$\hat{\mathbf{x}}^{\text{PL}} = \arg \max_{\mathbf{x} \in \mathcal{X}} \Phi(\mathbf{x}), \quad \Phi(\mathbf{x}) = L(\mathbf{x}) - \beta R(\mathbf{x}) \quad (5)$$

where the objective function  $\Phi$ , which can be nonconcave when  $r_i \neq 0$  [10], includes the log-likelihood

$$\begin{aligned} L(\mathbf{x}) &= \sum_{i=1}^N h_i([\mathbf{A}\mathbf{x}]_i) \\ h_i(l) &= y_i \log(b_i e^{-l} + r_i) - (b_i e^{-l} + r_i) \end{aligned}$$

and a roughness penalty

$$R(\mathbf{x}) = \frac{1}{2} \sum_{j=1}^p \sum_{k \in \mathcal{N}_j} w_{jk} \psi(x_j - x_k). \quad (6)$$

The feasible set is given by

$$\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^p : 0 \leq x_j \leq U, \quad \forall j\}. \quad (7)$$

In the box constraint set in (7), the upper bound  $U > 0$  is set by the user to be a value that is larger than the maximum attenuation coefficient conceivable for the object being scanned.

In the penalty function (6), the function  $\psi$  is a symmetric and convex potential function,  $\mathcal{N}_j$  represents a neighborhood of the  $j$ th pixel,  $\beta$  is a regularization parameter that controls the smoothness in reconstructed images, and  $w_{jk}$  are weights. We assume the potential function  $\psi$  satisfies some conditions given in [10], [18, p. 184]. We used the following edge-preserving nonquadratic potential function in our PL reconstruction results [19]:

$$\psi(t) = \delta^2 [ |t/\delta| - \log(1 + |t/\delta|) ] \quad (8)$$

for some  $\delta > 0$ .

#### B. Transmission Incremental Optimization Transfer (TRIOT)

We decompose the objective function  $\Phi$  into the following subobjective functions:

$$\Phi_m(\mathbf{x}) = \sum_{i \in S_m} h_i([\mathbf{A}\mathbf{x}]_i) - \frac{\beta}{M} R(\mathbf{x}), \quad m = 1, \dots, M,$$

TABLE I  
OUTLINE FOR INCREMENTAL OPTIMIZATION TRANSFER ALGORITHMS.

---

Initialize $\mathbf{x}^0, \bar{\mathbf{x}}_1^0, \dots, \bar{\mathbf{x}}_M^0 \in \mathcal{X}$	
<b>for</b> $n = 0, \dots, n_{\text{iter}} - 1$	
<b>for</b> $m = 1, \dots, M$	
$\mathbf{x}^{\text{new}} = \arg \max_{\mathbf{x} \in \mathcal{X}} F(\mathbf{x}; \bar{\mathbf{x}}_1^{n+1}, \dots, \bar{\mathbf{x}}_{m-1}^{n+1}, \bar{\mathbf{x}}_m^n, \bar{\mathbf{x}}_{m+1}^n, \dots, \bar{\mathbf{x}}_M^n)$	(T-1)
$\bar{\mathbf{x}}_m^{n+1} = \mathbf{x}^{\text{new}} = \arg \max_{\bar{\mathbf{x}}_m \in \mathcal{X}} F(\mathbf{x}^{\text{new}}; \bar{\mathbf{x}}_1^{n+1}, \dots, \bar{\mathbf{x}}_{m-1}^{n+1}, \bar{\mathbf{x}}_m, \bar{\mathbf{x}}_{m+1}^n, \dots, \bar{\mathbf{x}}_M^n)$	(T-2)
<b>end</b>	
$\mathbf{x}^{n+1} = \bar{\mathbf{x}}_M^{n+1}$	(T-3)
<b>end</b>	

---

where  $\{S_m\}_{m=1}^M$  is a partition of  $\{1, \dots, N\}$ . We use the usual subsets corresponding to downsampled projection angles [14]. Consider the following *separable* quadratic surrogate  $\phi_m$  for the subobjective function  $\Phi_m$ :

$$\begin{aligned} \phi_m(\mathbf{x}; \bar{\mathbf{x}}) &= \Phi_m(\bar{\mathbf{x}}) + \nabla \Phi_m(\bar{\mathbf{x}})'(\mathbf{x} - \bar{\mathbf{x}}) \\ &\quad - \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})' \check{\mathbf{C}}_m(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) \end{aligned} \quad (9)$$

with

$$\check{\mathbf{C}}_m(\mathbf{x}) = \text{diag}_j \{ \check{c}_{mj}(\mathbf{x}) \} \quad (10)$$

where  $\check{c}_{mj}(\cdot) > 0$  and  $\text{diag}\{\cdot\}$  denotes a diagonal matrix appropriately formed.

To make  $\phi_m$  satisfy the minorization conditions in (3), one has at least two choices for  $\check{c}_{mj}$ : “optimum curvature” (OC) and “maximum curvature” (MC). Those curvatures  $\check{c}_{mj}$  have the following form:

$$\begin{aligned} \check{c}_{mj}(\mathbf{x}) &= \max \left\{ \sum_{i \in S_m} a_{ij} a_i c_i([\mathbf{A}\mathbf{x}]_i)_+ \right. \\ &\quad \left. \frac{2\beta}{M} \sum_{k \in \mathcal{N}_j} w_{jk} \omega_\psi(x_j - x_k), \epsilon \right\} \end{aligned} \quad (11)$$

for some small value  $\epsilon > 0$  where  $a_i \triangleq \sum_{j=1}^p a_{ij}$  and  $\omega_\psi(t) \triangleq \dot{\psi}(t)/t$ . The functionals  $c_i(\cdot)$  are defined as follows. For OC, we define

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

and for MC,

$$c_i^{\text{MC}}(l) \triangleq \left[ -\ddot{h}_i(0) \right]_+, \quad (13)$$

where  $[x]_+ = \max\{x, 0\}$ . Detailed derivations of (11)–(13) can be found in [10]. The optimum curvature  $c_i^{\text{OC}}$  in (12) leads to faster convergence rates than the maximum curvature  $c_i^{\text{MC}}$  in (13), but it needs an “extra” backprojection per iteration.

The augmented objective function  $F$  defined in (4) with (9) is readily maximized with respect to  $\mathbf{x}$  over the box constraint  $\mathcal{X}$  as follows:

$$\begin{aligned} \mathbf{x}^{\text{new}} &= \left[ \left[ \sum_{m=1}^M \check{\mathbf{C}}_m(\bar{\mathbf{x}}_m) \right]^{-1} \right. \\ &\quad \left. \sum_{m=1}^M \left[ \check{\mathbf{C}}_m(\bar{\mathbf{x}}_m) \bar{\mathbf{x}}_m + \nabla \Phi_m(\bar{\mathbf{x}}_m) \right] \right]^+ \end{aligned} \quad (14)$$

where  $[\mathbf{x}]^+$  denotes the orthogonal projection of  $\mathbf{x} \in \mathbb{R}^p$  onto the box constraint  $\mathcal{X}$  and is easily computed componentwise as follows:  $[[\mathbf{x}]^+]_j = \text{median}\{0, x_j, U\}$  for all  $j$ . Using (14) in the step (T-1) leads to a new “transmission incremental optimization transfer (TRIOT)” algorithm. If  $M = 1$ , then TRIOT reduces to ordinary SPS [8].

### C. Acceleration

TRIOT-OC/MC<sup>1</sup> is slower initially than OS-SPS [8]. Here we discuss methods to accelerate TRIOT.

1) *Switch from OS-SPS to TRIOT*: It is a popular idea to switch from a nonconvergent yet initially fast OS type algorithm to a convergent non-OS algorithm at some point to take advantage of both fast initial convergence rates of OS methods and global convergence of non-OS methods.

We observed that it is very effective to switch to TRIOT from OS-SPS at the point where OS-SPS nearly gets to a limit cycle; even one single subiteration of TRIOT moves the iterate from the limit cycle very close to the optimal solution. The reason

<sup>1</sup>The second part of the algorithm name denotes a curvature.

is as follows: a group of the points in the limit cycle would be roughly centered around the optimal point and the update for TRIOT includes a weighted average of the points [see (14)].

However, in the early iterations, when OS-SPS is still far from the limit cycle and is making progress towards the optimal point, TRIOT is usually slower than OS-SPS due to the averaging of the past subiterates. So it is desirable to get to a limit cycle quickly using OS-SPS with *many* subsets and then switch to TRIOT.

2) *Precomputed Curvatures*: Forgoing monotonicity (in the augmented objective) and accordingly provable convergence, one can use for TRIOT the “precomputed curvatures (PC)” to obtain accelerated convergence rates. The “precomputed curvature (PC)” is obtained by approximating Newton’s curvatures [8], [10] as follows:

$$c_i^{\text{PC}} \triangleq \begin{cases} \frac{(y_i - r_i)^2}{y_i}, & y_i > r_i \\ 0, & \text{otherwise.} \end{cases}$$

It is an open question whether TRIOT-PC converges to an optimal solution. However, in our experiments, TRIOT-PC yielded the same limit as convergent algorithms like SPS-OC within numerical precision!

#### IV. RESULTS

To assess the performance of the proposed algorithms, we performed 2D attenuation map reconstructions from real PET data. We acquired PET data using a Siemens/CTI ECAT EX-ACT 921 PET scanner with rotating rod transmission sources [20]. We used an anthropomorphic thorax phantom (Data Spectrum, Chapel Hill, NC). The sinogram had 160 radial bins and 192 angles, and the reconstructed images were  $128 \times 128$  with 4.2 mm pixels. The system geometry was approximated with 3.375 mm wide strip integrals and 3.375 mm ray spacing. The total counts amounted to  $9.2 \times 10^5$ . We used the edge-preserving nonquadratic penalty (8) with  $\delta = 4 \times 10^{-4} \text{ mm}^{-1}$  and  $\beta = 2^{18.5}$ , chosen by visual inspection. A uniform image was used as a starting image. Images were reconstructed using SPS-MC/PC, OS-SPS(-PC), and TRIOT-MC/PC. For OS-SPS and TRIOT algorithms, we used 16 subsets (a moderate number) and 64 subsets (a little larger number than usual). For SPS and TRIOT, the performance (objective value or distance from the optimal image) with the optimum curvature (OC) in (12), which requires an extra backprojection per iteration, was between those with MC and PC; and the results with OC are not shown here. See [15] for reconstructed images not shown here.

Fig. 1 shows normalized  $\Phi$  difference versus iteration number for different algorithms using 16 subsets. The normalized  $\Phi$  difference is defined as  $(\Phi(\hat{x}^{\text{PL}}) - \Phi(\hat{x}^n)) / (\Phi(\hat{x}^{\text{PL}}) - \Phi(\hat{x}^0))$  where  $\hat{x}^{\text{PL}}$  is a maximizer of the PL objective. The optimal image  $\hat{x}^{\text{PL}}$  was estimated by 30 iterations of OS-SPS with 16 subsets followed by 800 iterations of SPS-OC. TRIOT algorithms were initialized by running one iteration of OS-SPS. So were the SPS algorithms for a fair comparison. Although OS-SPS showed a fast initial convergence rate, it became stuck

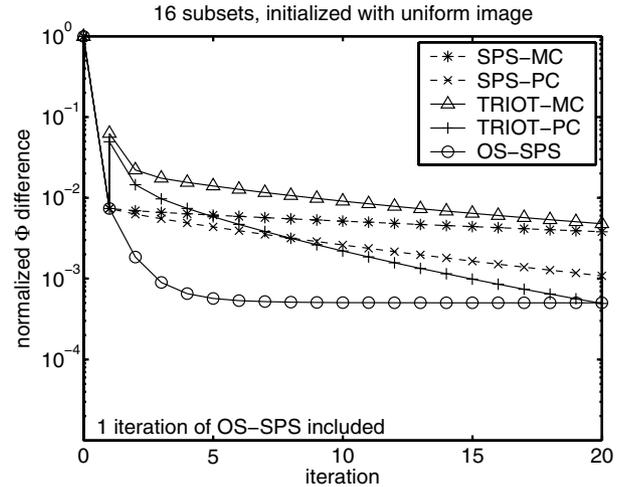


Fig. 1. Comparison of non-OS algorithms (SPS-MC/PC), an OS algorithm (OS-SPS), and incremental optimization transfer algorithms (TRIOT-MC/PC) for 2D attenuation map reconstruction using real PET data. This figure shows  $(\Phi(\hat{x}^{\text{PL}}) - \Phi(\hat{x}^n)) / (\Phi(\hat{x}^{\text{PL}}) - \Phi(\hat{x}^0))$  versus iteration number where  $\hat{x}^{\text{PL}}$  is the PL optimal image. The OS-SPS and TRIOT algorithms used 16 subsets, and TRIOT and SPS algorithms included one initial iteration of OS-SPS.

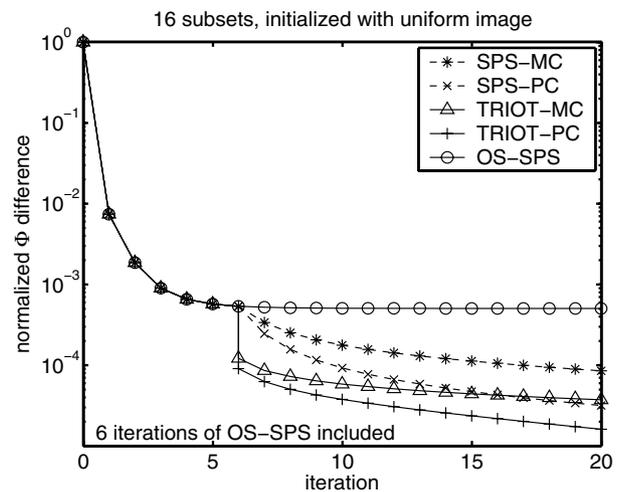


Fig. 2. Same as Fig. 1, but six initial iterations of OS-SPS were included for TRIOT and SPS algorithms.

at a suboptimal point whereas other methods continued to improve in terms of objective values. The TRIOT algorithms were outperformed by other algorithms in early iterations since the built-in averaging in TRIOT slows down convergence when a limit cycle has not reached yet. However, TRIOT-MC and TRIOT-PC eventually outrun SPS-MC and SPS-PC, respectively.

To investigate the performance of TRIOT algorithms after OS-SPS reaches a limit cycle, we performed 6 iterations of OS-SPS, which is sufficient to get close to a limit cycle, and then applied TRIOT (and SPS as well). Fig. 2 shows that TRIOT

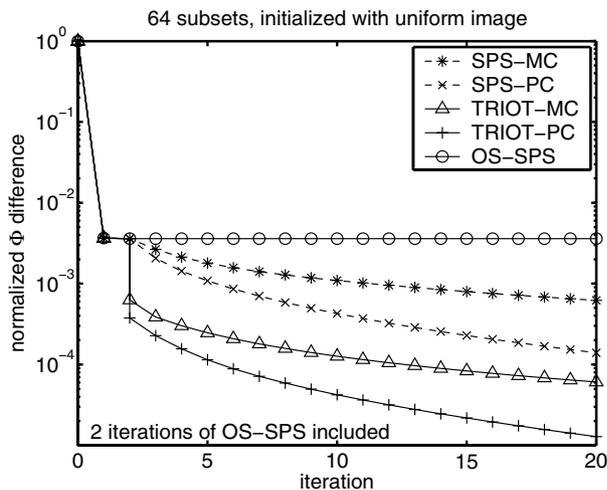


Fig. 3. Same as Fig. 1, but 64 subsets are used for OS-SPS and TRIOT algorithms, and two iterations of OS-SPS are included initially for TRIOT and SPS algorithms.

yielded considerable improvement at iteration 6 where TRIOT was first applied. TRIOT-MC and TRIOT-PC converge faster than SPS-MC and SPS-PC, respectively, which are similarly initialized by 6 iterations of OS-SPS. This shows that it is effective to switch from OS-SPS to TRIOT when OS-SPS almost reaches a limit cycle. However, it is inconvenient to predict how many iterations are required for OS-SPS to arrive at a limit cycle.

Fig. 3 shows normalized  $\Phi$  difference versus iteration number when 64 subsets are used. As the number of subsets increased to 64, the initial convergence rate of OS-SPS became faster (even a couple of iterations led to a limit cycle), but OS-SPS stagnated at a worse image. Meanwhile, the TRIOT algorithms were quite effective even though they used only a couple of iterations of OS-SPS as initialization, and they outperformed the SPS algorithms initialized similarly.

## V. CONCLUSION

We presented a broad family of incremental optimization transfer algorithms by generalizing the incremental EM family. The incremental optimization transfer algorithms usually show faster convergence rates than ordinary optimization transfer methods like EM or SPS, but they are convergent.

We also developed a particular incremental optimization transfer algorithm for transmission tomography by using separable quadratic surrogates: TRIOT algorithms. We found that it is very effective to switch from OS-SPS to TRIOT when OS-SPS nearly reaches a limit cycle. When reasonably many subsets are used, as few as a couple of iteration of OS-SPS can be sufficient to get close to a limit cycle (although it would depend on the degree of regularization and the size of the problem).

## REFERENCES

- [1] S. Ahn and J. A. Fessler, "Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 613–626, May 2003.
- [2] R. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse and other variants," in *Learning in Graphical Models*, M. I. Jordan, Ed. Dordrecht: Kluwer, 1998, pp. 255–268.
- [3] A. J. R. Gunawardana, "The information geometry of EM variants for speech and image processing," Ph.D. dissertation, The Johns Hopkins Univ., Baltimore, MD, Apr. 2001.
- [4] I. T. Hsiao, A. Rangarajan, and G. Gindi, "A provably convergent OS-EM like reconstruction algorithm for emission tomography," in *Proc. SPIE 4684, Medical Imaging 2002: Image Proc.*, 2002, pp. 10–19.
- [5] I. Hsiao, A. Rangarajan, and G. Gindi, "A new convergent MAP reconstruction algorithm for emission tomography using ordered subsets and separable surrogates," in *Proc. IEEE Intl. Symp. Biomedical Imaging*, 2002, pp. 409–412.
- [6] P. Khurd, I. T. Hsiao, A. Rangarajan, and G. Gindi, "A globally convergent regularized ordered-subset EM algorithm for list-mode reconstruction," *IEEE Trans. Nuc. Sci.*, vol. 51, no. 3, pp. 719–725, June 2004.
- [7] D. Blatt, A. Hero, and H. Gauchman, "An incremental gradient method that converges with a constant step size," *SIAM J. Optim.*, Submitted.
- [8] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–2851, Nov. 1999.
- [9] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, Feb. 2004.
- [10] H. Erdoğan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Trans. Med. Imag.*, vol. 18, no. 9, pp. 801–814, Sept. 1999.
- [11] W. Byrne and A. Gunawardana, "Comments on "Efficient training algorithms for HMMs using incremental estimation"," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 6, pp. 751–754, Nov. 2000.
- [12] K. Lange, D. R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," *J. Computational and Graphical Stat.*, vol. 9, no. 1, pp. 1–20, Mar. 2000.
- [13] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.
- [14] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–609, Dec. 1994.
- [15] S. Ahn, J. A. Fessler, D. Blatt, and A. O. Hero, "Convergent incremental optimization transfer algorithms: Application to tomography," *IEEE Trans. Med. Imag.*, Submitted.
- [16] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomogr.*, vol. 8, no. 2, pp. 306–316, Apr. 1984.
- [17] J. A. Fessler, "Statistical image reconstruction methods for transmission tomography," in *Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis*, M. Sonka and J. Michael Fitzpatrick, Eds. Bellingham: SPIE, 2000, pp. 1–70.
- [18] P. J. Huber, *Robust statistics*. New York: Wiley, 1981.
- [19] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Trans. Med. Imag.*, vol. 9, no. 4, pp. 439–446, Dec. 1990. Corrections, T-MI, 10:2(288), June 1991.
- [20] K. Wienhard, L. Eriksson, S. Grootoink, M. Casey, U. Pietrzyk, and W. D. Heiss, "Performance evaluation of a new generation positron scanner ECAT EXACT," *J. Comput. Assist. Tomogr.*, vol. 16, no. 5, pp. 804–813, Sept. 1992.