Convergent Incremental Optimization Transfer Algorithms: Application to Tomography

Sangtae Ahn, *Member, IEEE*, Jeffrey A. Fessler, *Senior Member, IEEE*, Doron Blatt, *Student Member, IEEE*, and Alfred O. Hero, *Fellow, IEEE*

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 [1], and methods based on the incremental expectation maximization (EM) approach [2]. This paper generalizes the incremental EM approach [3] 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 allows the use of a very broad family of non-EM surrogate functions, enabling the development of new algorithms [4]. 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) [5] which yield closed-form maximization steps. We found it is very effective to achieve fast convergence rates by starting with an OS algorithm with a large number of subsets and switching to the new "transmission incremental optimization transfer (TRIOT)" algorithm. We discuss how to determine the number of subsets and the switching point. Results show that TRIOT is faster than nonincremental ordinary SPS and even OS-SPS yet is convergent. We also show by simulation that nonconvergent OS algorithms lead to higher variance than convergent TRIOT particularly for a large number of subsets.

Index Terms

Statistical image reconstruction, maximum likelihood estimation, penalized-likelihood estimation, incremental optimization transfer, transmission tomography

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 tomographic image reconstruction due to their *remarkably fast "convergence" rates* [1], [5]–[13]. For example, ordered subsets expectation maximization (OS-EM) provides an order-of-magnitude acceleration over its non-OS counterpart, EM [6]. The incremental gradient type algorithms are also found in convex programming [14]–[17]. The ordered subsets (or incremental) idea is to perform the update iteration incrementally by sequentially (or sometimes randomly [15], [16]) using a subset of the data. Row-action methods [18] including algebraic reconstruction techniques (ART) [19], [20] can also be viewed as OS type algorithms in which each subset corresponds to a single measurement.

The OS algorithms apply successfully to problems where an objective function of interest is a sum of a large number of component functions. Because of the assumed statistical independence of tomographic data, such sums arise in statistical tomographic reconstruction problems including penalized-likelihood (PL) [equivalently, maximum *a posteriori* (MAP)] or maximum likelihood (ML) reconstruction. Typically, the OS methods partition the component functions into several subsets, each corresponding to a subset of the projection views, and each subset defines a subobjective function. Then they update the image estimate by using, in a specified cyclic order, the gradient of a subobjective function as an approximate gradient of the objective function.

If the subset gradients are suitably balanced, then the gradient approximation can be quite reasonable when the iterates are far from a maximizer. Thus OS methods initially accelerate convergence in the sense that less computation is required to achieve nearly the same level of objective increase as with non-OS methods. However, ordinary (unrelaxed) OS algorithms such as OS-EM [6], RBI-EM [8], and OS-SPS (or OSTR in a context of transmission tomography) [5] generally do not converge to an optimal solution but rather approach a suboptimal limit cycle that consists of as many points as there are subsets. In fact, due to their subset-dependent scaling (or preconditioning) matrices [1], OS-EM and RBI-EM in their original forms [6], [8] usually do not converge to the optimal point even if relaxed.

Convergence to an optimal solution is important for any algorithm for optimization problems, particularly in medical applications where reliability and stability are essential. Convergence is perhaps more

critical for PL (or MAP) reconstruction than for ML because one usually does not run ML reconstruction algorithms until convergence. For example, the image shown in Fig. 4(c), which corresponds to one point of a limit cycle generated by an OS algorithm, looks noticeably different from the PL solution image shown in Fig. 4(b) (see Section IV-A for details). Nonconvergent OS algorithms can lead to higher variance as the number of subsets increases as shown in Section IV-B. It is desirable to achieve both fast initial convergence rates (typical of OS algorithms) and global convergence. 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 [1], [7], [9]–[12], [14]–[16], [21]–[23]. Suitably relaxed algorithms can be shown to converge to an optimal solution under certain regularity conditions[1] [1]. However, since relaxation parameters should be scheduled to converge to zero to ensure global convergence, relaxed OS algorithms have slow asymptotic convergence rates. Also, inappropriately chosen (*e.g.*, too rapidly decreasing) relaxation parameters can make initial convergence rates even worse than those of non-OS algorithms. On the other hand, overly large relaxation parameters can lead to unstable or divergent behavior. Finding good relaxation parameters (in terms of convergence rates) may require some experimentation and trial-and-error; as a rule of thumb, for properly scaled OS algorithms such as modified BSREM and relaxed OS-SPS, one should initialize the relaxation parameter close to unity and decrease it gradually as convergence to a limit cycle nears [1]. One may optimize a few initial relaxation parameters by training when a training set is available for a particular task [7], [20]. Or one could use the dynamic stepsize rule in [15], [16], but that method needs to compute the objective value at every update, which is computationally expensive in tomographic reconstruction problems. Alternatively, to achieve convergence, one could decrease the number of subsets as iterations proceed or could use hybrid methods that combine OS and non-OS algorithms [25]. However, the schedules for decreasing the number of subsets and the parameters for the hybrid algorithms are as inconvenient to determine as relaxation parameters for relaxed OS algorithms.

Incremental EM algorithms do not require user-specified relaxation parameters [3]. They are convergent yet faster than ordinary EM algorithms although slower initially than nonconvergent OS-EM type algorithms [2], [26], [27]. Such incremental EM algorithms have been applied to emission tomography

---

[1]One of these conditions is the (strict) concavity of the objective function, which excludes the nonconcave transmission tomography problem [24].

[2], [13], [27], [28].

Recently, Blatt *et al.* proposed a convergent incremental gradient method, called incremental aggregated gradient (IAG), that does not require relaxation parameters [29]. The IAG method computes a single subset gradient for each update but aggregates it with the stored subset gradients that were computed in previous iterations. The use of the aggregated gradient to approximate the full gradient of the objective function leads to convergence. Similarly, as discussed below, the use of the sum of surrogate functions (rather than a single surrogate function) to approximate a minorizing function yields convergent algorithms.

In this paper we generalize the incremental EM algorithms by introducing an approach called "incremental optimization transfer"; this is akin to the generalization of the EM algorithms [30] by the optimization transfer principles [4]. In fact, the broad family of "incremental optimization transfer algorithms" includes the ordinary optimization transfer algorithms (*e.g.*, EM), also referred to as MM (minorize-maximize or majorize-minimize) algorithms in [31], as a special case where the objective function consists of only one subobjective function.

In the incremental optimization transfer approach, for *each* subobjective function, we define an augmented vector that has the same size as the parameter vector to be estimated. The augmented vector plays a role as an expansion point at which a minorizing surrogate function is defined for the subobjective function (see Section II for details). The sum of the surrogate functions defines an augmented objective that is a function of the parameter vector and the augmented vectors. With surrogate functions satisfying usual minorization conditions [4], [24], a solution to the problem of maximizing the original objective can be found by maximizing the augmented objective instead. Applying a block coordinate ascent approach to the augmented problem leads to a new class of "incremental optimization transfer algorithms." By construction, incremental optimization transfer algorithms are monotonic in the augmented objective though not necessarily in the original objective; nevertheless, global convergence is ensured under mild regularity conditions. Incremental optimization transfer algorithms show faster convergence rates than their nonincremental counterparts like EM [2], [3], [27].

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. These methods are particularly useful for large-scale problems where the objective function is expressed as a sum of several subobjective 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) [5]. Such quadratic surrogates simplify the maximization. In contrast, the standard EM surrogates for transmission tomography do not

have a closed-form maximizer in the "M-step" [32].

The proposed "transmission incremental optimization transfer (TRIOT)" algorithm is convergent yet converges faster than ordinary SPS [5]; it can be further accelerated by the enhancement method in [33] or by initializing through a few iterations of OS-SPS (see Section III for details). It is parallelizable, and the nonnegativity constraint is naturally enforced. In addition, it is easily implemented for system models that use factored system matrices [34], [35] whereas pixel-grouped coordinate ascent based methods require column access of the system matrix [36]–[39].

Section II describes the incremental optimization transfer algorithms in a general framework and discusses their convergence properties. Section III develops incremental optimization transfer algorithms for transmission tomography, and addresses acceleration methods. Section IV provides simulation and real PET data results, and Section V gives conclusions.

II. Incremental Optimization Transfer

A. Incremental Optimization Transfer Algorithms

Most objective functions of interest in image reconstruction can be expressed as a sum of subobjective functions:[2]

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

where $\Phi_m : \mathcal{X} \subset \mathbb{R}^p \to \mathbb{R}$ is a continuously differentiable function whose domain $\mathcal{X}$ is a nonempty, convex and closed set. The $p \times 1$ vector $\boldsymbol{x}$ represents an image, *e.g.*, attenuation coefficients or radioactivity, depending on applications. Consider the following optimization problem:

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

Usually there exists no closed-form solution to the above problem, so one must apply iterative algorithms. Assume that for each subobjective function Φ_m, we find a surrogate function $\phi_m(\cdot;\cdot) : \mathcal{X}^2 \subset \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$, where $\mathcal{X}^n$ denotes the n-ary Cartesian product over the set $\mathcal{X}$, such that (i) it is easier to maximize with respect to the first argument than Φ_m and (ii) it satisfies the following "minorization" conditions [24], [41]:

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

---

[2]Such functions are said to be *additive-separable* in [14]; and to be *partially separable* [40] when each $\Phi_m(\boldsymbol{x})$ is a function of fewer components of $\boldsymbol{x} \in \mathbb{R}^p$ than p.

where $\bar{\boldsymbol{x}}$ is an augmented variable that is called an expansion point in [39]. It follows from the above conditions that

$$\Phi_m(\boldsymbol{x}) - \Phi_m(\bar{\boldsymbol{x}}) \geq \phi_m(\boldsymbol{x}; \bar{\boldsymbol{x}}) - \phi_m(\bar{\boldsymbol{x}}; \bar{\boldsymbol{x}}), \qquad \forall \boldsymbol{x}, \bar{\boldsymbol{x}} \in \mathcal{X}. \tag{4}$$

In other words, choosing $\boldsymbol{x}$ such that $\phi_m(\boldsymbol{x}; \bar{\boldsymbol{x}}) \geq \phi_m(\bar{\boldsymbol{x}}; \bar{\boldsymbol{x}})$ ensures that $\Phi_m(\boldsymbol{x}) \geq \Phi_m(\bar{\boldsymbol{x}})$, which is called a "monotonicity" property. When there is only one subobjective function ($M = 1$, $\Phi = \Phi_1$, and $\phi = \phi_1$), one can construct the following "optimization transfer" algorithm:

$$\boldsymbol{x}^{n+1} = \arg \max_{\boldsymbol{x} \in \mathcal{X}} \phi(\boldsymbol{x}; \boldsymbol{x}^n).$$

Then we have $\Phi(\boldsymbol{x}^{n+1}) \geq \Phi(\boldsymbol{x}^n)$ due to the monotonicity property of the surrogate. The reader who would like more details on optimization transfer principles is referred to [4].

Now we construct an incremental version of the optimization transfer algorithm. First, define the following "augmented" objective function:

$$F(\boldsymbol{x}; \bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_M) = \sum_{m=1}^{M} \phi_m(\boldsymbol{x}; \bar{\boldsymbol{x}}_m). \tag{5}$$

Noting that $\arg \max_{\bar{\boldsymbol{x}} \in \mathcal{X}} \phi_m(\boldsymbol{x}; \bar{\boldsymbol{x}}) = \Phi_m(\boldsymbol{x})$ due to (3), one can rewrite the optimization problem in (2) as follows:

$$\begin{aligned} &\text{maximize } F(\boldsymbol{x}; \bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_M) \\ &\text{subject to } (\boldsymbol{x}; \bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_M) \in \mathcal{X}^{M+1}. \end{aligned} \tag{6}$$

By construction, $\boldsymbol{x}^* \in \mathcal{X}$ is an optimal solution of (2) if and only if $(\boldsymbol{x}^*; \bar{\boldsymbol{x}}_1^*, \ldots, \bar{\boldsymbol{x}}_M^*) \in \mathcal{X}^{M+1}$ is an optimal solution of (6) for some $(\bar{\boldsymbol{x}}_1^*, \ldots, \bar{\boldsymbol{x}}_M^*) \in \mathcal{X}^M$. Therefore, we can find a solution to problem (2) by maximizing F with respect to $(\boldsymbol{x}; \bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_M)$. By alternating between updating $\boldsymbol{x}$ and one of the $\bar{\boldsymbol{x}}_m$'s, we obtain an "incremental optimization transfer algorithm" outlined in Table I, where we assume that there exists one or possibly more maximizers in (T-1), and "$\arg \max$" denotes one of those maximizers.

In many applications where (T-1) has a closed-form solution, the computational cost for computing $\boldsymbol{x}^{\text{new}}$ is smaller, usually by a factor of M, when only one of $\bar{\boldsymbol{x}}_m$'s has been incrementally updated than when all $\bar{\boldsymbol{x}}_m$'s have been simultaneously updated. For example, in tomographic reconstruction problems we focus on, computing $\boldsymbol{x}^{\text{new}}$ usually involves forward projections of $\bar{\boldsymbol{x}}_m$'s, and we have only to project one of $\bar{\boldsymbol{x}}_m$'s and use previously computed projections for other $\bar{\boldsymbol{x}}_m$'s.

The incremental optimization transfer algorithm shown in Table I can be viewed as a block coordinate ascent algorithm for maximizing F with respect to $(\boldsymbol{x}; \bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_M)$ [42, p. 270]. It monotonically increases the augmented objective function F, but not necessarily the original objective function Φ [43]. The incremental approach ($M > 1$) usually leads to faster convergence rates than nonincremental

methods ($M = 1$) [3]. The incremental EM algorithms [3], [28] including COSEM [2], [27] are a special case where the surrogates ϕ_m are constructed by EM principles as described in Appendix A.

If one were to maximize just one of the ϕ_m's instead of the sum shown in (5), then one would have ordinary OS type algorithms. Although this greedy approach usually yields fast initial convergence rates, the OS type algorithms are not monotonic in F nor in Φ. Since OS algorithms consider only a part of the objective function for each update and an optimal point is characterized only by the whole objective function, OS algorithms usually cannot converge to the optimal point. In contrast, the incremental optimization transfer methods involve the augmented objective function that reflects the whole objective function yet is updated incrementally.

Although we focus on monoenergetic transmission tomography in Section III, the incremental optimization transfer is a very general method which can be applied to a variety of problems where an objective function is a sum of functions as in (1) and the OS approach applies: for example, polyenergetic transmission tomography [44], confocal microscopy [45], and emission tomography [46].

For incremental optimization transfer algorithms one must store M vectors $\{\bar{x}_m\}_{m=1}^{M}$, so one needs more memory compared to ordinary OS algorithms. This can be a practical limitation when M is very large for large-sized problems.

B. Convergence Properties

Since incremental optimization transfer algorithms monotonically increase the augmented objective F, the sequence of augmented objective values converges to some value in the usual case where F has an upper bound. The question of whether the algorithms really converge to a maximizer of (2) is addressed next.

Define a *solution set* as the collection of stationary points of (2):

$$\Gamma \triangleq \{\boldsymbol{x}^* \in \mathcal{X} : \nabla\Phi(\boldsymbol{x}^*)'(\boldsymbol{x} - \boldsymbol{x}^*) \leq 0, \quad \forall \boldsymbol{x} \in \mathcal{X}\}, \tag{7}$$

where $'$ denotes matrix or vector transpose, and we assume $\Gamma \neq \emptyset$. Each element of the solution set Γ satisfies the first-order necessary condition for a local maximizer of Φ over $\mathcal{X}$ [42, p. 194]. We want algorithms to converge to some point in Γ. If the objective function Φ is concave, then the condition defining Γ is sufficient for optimality, and Γ is the set of (possibly multiple) global maximizers of Φ over $\mathcal{X}$ [42, p. 194]. If Φ is strictly concave, then Γ is the singleton set whose element is a unique global maximizer [42, p. 685]. On the other hand, for a nonconcave objective function Φ (as in Section III), the solution set Γ could contain local maximizers and even local minimizers. It is difficult to guarantee finding

a global maximizer of a nonconcave objective function that may have multiple local maxima. However, the hope is that, with an initial point reasonably close to a global maximizer, the iterates generated by a monotonic algorithm will approach the global maximizer (see [39] for discussion about convergence to a globally optimal point).

In Appendix B, we show that every limit point[3] of the sequence generated by an incremental optimization transfer algorithm is an element of the solution set Γ of stationary points regardless of initial points[4] when the following general sufficient conditions hold: (i) each Φ_m is differentiable, and each $\phi_m(\cdot;\cdot)$ is differentiable with respect to the first argument and is continuous with respect to the second argument, (ii) the iterates are bounded (*e.g.*, $\mathcal{X}$ is a bounded set), (iii) the surrogates ϕ_m satisfy the minorization conditions in (3), (iv) the gradients of Φ_m and $\phi_m(\cdot;\bar{x})$ with respect to the first argument match (see Condition 2 in Appendix B), (v) the maximizer in (T-1) is defined uniquely (*e.g.*, $\phi_m(\cdot;\bar{x}_m)$ is strictly concave for any $\bar{x}_m$), and (vi) the maximizer of the augmented objective function is bounded (see Condition 4 in Appendix B). Consequently, if the objective function Φ is strictly concave, then the algorithm converges to the global maximizer. For a nonconcave objective function Φ, if the points in Γ are isolated, the algorithm will still converge to some stationary point in Γ that we hope is a global maximizer or at least a local maximizer (see Appendix B). It is an open question whether optimization transfer algorithms converge to nonisolated stationary points (see [39] for a discussion of this issue).

Although we focus on transmission tomography in the next section, we briefly discuss the applicability of our convergence proofs to the emission case. The convergence proofs in Appendix B do not apply to classical ML-EM and COSEM for the emission case in their original forms in [32], [49] and [2], [27] respectively since the EM surrogates used in those algorithms blow up to (negative) infinity on the boundary of the nonnegativity constraint set and therefore they violate the aforementioned sufficient conditions. The readers are referred to [50] and [51] for convergence proofs for ML-EM and COSEM respectively for the emission case. However, to avoid the boundary problem one can use a slightly modified EM surrogate in [38, Eq. (20)] for the usual case where there are nonzero contributions from background events such as scatter and randoms. Our convergence proofs apply to those modified ML-EM

---

[3]Recall the distinction between a limit and a limit point. A point $\check{x}$ is called *a limit* of a sequence $\{x^n\}$ if $\forall \epsilon > 0$, $\exists N$ such that $\forall n > N$, $\|\check{x} - x^n\| < \epsilon$. On the other hand, a point $\bar{x}$ is called *a limit point* of a sequence $\{x^n\}$ if $\forall \epsilon > 0$, $\forall N$, $\exists n > N$ such that $\|\bar{x} - x^n\| < \epsilon$, in other words, if there exists a subsequence $\{x^{n_k}\}$ whose limit is $\bar{x}$.

[4]Some authors define *global convergence* as the property that limit points of the sequence generated by an algorithm are stationary points of the problem [47, p. 228] or that limits are stationary points [48, p. 312], irrespective of starting points. We adopt the former convention here.

(called "ML-EM-3" in [38]) and COSEM algorithms. Moreover, the modified EM surrogate is known to accelerate convergence rates [38].

See [52, Appendix F] for an asymptotic local convergence rate analysis and an illustrative one-parameter example for a comparison of the convergence rates of incremental and nonincremental algorithms.

III. APPLICATION TO TRANSMISSION TOMOGRAPHY

In this section we develop a particular incremental optimization transfer algorithm for transmission tomographic reconstruction. We use quadratic surrogates [5], [24] rather than EM surrogates in (23) because the standard complete-data proposed in [32] for transmission tomography does not yield a closed-form solution for M-step [53]. Using quadratic surrogates is not limited to the transmission case [54]–[56].

A. Problem

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

$$y_i \sim \text{Poisson}\left\{b_i e^{-[\boldsymbol{Ax}]_i} + r_i\right\}, \quad i = 1, \ldots, N \tag{8}$$

where y_i denotes the transmission measurement of the ith detector, b_i denotes the blank scan counts of the ith detector, r_i denotes the mean number of background counts, and $[\boldsymbol{Ax}]_i = \sum_{j=1}^{p} a_{ij} x_j$ represents the ith line integral of the attenuation map in which x_j is the unknown attenuation coefficient in the jth pixel, $\boldsymbol{A} = \{a_{ij}\}$ 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 $\hat{\boldsymbol{x}}^{\text{PL}}$ which is defined by

$$\hat{\boldsymbol{x}}^{\text{PL}} = \arg \max_{\boldsymbol{x} \in \mathcal{X}} \Phi(\boldsymbol{x}), \quad \Phi(\boldsymbol{x}) = L(\boldsymbol{x}) - \beta R(\boldsymbol{x}) \tag{9}$$

where the objective function Φ, which can be nonconcave when $r_i \neq 0$ [24], includes the log-likelihood

$$L(\boldsymbol{x}) = \sum_{i=1}^{N} h_i([\boldsymbol{Ax}]_i)$$

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

and a roughness penalty

$$R(\boldsymbol{x}) = \frac{1}{2} \sum_{j=1}^{p} \sum_{k \in \mathcal{N}_j} w_{jk} \psi(x_j - x_k). \tag{10}$$

The box constraint set is defined by

$$\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^p : 0 \leq x_j \leq U, \quad \forall j\}. \tag{11}$$

The nonnegativity restriction is imposed on physical grounds, and 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. The reason for using the box constraint rather than the usual nonnegativity constraint is that the convergence proofs in Appendix B need the iterates to be bounded. However, imposing upper bounds is not overly restrictive in a sense that one can choose a physically meaningful upper bound for attenuation coefficients, and the image estimate $\hat{x}$ is unlikely to be affected by U if one chooses an arbitrarily large U. In practice, if the upper bound happens to be active by some iterate, then the user could re-run the algorithm with a larger bound.

In the penalty function (10), the function ψ is a symmetric and convex potential function, $\mathcal{N}_j$ represents a neighborhood of the jth pixel, β is a regularization parameter that controls the smoothness in reconstructed images, and w_{jk} are weights (ordinarily, $w_{jk} = 1$ for horizontal and vertical neighboring pixels, and $w_{jk} = 1/\sqrt{2}$ for diagonal neighboring pixels). We assume the potential function ψ satisfies some conditions given in [24], [57, p. 184]. We used the following edge-preserving nonquadratic potential function[5] in our PL reconstruction results [58]:

$$\psi(t) = \delta^2[|t/\delta| - \log(1 + |t/\delta|)] \tag{12}$$

for some $\delta > 0$. We assume that appropriate β and δ are prespecified. To design optimal regularization functions is application-dependent and is beyond the scope of this paper. See [59]–[64] for quadratic penalty design.

B. Transmission Incremental Optimization Transfer (TRIOT)

We decompose the objective function Φ into the following subobjective functions:

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

where $\{S_m\}_{m=1}^M$ is a partition of $\{1, \ldots, N\}$. We use the usual subsets corresponding to downsampled projection angles [6]. Consider the following *separable* quadratic surrogate ϕ_m for the subobjective function Φ_m:

$$\phi_m(\boldsymbol{x}; \bar{\boldsymbol{x}}) = \Phi_m(\bar{\boldsymbol{x}}) + \nabla\Phi_m(\bar{\boldsymbol{x}})'(\boldsymbol{x} - \bar{\boldsymbol{x}}) - \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})'\breve{\boldsymbol{C}}_m(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}) \tag{13}$$

with

$$\breve{\boldsymbol{C}}_m(\boldsymbol{x}) = \mathrm{diag}_j\{\breve{c}_{mj}(\boldsymbol{x})\} \tag{14}$$

---

[5]Note that ψ is twice differentiable: $\dot{\psi}(t) = \delta \cdot t/(\delta + |t|)$ and $\ddot{\psi}(t) = \delta^2/(\delta + |t|)^2$.

where $\check{c}_{mj}(\cdot) > 0$ and diag$\{\cdot\}$ denotes a diagonal matrix appropriately formed. The surrogates ϕ_m in (13) satisfy Conditions 2 and 3 in Appendix B.

To make ϕ_m additionally 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:

$$\check{c}_{mj}(\boldsymbol{x}) = \max\left\{ \sum_{i \in S_m} a_{ij} a_i c_i([\boldsymbol{Ax}]_i) + \frac{2\beta}{M} \sum_{k \in \mathcal{N}_j} w_{jk} \omega_\psi(x_j - x_k), \ \epsilon \right\} \tag{15}$$

for some small value $\epsilon > 0$ where $a_i \overset{\triangle}{=} \sum_{j=1}^p a_{ij}$ and $\omega_\psi(t) \overset{\triangle}{=} \dot{\psi}(t)/t$. The use of ϵ keeps the curvatures $\check{c}_{mj}$ positive and so the augmented objective function defined in (5) has a unique maximizer with respect to the first argument due to strict concavity. The functions $c_i(\cdot)$ are defined as follows. For OC, we define[6]

$$c_i^{\mathrm{OC}}(l) \overset{\triangle}{=} \begin{cases} \left[-2\dfrac{h_i(0) - h_i(l) + \dot{h}_i(l) \cdot l}{l^2} \right]_+ , & l > 0 \\ \left[-\ddot{h}_i(0) \right]_+ , & l = 0, \end{cases} \tag{16}$$

and for MC,

$$c_i^{\mathrm{MC}}(l) \overset{\triangle}{=} \left[-\ddot{h}_i(0) \right]_+ , \tag{17}$$

where $[x]_+ = \max\{x, 0\}$. On the right side in (15), the first term corresponds to the curvature of a quadratic surrogate for the log-likelihood part, and the second term for the penalty part. The optimum curvature c_i^{OC} in (16) is computed as the lowest curvature of 1D quadratic surrogates, satisfying the minorization conditions, for the marginal log-likelihood $h_i(l)$. A low curvature of a surrogate implies a wide paraboloid which in turn implies a large stepsize, that is, fast convergence rate [24]. Therefore, the optimal curvature is "optimal" in a sense that it leads to the fastest convergence rate (per iteration). However, one needs an "extra" backprojection[7] for computing the first summation in (15). On the other hand, the c_i^{MC} is the maximum curvature of the marginal log-likelihood $h_i(l)$ over $l \geq 0$. Therefore, c_i^{MC} is a constant and the first summation in (15) can be precomputed and stored. Detailed derivations of (15)–(17) can be found in [24]. We leave the second summation in (15) as a function of $\boldsymbol{x}$ even for MC since its computation is usually cheap compared to projection and backprojection operations unless M is too large.

---

[6]Using L'Hôpital's rule [65, p. 234], it can be shown that $c_i^{\mathrm{OC}}(l)$ is continuous for $l \geq 0$.

[7]For some "on-the-fly" projector/backprojector pairs, most of the overhead is computing the a_{ij} values for each i, so it may be possible to evaluate the "extra" backprojection with modest additional computation.

The augmented objective function F defined in (5) with (13) is readily maximized with respect to $\boldsymbol{x}$ over the box constraint $\mathcal{X}$ as follows:

$$\boldsymbol{x}^{\text{new}} = \mathcal{P}_{\mathcal{X}}\left(\left[\sum_{m=1}^{M} \breve{\boldsymbol{C}}_m(\bar{\boldsymbol{x}}_m)\right]^{-1} \sum_{m=1}^{M}\left[\breve{\boldsymbol{C}}_m(\bar{\boldsymbol{x}}_m)\bar{\boldsymbol{x}}_m + \nabla\Phi_m(\bar{\boldsymbol{x}}_m)\right]\right) \tag{18}$$

where $\mathcal{P}_{\mathcal{X}}(\boldsymbol{x})$ is the orthogonal projection of $\boldsymbol{x} \in \mathbb{R}^p$ onto $\mathcal{X}$ and is easily computed componentwise as follows: $[\mathcal{P}_{\mathcal{X}}(\boldsymbol{x})]_j = \text{median}\{0, x_j, U\}$ for all j. Using (18) in the step (T-1) leads to a new "transmission incremental optimization transfer (TRIOT)" algorithm, which is outlined in Table II. When $M = 1$, TRIOT reduces to ordinary SPS [5]. The TRIOT update begins after $n_{\text{iter}}^{\text{OS}}$ (≥ 1) iteration(s) of OS-SPS [5] (see the next subsection for OS-SPS in detail). The strategy to switch from OS-SPS to TRIOT is discussed in Section III-D. Running initially (at least) one iteration of OS-SPS is more effective than initializing all $\bar{\boldsymbol{x}}_m$'s to be the same image (*e.g.*, a FBP or uniform image) because both cases require nearly the same computation yet one can take advantage of fast initial convergence rates of OS-SPS.

In Table II, a TRIOT using MC in (17), called TRIOT-MC[8], is outlined; however, the OC case in (16) can be easily included. The two steps (T-1) and (T-2) in Table I are combined in Table II. In (T-5), one can avoid the summation $\sum_{l=1}^{M}$ at every subiteration by maintaining that sum as a state vector that is updated incrementally as in [2], [27], [33]. One iteration, indexed by n, of TRIOT-MC requires one projection and one backprojection operation while TRIOT-OC needs an extra backprojection [see (15) and (16)].

The discussion and proofs for global convergence given in Section II-B and Appendix B apply to TRIOT. When $r_i = 0$ for all i, the algorithm converges to the optimal solution under mild conditions[9] since the PL objective for transmission tomography is strictly concave [66]. In the case where $r_i \neq 0$, the objective function is not necessarily concave [24], and we have a weaker conclusion that every limit point of a sequence generated by TRIOT is a stationary point. However, in our practical experience, we obtained the same limit in all experiments with different initializations, suggesting that suboptimal local maxima are rare, or are far from reasonable starting images.

C. OS-SPS

Since we use OS-SPS in initializing and accelerating TRIOT, we briefly review OS-SPS [5] for completeness. For each subiteration, indexed by m, maximizing the mth subobjective $\phi_m(\cdot; \bar{\boldsymbol{x}}_m)$ in (13)

---

[8]The second part denotes a specific curvature used (*e.g.*, SPS-OC).

[9]The potential function ψ is strictly convex, and $\boldsymbol{A}'\boldsymbol{y} \neq \boldsymbol{0}$.

instead of the augmented objective $F(\cdot; \bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_M)$ in (5) leads to the following OS-SPS update:

$$\bar{\boldsymbol{x}}^{\text{new}}_{(m \bmod M)+1} = \mathcal{P}_{\mathcal{X}}\left(\bar{\boldsymbol{x}}_m + \left[\breve{\boldsymbol{C}}_m(\bar{\boldsymbol{x}}_m)\right]^{-1} \nabla \Phi_m(\bar{\boldsymbol{x}}_m)\right) \tag{19}$$

for $m = 1, \ldots, M$. This greedy approach does not ensure monotonicity in the augmented objective F nor the PL objective Φ, so we need not insist that the curvatures satisfy the minorization conditions. A natural choice for $c_i(\cdot)$ is the Newton's curvature $-\ddot{h}_i(\cdot)$; this can be approximated as follows:

$$-\ddot{h}_i(l) \approx c_i^{\text{PC}} \triangleq -\ddot{h}_i\left(\arg\max_{\tilde{l} \geq 0} h_i(\tilde{l})\right)$$

$$= \begin{cases} \dfrac{(y_i - r_i)^2}{y_i}, & y_i > r_i \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

This choice is called "precomputed curvature (PC)" [5], [24]. For OS-SPS, the following subset-independent preconditioning matrix using PC is usually used in place of $\breve{\boldsymbol{C}}_m(\bar{\boldsymbol{x}}_m)$ in (19):

$$\breve{\boldsymbol{C}}^{\text{PC}}(\boldsymbol{x}) = \text{diag}_j\left\{\breve{c}_j^{\text{PC}}(\boldsymbol{x})\right\}$$

$$\breve{c}_j^{\text{PC}}(\boldsymbol{x}) = \max\left\{\frac{1}{M}\sum_{i=1}^N a_{ij} a_i c_i^{\text{PC}} + \frac{2\beta}{M}\sum_{k \in \mathcal{N}_j} w_{jk}\omega_\psi(x_j - x_k), \epsilon\right\}. \tag{21}$$

The first term on the right side in (21) can be precomputed and stored like the maximum curvatures (MC). The benefit of using PC is that it leads to faster convergence rates than MC since $c_i^{\text{MC}} \geq c_i^{\text{PC}}$. The update for OS-SPS is shown in (T-4) in Table II.

The OS-SPS shows very fast initial convergence rates but becomes eventually stuck at a limit cycle. Using more subsets leads to a faster initial convergence rate but causes the points in the limit cycle to be farther from the optimal solution.

It is worth noting that, for each update, OS-SPS uses the gradient and curvature for only one subobjective function at the previous subiterate in (19) whereas TRIOT uses the gradients and curvatures for all subobjective functions at previous M subiterates in (18). When the number of subobjective functions is $M = 1$, then both OS-SPS and TRIOT reduce to SPS.

D. Acceleration

TRIOT-OC/MC is usually slower than the nonconvergent OS-SPS. 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 the OS-SPS algorithm nearly gets to a limit cycle; even one single subiteration of TRIOT moves the iterate from the limit cycle to some point very close to the optimal solution. The reason 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 the first term on the right side in (18) or (T-5)].

To obtain further insight into this property, consider a simple unconstrained quadratic problem where the objective function and the subobjective functions are

$$\Phi(\boldsymbol{x}) = -\frac{1}{2}\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x}, \quad \Phi_m(\boldsymbol{x}) = -\frac{1}{2}\boldsymbol{x}'\boldsymbol{Q}_m\boldsymbol{x} + \boldsymbol{b}_m'\boldsymbol{x}$$

for $m = 1, \cdots, M$ where $\sum_{m=1}^M \boldsymbol{Q}_m = \boldsymbol{Q}$ and $\sum_{m=1}^M \boldsymbol{b}_m = \boldsymbol{b}$. Assume that each surrogate function $\phi_m(\boldsymbol{x}; \bar{\boldsymbol{x}})$ is equal to its corresponding subobjective $\Phi_m(\boldsymbol{x})$ so it has a closed-form maximizer $\hat{\boldsymbol{x}}_m = \boldsymbol{Q}_m^{-1}\boldsymbol{b}_m$ where we assume each $\boldsymbol{Q}_m$ is invertible. Then the OS approach will generate a limit cycle that consists of those $\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_M$. Now applying just one iteration of the incremental optimization transfer method as in (18) leads to

$$\hat{\boldsymbol{x}} = \left(\sum_{m=1}^M \boldsymbol{Q}_m\right)^{-1} \sum_{m=1}^M \boldsymbol{Q}_m\hat{\boldsymbol{x}}_m \;=\; \boldsymbol{Q}^{-1}\sum_{m=1}^M \boldsymbol{Q}_m(\boldsymbol{Q}_m^{-1}\boldsymbol{b}_m)$$
$$=\; \boldsymbol{Q}^{-1}\boldsymbol{b},$$

which is the maximizer of the original objective Φ [the second term on the right side in (18) equals zero]. This example suggests that the built-in averaging operation in TRIOT helps iterates escape from a limit cycle, generated by nonconvergent OS algorithms, toward the optimal solution.

However, in the early iterations, when OS-SPS is still far from the limit cycle and is making progress toward the optimal point, TRIOT is usually slower than OS-SPS due to the averaging of the past subiterates because the incremental optimization transfer approach updates the surrogates incrementally, that is, conservatively to ensure monotonicity. So it is desirable to get to a limit cycle quickly using OS-SPS with *many* subsets and then switch to TRIOT. In a 2D reconstruction case in Section IV, the use of 64 subsets is sufficient to reach a limit cycle within a couple of iterations.

2) Precomputed Curvatures: Forgoing monotonicity (in the augmented objective) and accordingly provable convergence, one can use for TRIOT the "precomputed curvatures (PC)" in (20). TRIOT-PC is faster than provably convergent TRIOT-OC/MC. 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

A. Algorithms Performance Evaluation: Real Data Study

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 EXACT 921 PET scanner with rotating rod transmission sources [67]. 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 128×128 with 4.2 mm pixels. The system geometry was approximated with 3.375 mm wide strip integrals and 3.375 mm ray spacing; the system matrix was generated using ASPIRE [68]. The total counts amounted to 9.2×10^5. We used the edge-preserving nonquadratic penalty (12) with $\delta = 4 \times 10^{-4}$ mm^{-1} and $\beta = 2^{18.5}$, chosen by visual inspection. A uniform image was used as a starting image. The results obtained by using a FBP reconstruction as a starting image were similar and are not shown here. We set U in (11) to be 7 mm^{-1}, which was much higher than reconstructed values (< 0.02 mm^{-1}), and the upper bound was never active.

Images were reconstructed using SPS-MC/PC, OS-SPS, 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 (16), which requires an extra backprojection per iteration, was between those with MC and PC (see [52, pp. 85–86] for the results with OC).

Fig. 1 shows normalized Φ difference versus iteration number for different algorithms using 16 subsets. The normalized Φ difference is defined as $(\Phi(\hat{x}^{\mathrm{PL}}) - \Phi(\hat{x}^n))/(\Phi(\hat{x}^{\mathrm{PL}}) - \Phi(\hat{x}^0))$ where $\hat{x}^{\mathrm{PL}}$ is a maximizer of the PL objective and $\hat{x}^0$ is the initial uniform image; a small value of the normalized difference means the image is closer to the optimal image $\hat{x}^{\mathrm{PL}}$. The optimal image $\hat{x}^{\mathrm{PL}}$ [shown in Fig. 4(b)] was estimated by 30 iterations of OS-SPS-16 (where "16" means the number of subsets) followed by 800 iterations of the SPS-OC algorithm that is monotonic and convergent (to a stationary point). For TRIOT, as described in Section III-B, we ran one iteration of OS-SPS, and then performed TRIOT iterations. Only for the last iteration of OS-SPS (in this case, iteration 1), we performed an additional TRIOT update in (T-6) in Table II that takes negligible computation. To see the effects of "this" TRIOT update, we plotted normalized Φ difference values both before and after (T-6) at iteration

1, leading to the vertical lines at iteration 1 for TRIOT algorithms in Fig. 1. For TRIOT, the jump in the unfavorable direction is due to the fact that built-in averaging slows down the algorithm when a limit cycle has not yet been reached as discussed in Section III-D. For a fair comparison, the SPS algorithms were performed after one iteration of OS-SPS was run. That is, for all methods shown in Fig. 1, $\hat{x}^1$ denotes the reconstructed image obtained using one iteration of OS-SPS. Although OS-SPS showed a fast initial convergence rate, it became stuck 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. However, TRIOT-MC and TRIOT-PC eventually outrun SPS-MC and SPS-PC, respectively. Although global convergence is not provably ensured for TRIOT-PC, the limit of TRIOT-PC (say, obtained by 1000 iterations) was the same as that of SPS-OC (obtained similarly) within numerical precision, which suggests TRIOT-PC has desirable convergence properties.

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 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, (SPS-PC outperforms TRIOT-MC from iteration 16 on.) This shows that it is effective to switch from OS-SPS to TRIOT, as described in Section III-D, when OS-SPS almost reaches a limit cycle. However, it is still inconvenient to predict how many iterations are required for OS-SPS to arrive at a limit cycle.

Fig. 3 shows normalized Φ 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, for the TRIOT algorithms, a significant improvement was obtained at iteration 2 when OS-SPS was switched to TRIOT, and the TRIOT algorithms outperformed the SPS algorithms. The switching point (iteration 2) was determined by some experimentation. In light of the effectiveness of the built-in averaging in TRIOT, to make SPS a stronger competitor, prior to switching to SPS (at iteration 2), we averaged the 64 previous subiterates that approximately comprise the limit cycle. As shown in Fig. 3, this averaging yielded improvements for SPS algorithms. However, convergence rates of TRIOT were still faster than those of SPS with such averaging. A plot of the normalized distance from the optimal image, $\|\hat{x}^n - \hat{x}^{\mathrm{PL}}\|/\|\hat{x}^{\mathrm{PL}}\|$, versus iteration number showed a similar trend (not shown here).

Fig. 4(c) shows the image to which OS-SPS with 64 subsets converged. It represents one point of the limit cycle generated by the OS-SPS, and looks visually different from the true PL optimal image $\hat{x}^{\mathrm{PC}}$ in Fig. 4(b). In contrast, 18 iterations of TRIOT-PC initialized by 2 iterations of OS-SPS yielded the

image in Fig. 4(d) which is indistinguishable from the optimal image in Fig. 4(b). TRIOT-MC and -OC (18 iterations with initial 2 iterations of OS-SPS) in Figs. 4(e) and (f) also yielded images very similar to the optimal image. The normalized distance $\|\hat{x}^{\mathrm{PL}} - \cdot\|/\|\hat{x}^{\mathrm{PL}}\|$ from the optimal image amounted to 1.1% for TRIOT-PC, 1.5% for TRIOT-OC, 2.0% for TRIOT-MC, 6.9% for OS-SPS, and 17.9% for FBP.

B. Convergence Really Matters?: Simulation Study

Generally, ordered subsets or incremental algorithms converge to a limit cycle that consists of sub-optimal solutions. In this paper we focus on developing incremental algorithms that converge to a true optimal solution. To examine how "suboptimal" the limit cycle solutions are, we compare nonconvergent OS algorithms (OS-SPS) and convergent TRIOT through a simulation study for 2D attenuation map reconstruction.

We used the same scanner geometry as in the previous subsection. The synthetic digital phantom shown in Fig. 5 was used; its warm background, left cold disc, and right hot disc had 0.004mm^{-1}, 0.009mm^{-1}, and 0.001mm^{-1}. The total counts amounted to 1M, and the known background contribution r_i corresponded to a uniform field of 10%. We used the penalty in (12) with $\delta = 5 \times 10^{-4}$ mm^{-1} and $\beta = 2^{17.5}$.

The aim of the simulation is to compare the bias and variance of the suboptimal images obtained by nonconvergent OS-SPS algorithms and the (nearly) optimal image obtained by TRIOT. We generated 400 realizations of pseudorandom transmission measurements according to (8). For each realization, we obtained a limit cycle image for OS-SPS-8 and OS-SPS-16 by running 50 iterations, and for OS-SPS-32 and OS-SPS-64 by running 20 iterations (note that it takes more iterations for OS algorithms with fewer subsets to reach a limit cycle); we also obtained a (nearly) optimal image by 18 iterations of TRIOT-PC with 64 subsets after 2 iterations of OS-SPS-64 as in the previous subsection.

A limit cycle consists of as many images as the number of subsets. We took only one image, say the first one, from the limit cycle for each OS method and computed the sample mean and variance over 400 realizations. Taking another image (*e.g.*, the second one) from the limit cycle led to similar results. Fig. 6 shows the profiles through the sample mean images for different methods; all images were nearly free of systematic biases. For each pixel, we computed the ratios of the sample standard deviation of OS methods to the sample standard deviation of TRIOT, and Fig. 7 shows the histogram of the ratios. Overall, OS-SPS-8 showed almost the same level of standard deviation as TRIOT; and OS-SPS-16, -32, and -64 yielded larger standard deviation than TRIOT by about 5%, 20%, and 58%, respectively. As the number of subsets decreases, the variance decreases, but the (initial) convergence rate also decreases and

it takes longer to converge to a limit cycle. We also tried averaging the limit cycle images; the resulting sample variance decreased but they were still larger than or equal to that of TRIOT depending on the number of subsets.

This example illustrates the importance of convergence for regularized methods. To summarize, as the number of subsets increases, a nonconvergent OS algorithm leads to higher variance although it does not increase bias significantly; therefore, it gives lower contrast-to-noise ratios. The limit cycle behavior introduces additional noise in reconstructed images.

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, and they are globally 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. The switching idea is also found in [69]. But we need to determine when to switch from OS-SPS to TRIOT as we must determine the relaxation parameters for relaxed OS algorithms such as RAMLA, BSREM, and relaxed OS-SPS. However, from our experience, it seems more convenient to determine when to switch than to choose relaxation parameters. In fact, when reasonably many subsets were used, as few as a couple of iterations of OS-SPS were sufficient to get close to a limit cycle for our 2D reconstruction. This switching strategy seems robust since we obtained similar results from a 2D simulation study using a different digital phantom [52]. Recently, Li *et al.* have proposed a method to determine automatically the switching point by fitting an exponential function to (sub)objective function values computed at previous iterates and by determining how close to a limit cycle a current iterate is [46].

TRIOT may be preferable to reducing the number of subsets with iteration since the consistent data flow in OS-SPS and TRIOT could be beneficial and it would be inconvenient to determine an optimal schedule for reducing the number of subsets.

One iteration of TRIOT-MC/PC or OS-SPS requires computing one projection and one backprojection plus the penalty related gradients and curvatures (the use of OC needs an extra backprojection); so the computational cost is almost the same as classic "emission" ML-EM except for the contribution of the penalty part. As the number of subsets increases, computation per iteration also increases due to the penalty part being updated for each subiteration. Although the computational contribution of the penalty

function is usually small compared to projection/backprojection, further investigation could help reduce this computation, *e.g.*, by subsetizing the penalty part.

Although the TRIOT-PC was numerically found to be convergent, if one really wants provable convergence, one could switch to TRIOT-MC or OC at some point.

One of the limitations of TRIOT is that for each update it uses previous M subiterates (where M is the number of subsets), so it requires more memory than ordinary OS algorithms by a factor of M. Therefore, the number of subsets is practically limited by memory availability particularly for large-sized problems. It will be an interesting challenging problem to resolve the memory issues while keeping TRIOT's global convergence and convergence rates.

APPENDIX A

DERIVATION OF INCREMENTAL EM ALGORITHMS

In this appendix we show that the incremental EM algorithms are a special case of the incremental optimization transfer algorithms.

For maximum likelihood (ML) estimation, one must maximize a log-likelihood function

$$\Phi(\boldsymbol{x}) = \log f(\boldsymbol{y}; \boldsymbol{x})$$

with respect to parameter $\boldsymbol{x} \in \mathbb{R}^p$ over a feasible set $\mathcal{X} \subset \mathbb{R}^p$ where $\boldsymbol{y} \in \mathbb{R}^N$ denotes a realization of an observable random vector $\boldsymbol{Y}$ with probability distribution $f(\boldsymbol{y}; \boldsymbol{x}^{\text{true}})$, and $\boldsymbol{x}^{\text{true}} \in \mathbb{R}^p$ is the true value of the unknown parameter. Assume that we identify an admissible complete-data[10] random vector $\boldsymbol{Z}$ for $f(\boldsymbol{y}; \boldsymbol{x})$. Then the following EM surrogate function satisfies the minorization conditions in (3) [30]:

$$\phi(\boldsymbol{x}; \bar{\boldsymbol{x}}) \overset{\triangle}{=} E[\log f(\boldsymbol{Z}; \boldsymbol{x})|\boldsymbol{Y} = \boldsymbol{y}; \bar{\boldsymbol{x}}] \tag{22}$$

for all $\bar{\boldsymbol{x}} \in \mathcal{X}$. But in many applications including imaging problems, the observed data is independent so the log-likelihood objective is additive-separable, that is,

$$\Phi(\boldsymbol{x}) = \sum_{m=1}^{M} \Phi_m(\boldsymbol{x}), \qquad \Phi_m(\boldsymbol{x}) = \log f(\boldsymbol{y}_m; \boldsymbol{x}),$$

and the complete data is conditionally independent, so for each $\Phi_m(\boldsymbol{x})$, one can obtain the following EM surrogate:

$$\phi_m(\boldsymbol{x}; \bar{\boldsymbol{x}}) = E[\log f(\boldsymbol{Z}_m; \boldsymbol{x})|\boldsymbol{Y}_m = \boldsymbol{y}_m; \bar{\boldsymbol{x}}], \tag{23}$$

---

[10]A random vector $\boldsymbol{Z}$ with probability distribution $f(\boldsymbol{z}; \boldsymbol{x})$ is called an admissible complete-data vector for $f(\boldsymbol{y}; \boldsymbol{x})$ if $f(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{x}) = f(\boldsymbol{y}|\boldsymbol{z})f(\boldsymbol{z}; \boldsymbol{x})$ [37], [38]. A special case is that $\boldsymbol{Y}$ is a deterministic function of $\boldsymbol{Z}$.

which also satisfies the minorization conditions in (3) where $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_M)$ and $\boldsymbol{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M)$ are some decompositions of the incomplete data and the complete data, respectively. Defining the augmented objective function as in (5) and then alternating between updating $\boldsymbol{x}$ and one of the $\bar{\boldsymbol{x}}_m$'s as in Table I leads to the incremental EM algorithms [3], [28]. The COSEM algorithm [2], [27], a special case of the incremental EM for emission tomography, can be readily derived.

APPENDIX B
GLOBAL CONVERGENCE PROOF

In this appendix we prove the convergence of the incremental optimization transfer algorithm given in Table I. Define $\boldsymbol{z} \overset{\triangle}{=} (\boldsymbol{x}; \bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_M) \in \mathcal{X}^{M+1}$, and define a map $\mathcal{M} : \mathcal{X}^{M+1} \to \mathcal{X}^{M+1}$ such that $\boldsymbol{z}^{n+1} = \mathcal{M}(\boldsymbol{z}^n)$ where $\boldsymbol{z}^{n+1} = (\boldsymbol{x}^{n+1}; \bar{\boldsymbol{x}}_1^{n+1}, \ldots, \bar{\boldsymbol{x}}_M^{n+1})$ is computed by (T-1)–(T-3) for $\boldsymbol{z}^n = (\boldsymbol{x}^n; \bar{\boldsymbol{x}}_1^n, \ldots, \bar{\boldsymbol{x}}_M^n)$. Suppose that the algorithm generates a sequence $\{\boldsymbol{z}^n\}$ (or a sequence $\{\boldsymbol{x}^n\}$ by taking the first component of $\boldsymbol{z}^n$), given some initial point $\boldsymbol{z}^0 \in \mathcal{X}^{M+1}$. Define an augmented solution set as follows:

$$\Lambda \overset{\triangle}{=} \{\boldsymbol{z} = (\boldsymbol{x}; \boldsymbol{x}, \ldots, \boldsymbol{x}) \in \mathcal{X}^{M+1} : \boldsymbol{x} \in \Gamma\} \tag{24}$$

where Γ is defined in (7). We impose the following assumptions.

Assumption 1: Each $\Phi_m(\boldsymbol{x})$ and $\phi_m(\boldsymbol{x}; \bar{\boldsymbol{x}})$ is differentiable with respect to $\boldsymbol{x}$ on a nonempty, closed, and convex set $\mathcal{X} \subset \mathbb{R}^p$ for all $\bar{\boldsymbol{x}} \in \mathcal{X}$, and each $\phi_m(\boldsymbol{x}; \bar{\boldsymbol{x}})$ is continuous with respect to $\bar{\boldsymbol{x}}$ on $\mathcal{X}$ for all $\boldsymbol{x} \in \mathcal{X}$.

Assumption 2: The solution set Γ of stationary points defined in (7) is non-empty, and therefore so is the augmented solution set Λ defined in (24).

Assumption 3: The iterates $\{\boldsymbol{z}^n\}$ are bounded where $\boldsymbol{z}^n = (\boldsymbol{x}^n; \bar{\boldsymbol{x}}_1^n, \ldots, \bar{\boldsymbol{x}}_M^n)$.

Assumption 3 is ensured by either of the following sufficient conditions.

Assumption 3′: A level set defined by $\{\boldsymbol{z} \in \mathcal{X}^{M+1} : F(\boldsymbol{z}) \geq F(\boldsymbol{z}^0)\}$ is bounded where F is the augmented objective function defined in (5).

Assumption 3″: The feasible set $\mathcal{X}$ is bounded.

Note that Assumptions 1 and 3″ imply Assumption 2 by the Weierstrass' Theorem [42, p. 654]. We assume that the surrogates ϕ_m satisfy the following conditions.

Condition 1: The functionals ϕ_m satisfy the minorization conditions in (3).

Condition 2: The following derivatives match for all m and $\boldsymbol{x} \in \mathcal{X}$:

$$\nabla \Phi_m(\boldsymbol{x}) = \nabla^{10} \phi_m(\boldsymbol{x}; \boldsymbol{x}) \tag{25}$$

where ∇^{10} is the column gradient operator with respect to the first argument[11] (see [39] for less restrictive conditions).

Condition 3: There exists a *unique* maximizer in (T-1).

The following is sufficient for Condition 3.

Condition 3': Each $\phi_m(\cdot; \bar{\boldsymbol{x}}_m)$ is strictly concave for all $\bar{\boldsymbol{x}}_m \in \mathcal{X}$, and there exists a maximizer of $F(\cdot; \bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_M)$ over $\mathcal{X}$ for all $\bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_M \in \mathcal{X}$.

Condition 4: If A is a bounded subset of $\mathcal{X}^M$, then $\{\arg\max_{\boldsymbol{x} \in \mathcal{X}} F(\boldsymbol{x}; \bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_M) : (\bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_M) \in A\}$ is also bounded.

Note that Assumption $3''$ implies Condition 4. Using the above assumptions and conditions, we prove a series of lemmas necessary for proving convergence.

Lemma 1: The map $\mathcal{M}$ such that $\boldsymbol{z}^{n+1} = \mathcal{M}(\boldsymbol{z}^n)$ is continuous.

Proof: The map $\mathcal{M}$ is the composition of $2M$ maps: $\mathcal{M} = \mathcal{S}^M \mathcal{T} \mathcal{S}^{M-1} \mathcal{T} \cdots \mathcal{S}^1 \mathcal{T}$ where $\mathcal{S}^m(\boldsymbol{x}; \bar{\boldsymbol{x}}_1, \cdots, \bar{\boldsymbol{x}}_M) \triangleq (\boldsymbol{x}; \bar{\boldsymbol{x}}_1, \cdots, \bar{\boldsymbol{x}}_{m-1}, \boldsymbol{x}, \bar{\boldsymbol{x}}_{m+1}, \cdots, \bar{\boldsymbol{x}}_M)$, and $\mathcal{T}(\boldsymbol{x}; \bar{\boldsymbol{x}}_1, \cdots, \bar{\boldsymbol{x}}_M) \triangleq (\boldsymbol{w}; \bar{\boldsymbol{x}}_1, \cdots, \bar{\boldsymbol{x}}_M)$ with $\boldsymbol{w} = \arg\max_{\boldsymbol{x} \in \mathcal{X}} F(\boldsymbol{x}; \bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_M)$. Since each $\mathcal{S}^m$ is continuous, it is only necessary to show that $\mathcal{T}$ is continuous [47, p. 187]. Let $\{\bar{\boldsymbol{x}}_m^k\}$ be a sequence from $\mathcal{X}$ with $\bar{\boldsymbol{x}}_m^k \to \bar{\boldsymbol{x}}_m^*$ for all m. Suppose that $\boldsymbol{w}^k = \arg\max_{\boldsymbol{x} \in \mathcal{X}} F(\boldsymbol{x}; \bar{\boldsymbol{x}}_1^k, \ldots, \bar{\boldsymbol{x}}_M^k)$ and that $\boldsymbol{w}^* = \arg\max_{\boldsymbol{x} \in \mathcal{X}} F(\boldsymbol{x}; \bar{\boldsymbol{x}}_1^*, \ldots, \bar{\boldsymbol{x}}_M^*)$. It suffices to show that $\boldsymbol{w}^k \to \boldsymbol{w}^*$. Suppose that $\{\boldsymbol{w}^k\}$ does not converge to $\boldsymbol{w}^*$. Then there exists a subsequence $\{\boldsymbol{w}^{k_l}\}$ of $\{\boldsymbol{w}^k\}$ such that $\boldsymbol{w}^{k_l} \to \boldsymbol{w}^{**}$ for some $\boldsymbol{w}^{**} \neq \boldsymbol{w}^*$ by Condition 4. By the definition of $\boldsymbol{w}^{k_l}$, $F(\boldsymbol{w}^{k_l}; \bar{\boldsymbol{x}}_1^{k_l}, \cdots, \bar{\boldsymbol{x}}_M^{k_l}) \geq F(\boldsymbol{w}^*; \bar{\boldsymbol{x}}_1^{k_l}, \cdots, \bar{\boldsymbol{x}}_M^{k_l})$. By letting $l \to \infty$, we have $F(\boldsymbol{w}^{**}; \bar{\boldsymbol{x}}_1^*, \cdots, \bar{\boldsymbol{x}}_M^*) \geq F(\boldsymbol{w}^*; \bar{\boldsymbol{x}}_1^*, \cdots, \bar{\boldsymbol{x}}_M^*)$ since F is continuous by Assumption 1. But this is a contradiction in view of the definition of $\boldsymbol{w}^*$ and Condition 3. Therefore, it must be the case that $\boldsymbol{w}^k \to \boldsymbol{w}^*$, and the conclusion follows. ∎

Lemma 2: The iterates $\{\boldsymbol{z}^n\}$ generated by (T-1)–(T-3) yield monotonic increases in F, that is, $F(\boldsymbol{z}^{n+1}) \geq F(\boldsymbol{z}^n)$ for all n.

Proof: It follows from the cyclic block coordinate ascent updates in (T-1) and (T-2). ∎

Lemma 3: Suppose that $\boldsymbol{z}^* \in \mathcal{X}^{M+1}$ is a fixed point of $\mathcal{M}$, that is, $\mathcal{M}(\boldsymbol{z}^*) = \boldsymbol{z}^*$. Then $\boldsymbol{z}^* \in \Lambda$ where Λ is defined in (24).

Proof: For the fixed point $\boldsymbol{z}^* = (\boldsymbol{x}^*; \bar{\boldsymbol{x}}_1^*, \ldots, \bar{\boldsymbol{x}}_M^*)$, in view of Condition 3, one can show that $\boldsymbol{x}^* = \bar{\boldsymbol{x}}_1^* = \ldots = \bar{\boldsymbol{x}}_M^*$. Since $\boldsymbol{x}^*$ is a maximizer of $\sum_{m=1}^{M} \phi_m(\cdot; \boldsymbol{x}^*)$ over $\mathcal{X}$, it follows that $\sum_{m=1}^{M} \nabla^{10} \phi_m(\boldsymbol{x}^*; \boldsymbol{x}^*)'(\boldsymbol{x} - \boldsymbol{x}^*) \leq 0$ for all $\boldsymbol{x} \in \mathcal{X}$ [42, p. 194]. Therefore, by Condition 2, $\nabla\Phi(\boldsymbol{x}^*)'(\boldsymbol{x} - \boldsymbol{x}^*) \leq 0$ for all

---

[11]If $\boldsymbol{x}$ is an interior point of $\mathcal{X}$, Condition 2 is implied by Condition 1 [52, Lemma 3.3].

$x \in \mathcal{X}$, and it follows that $x^* \in \Gamma$. ∎

Lemma 4: If $z \notin \Lambda$, then $F(\mathcal{M}(z)) > F(z)$.

Proof: If $z \notin \Lambda$, then z is not a fixed point of $\mathcal{M}$ by Lemma 3. Combining Condition 3 and Lemma 2 leads to the conclusion. ∎

Now we prove the following theorem on the convergence of the incremental optimization transfer algorithm.

Theorem 1: Suppose that $\{z^n\}$ is a sequence generated by (T-1)–(T-3) with $z^0 \in \mathcal{X}^{M+1}$ and that Assumptions 1–3 and Conditions 1–4 hold. Then any limit point of $\{z^n\}$ is an element of Λ.

Proof: The conclusion follows from the Zangwill's Convergence Theorem [70, p. 91] with Assumption 3; and Lemmas 1, 2 and 4. ∎

The following corollaries and lemmas also hold when "$\{x^n\}$" is replaced with "$\{\bar{x}_m^n\}$" for all m.

Corollary 1: Suppose $\{x^n\}$ is a sequence obtained by taking the first component from z^n in Theorem 1. Then any limit point of $\{x^n\}$ is an element of Γ.

Proof: Use Theorem 1, Assumption 3, and the definition of Λ in (24). ∎

Corollary 2: If Φ is concave, then any limit point of $\{x^n\}$ is a global maximizer of Φ over $\mathcal{X}$. Moreover, if Φ is strictly concave, then $\{x^n\}$ converges to *the* global maximizer of Φ over $\mathcal{X}$.

Proof: Use Corollary 1 and [42, Proposition 2.1.2]. ∎

When Φ is not strictly concave, there is no guarantee that the algorithm converges to a limit. However, convergence can be established by additionally assuming that the solution set Γ is discrete.

Lemma 5: Suppose $\{x^n\}$ is a sequence from Corollary 1. Then $\left\| x^{n+1} - x^n \right\| \to 0$.

Proof: It follows from [71, Theorem 3.1] that $\|z^{n+1} - z^n\| \to 0$. Since $\|z^{n+1} - z^n\|^2 = \|x^{n+1} - x^n\|^2 + \sum_{m=1}^M \|\bar{x}_m^{n+1} - \bar{x}_m^n\|^2$, it must be the case that $\left\| x^{n+1} - x^n \right\| \to 0$. ∎

Lemma 6: Suppose $\{x^n\}$ is a sequence from Corollary 1. Additionally, suppose that the set Γ is discrete. Then $\{x^n\}$ converges to an element in Γ.

Proof: Let S be a set of limit points of $\{x^n\}$. Then $S \subset \Gamma$ by Theorem 1. But, by Lemma 5, S is connected [72, p. 173]. Since S is both discrete and connected, it is a singleton. ∎

The above lemma implies that if stationary points of (2) are isolated, then the algorithm converges to one of them.

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] 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.

[3] 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., pp. 255–268. Kluwer, Dordrencht, 1998.

[4] 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.

[5] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–2851, Nov. 1999.

[6] 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.

[7] J. A. Browne and A. R. De Pierro, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography," *IEEE Trans. Med. Imag.*, vol. 15, no. 5, pp. 687–699, Oct. 1996.

[8] C. L. Byrne, "Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative methods," *IEEE Trans. Image Processing*, vol. 7, no. 1, pp. 100–109, Jan. 1998.

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

[10] H. Kudo, H. Nakazawa, and T. Saito, "Block-gradient method for image reconstruction in emission tomography," *Trans. IEICE*, vol. J83-D-II, no. 1, pp. 63–73, Jan. 2000, In Japanese.

[11] A. R. De Pierro and M. E. B. Yamagishi, "Fast EM-like methods for maximum 'a posteriori' estimates in emission tomography," *IEEE Trans. Med. Imag.*, vol. 20, no. 4, pp. 280–288, Apr. 2001.

[12] Q. Li, E. Asma, and R. M. Leahy, "A fast fully 4D incremental gradient reconstruction algorithm for list mode PET data," in *Proc. IEEE Intl. Symp. Biomedical Imaging*, 2004, pp. 555–558.

[13] 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.

[14] V. M. Kibardin, "Decomposition into functions in the minimization problem," *Automat. Remote Control*, vol. 40, pp. 1311–1323, 1980, Translation of *Avtomatika i Telemekhanika*, vol. 9, pp. 66–79, Sept. 1979.

[15] A. Nedić and D. P. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM J. Optim.*, vol. 12, no. 1, pp. 109–138, 2001.

[16] A. Nedić 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.

[17] K. C. Kiwiel, "Convergence of approximate and incremental subgradient methods for convex optimization," *SIAM J. Optim.*, vol. 14, no. 3, pp. 807–840, 2004.

[18] Y. Censor, "Row-action methods for huge and sparse systems and their applications," *SIAM Review*, vol. 23, no. 4, pp. 444–466, Oct. 1981.

[19] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for the three-dimensional electron microscopy and X-ray photography," *J. Theor. Biol.*, vol. 29, pp. 471–481, 1970.

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

[21] Y. Censor, P. P. B. Eggermont, and D. Gordon, "Strong underrelaxation in Kaczmarz's method for inconsistent systems," *Numerische Mathematik*, vol. 41, pp. 83–92, 1983.

[22] Y. Censor, D. Gordon, and R. Gordon, "Component averaging: An efficient iterative parallel algorithm for large and sparse unstructured problems," *Parallel Computing*, vol. 27, no. 6, pp. 777–808, May 2001.

[23] S. Sotthivirat and J. A. Fessler, "Relaxed ordered-subsets algorithm for penalized-likelihood image restoration," *J. Opt. Soc. Amer. A*, vol. 20, no. 3, pp. 439–449, Mar. 2003.

[24] H. Erdoğan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Trans. Med. Imag.*, vol. 18, no. 9, pp. 801–814, Sept. 1999.

[25] D. P. Bertsekas, "A new class of incremental gradient methods for least squares problems," *SIAM J. Optim.*, vol. 7, no. 4, pp. 913–926, Nov. 1997.

[26] A. Gunawardana and W. Byrne, "Convergence of EM variants," Tech. Rep. CLSP Research Note No. 32, ECE Dept., Johns Hopkins University, Feb. 1999.

[27] 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.

[28] A. J. R. Gunawardana, *The information geometry of EM variants for speech and image processing*, Ph.D. thesis, Johns Hopkins University, Baltimore, MD., 2001.

[29] D. Blatt, A. Hero, and H. Gauchman, "An incremental gradient method that converges with a constant step size," *SIAM J. Optim.*, 2004, Submitted.

[30] 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.

[31] D. R. Hunter and K. Lange, "Rejoinder to discussion of "Optimization transfer using surrogate objective functions"," *J. Computational and Graphical Stat.*, vol. 9, no. 1, pp. 53–59, Mar. 2000.

[32] 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.

[33] I. T. Hsiao, A. Rangarajan, P. Khurd, and G. Gindi, "An accelerated convergent ordered subsets algorithm for emission tomography," *Phys. Med. Biol.*, vol. 49, no. 11, pp. 2145–2156, June 2004.

[34] J. Qi, R. M. Leahy, S. R. Cherry, A. Chatziioannou, and T. H. Farquhar, "High resolution 3D Bayesian image reconstruction using the microPET small-animal scanner," *Phys. Med. Biol.*, vol. 43, no. 4, pp. 1001–14, Apr. 1998.

[35] R. M. Leahy and J. Qi, "Statistical approaches in quantitative positron emission tomography," *Statistics and Computing*, vol. 10, no. 2, pp. 147–165, Apr. 2000.

[36] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Sig. Proc.*, vol. 41, no. 2, pp. 534–548, Feb. 1993.

[37] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. Sig. Proc.*, vol. 42, no. 10, pp. 2664–2677, Oct. 1994.

[38] J. A. Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Trans. Image Processing*, vol. 4, no. 10, pp. 1417–1429, Oct. 1995.

[39] M. W. Jacobson and J. A. Fessler, "Properties of optimization transfer algorithms on convex feasible sets," *SIAM J. Optim.*, 2003, Submitted.

[40] J. Nocedal, "Large scale unconstrained optimization," in *The State of the Art in Numerical Analysis*, I. S. Duff and G. A. Watson, Eds., pp. 311–338. Clarendon Press, Oxford, 1997.

[41] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, Feb. 2004.

[42] D. P. Bertsekas, *Nonlinear programming*, Athena Scientific, Belmont, 2 edition, 1999.

[43] 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.

[44] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 89–99, Feb. 2002.

[45] S. Sotthivirat and J. A. Fessler, "Image recovery using partitioned-separable paraboloidal surrogate coordinate ascent algorithms," *IEEE Trans. Image Processing*, vol. 11, no. 3, pp. 306–17, Mar. 2002.

[46] Q. Li, S. Ahn, and R. M. Leahy, "Fast hybrid algorithms for PET image reconstruction," in *Proc. IEEE Nuclear Science Symp. Medical Imaging Conf.*, 2005, Submitted.

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

[48] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear programming: Theory and algorithms*, Wiley, New York, 1993.

[49] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, no. 2, pp. 113–122, Oct. 1982.

[50] F. Natterer and F. Wübbeling, *Mathematical methods in image reconstruction*, SIAM, Philadelphia, 2001.

[51] A. Rangarajan, P. Khurd, I.-T. Hsiao, and G. Gindi, "Convergence proofs for the COSEM-ML and COSEM-MAP algorithms," Tech. Rep. 03-1, Medical Imag. Proc. Lab., State Univ. of New York, Stony Brook, NY, Dec. 2003.

[52] S. Ahn, *Convergent algorithms for statistical image reconstruction in emission tomography*, Ph.D. thesis, Univ. of Michigan, Ann Arbor, MI., 2004.

[53] 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., pp. 1–70. SPIE, Bellingham, 2000.

[54] J. A. Fessler and H. Erdoğan, "A paraboloidal surrogates algorithm for convergent penalized-likelihood emission image reconstruction," in *Proc. IEEE Nuclear Science Symp. Medical Imaging Conf.*, 1998, vol. 2, pp. 1132–1135.

[55] S. Ahn and J. A. Fessler, "Emission image reconstruction for randoms-precorrected PET allowing negative sinogram values," *IEEE Trans. Med. Imag.*, vol. 23, no. 5, pp. 501–601, May 2004.

[56] S. Sotthivirat and J. A. Fessler, "Penalized-likelihood image reconstruction for digital holography," *J. Opt. Soc. Am. A*, vol. 21, no. 5, pp. 737–50, May 2004.

[57] P. J. Huber, *Robust statistics*, Wiley, New York, 1981.

[58] 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.

[59] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Trans. Image Processing*, vol. 5, no. 9, pp. 1346–1358, Sept. 1996.

[60] J. Qi and R. M. Leahy, "A theoretical study of the contrast recovery and variance of MAP reconstructions with applications to the selection of smoothing parameters," *IEEE Trans. Med. Imag.*, vol. 18, no. 4, pp. 293–305, Apr. 1999.

[61] J. W. Stayman and J. A. Fessler, "Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction," *IEEE Trans. Med. Imag.*, vol. 19, no. 6, pp. 601–615, June 2000.

[62] J. Qi and R. M. Leahy, "Resolution and noise properties MAP reconstruction for fully 3D PET," *IEEE Trans. Med. Imag.*, vol. 19, no. 5, pp. 493–506, May 2000.

[63] J. A. Fessler, "Analytical approach to regularization design for isotropic spatial resolution," in *Proc. IEEE Intl. Symp. Biomedical Imaging*, 2003, pp. 2202–2206.

[64] J. W. Stayman and J. A. Fessler, "Compensation for nonuniform resolution using penalized-likelihood reconstruction in space-variant imaging systems," *IEEE Tr. Med. Im.*, vol. 23, no. 3, pp. 269–284, Mar. 2004.

[65] A. F. Bermant, *A course of mathematical analysis part I*, Pergamon Press, Oxford, England, 1963.

[66] K. Lange and J. A. Fessler, "Globally convergent algorithms for maximum a posteriori transmission tomography," *IEEE Trans. Image Processing*, vol. 4, no. 10, pp. 1430–1438, Oct. 1995.

[67] K. Wienhard, L. Eriksson, S. Grootoonk, 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.

[68] J. A. Fessler, "ASPIRE 3.0 user's guide: A sparse iterative reconstruction library," Tech. Rep. 293, Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 48109-2122, July 1995, Available from `http://www.eecs.umich.edu/~fessler`.

[69] A. Rahmim, M. Lenox, A. J. Reader, C. Michel, Z. Burbar, T. J. Ruth, and V. Sossi, "Statistical list-mode image reconstruction for the high resolution research tomograph," *Phys. Med. Biol.*, vol. 49, pp. 4239–4258, Aug. 2004.

[70] W. Zangwill, *Nonlinear programming, a unified approach*, Prentice-Hall, NJ, 1969.

[71] R. R. Meyer, "Sufficient conditions for the convergence of monotonic mathematical programming algorithms," *J. Comput. System. Sci.*, vol. 12, no. 1, pp. 108–121, 1976.

[72] A. M. Ostrowski, *Solution of equations in Euclidean and Banach spaces*, Academic, New York, 3 edition, 1973.

TABLE I

Initialize $\boldsymbol{x}^0, \bar{\boldsymbol{x}}_1^0, \ldots, \bar{\boldsymbol{x}}_M^0 \in \mathcal{X}$

for $n = 0, \ldots, n_{\text{iter}} - 1$

    for $m = 1, \ldots, M$

$$\boldsymbol{x}^{\text{new}} \quad = \quad \arg\max_{\boldsymbol{x} \in \mathcal{X}} F\left(\boldsymbol{x}; \bar{\boldsymbol{x}}_1^{n+1}, \ldots, \bar{\boldsymbol{x}}_{m-1}^{n+1}, \bar{\boldsymbol{x}}_m^n, \bar{\boldsymbol{x}}_{m+1}^n, \ldots \bar{\boldsymbol{x}}_M^n\right) \tag{T-1}$$

$$\bar{\boldsymbol{x}}_m^{n+1} \quad = \quad \boldsymbol{x}^{\text{new}} = \arg\max_{\bar{\boldsymbol{x}}_m \in \mathcal{X}} F\left(\boldsymbol{x}^{\text{new}}; \bar{\boldsymbol{x}}_1^{n+1}, \ldots, \bar{\boldsymbol{x}}_{m-1}^{n+1}, \bar{\boldsymbol{x}}_m, \bar{\boldsymbol{x}}_{m+1}^n, \ldots \bar{\boldsymbol{x}}_M^n\right) \tag{T-2}$$

    end

$$\boldsymbol{x}^{n+1} = \bar{\boldsymbol{x}}_M^{n+1} \tag{T-3}$$

end

TABLE II

<small>OUTLINE FOR TRANSMISSION INCREMENTAL OPTIMIZATION TRANSFER (TRIOT) ALGORITHM USING MAXIMUM CURVATURE (MC).</small>

Initialize: $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}^0 = \left[\text{FBP} \left\{ \log \left(\frac{b_i}{y_i - r_i} \right) \right\}_{i=1}^N \right]^+$

Precompute: $d_{mj}^{\text{MC}} = \sum_{i \in S_m} a_{ij} a_i \left[\left(1 - \frac{y_i r_i}{(b_i + r_i)^2} \right) b_i \right]_+$ and $d_j^{\text{PC}} = \frac{1}{M} \sum_{i=1}^N a_{ij} a_i c_i^{\text{PC}}, \quad \forall m, \ j$

for each iteration $n = 1, \ldots, n_{\text{iter}}$

    for each subset (subiteration) $m = 1, \ldots, M$

$$\hat{l}_i = \sum_{j=1}^p a_{ij} \hat{x}_j, \quad \dot{h}_i = \left(1 - \frac{y_i}{b_i e^{-\hat{l}_i} + r_i} \right) b_i e^{-\hat{l}_i}, \quad \forall i \in S_m$$

$$\dot{L}_{mj} = \sum_{i \in S_m} a_{ij} \dot{h}_i, \quad r_{mj} = \frac{2\beta}{M} \sum_{k \in \mathcal{N}_j} w_{jk} \omega_\psi (\hat{x}_j - \hat{x}_k), \quad \forall j$$

$$\bar{x}_{mj} = \hat{x}_j, \quad \forall j$$

        if $n \leq n_{\text{iter}}^{\text{OS}}$, perform the following OS-SPS update:

$$\hat{x}_j = \left[\bar{x}_{mj} + \frac{\dot{L}_{mj} - \frac{\beta}{M} \sum_{k \in \mathcal{N}_j} w_{jk} \dot{\psi}(\bar{x}_{mj} - \bar{x}_{mk})}{\max \left\{ d_j^{\text{PC}} + r_{mj}, \ \epsilon \right\}} \right]^+, \quad \forall j \tag{T-4}$$

        else, perform the following TRIOT-MC update:

$$\hat{x}_j = \left[\frac{\sum_{l=1}^M \left[\bar{x}_{lj} \max\{d_{lj}^{MC} + r_{lj}, \ \epsilon\} + \left(\dot{L}_{lj} - \frac{\beta}{M} \sum_{k \in \mathcal{N}_j} w_{jk} \dot{\psi}(\bar{x}_{lj} - \bar{x}_{lk}) \right) \right]}{\sum_{l=1}^M \max\{d_{lj}^{MC} + r_{lj}, \ \epsilon\}} \right]^+, \quad \forall j \tag{T-5}$$

        end

    end

    if $n = n_{\text{iter}}^{\text{OS}}$ (the last iteration of OS-SPS),

$$\text{perform (T-5)} \tag{T-6}$$

    end

    $\hat{\boldsymbol{x}}^n = \hat{\boldsymbol{x}}$

end

Here ϵ is some small positive value; c_i^{PC} is defined in (20); and $[x]^+ \triangleq \text{median}\{0, x, U\}$, which should not be confused with $[x]_+ \triangleq \max\{x, 0\}$.
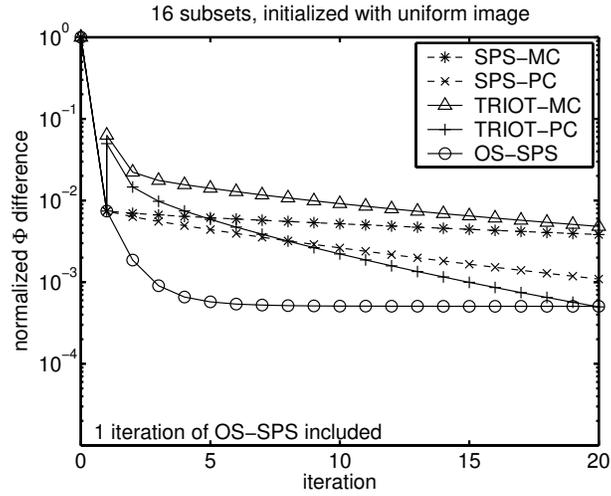
Fig. 1.   Comparison of non-OS algorithms (SPS-MC/PC), an OS algorithm (OS-SPS), and incremental optimization transfer algorithms (TRIOT-MC/PC) for 2D attenuation map reconstruction using real PET data. This figure shows $(\Phi(\hat{x}^{\mathrm{PL}}) - \Phi(\hat{x}^n))/(\Phi(\hat{x}^{\mathrm{PL}}) - \Phi(\hat{x}^0))$ versus iteration number where $\hat{x}^{\mathrm{PL}}$ 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. The starting image was a uniform image for all cases.
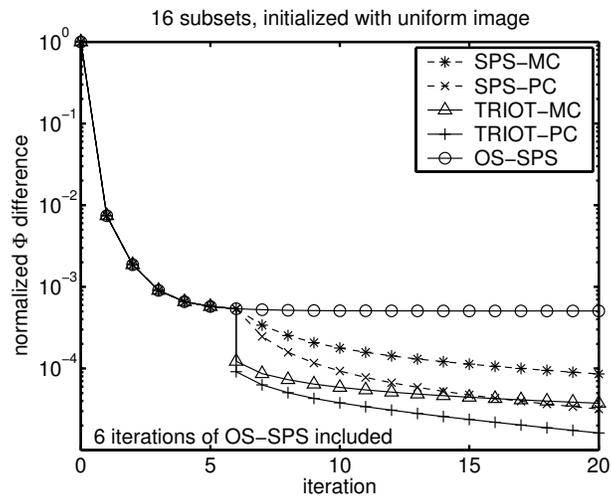
Fig. 2.   Same as Fig. 1, but six initial iterations of OS-SPS were included for TRIOT and SPS algorithms.
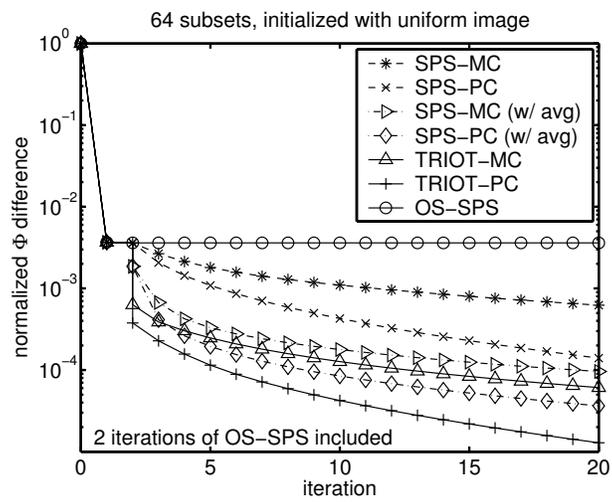
Fig. 3. Comparison of $(\Phi(\hat{\boldsymbol{x}}^{\mathrm{PL}}) - \Phi(\hat{\boldsymbol{x}}^{n}))/(\Phi(\hat{\boldsymbol{x}}^{\mathrm{PL}}) - \Phi(\hat{\boldsymbol{x}}^{0}))$ versus iteration number. For this figure, 64 subsets are used for OS-SPS and TRIOT algorithms, and two iterations of OS-SPS are included initially for TRIOT and SPS algorithms. This figure also shows the performance of SPS algorithms that include averaging 64 subiterates after 2 iterations of OS-SPS.
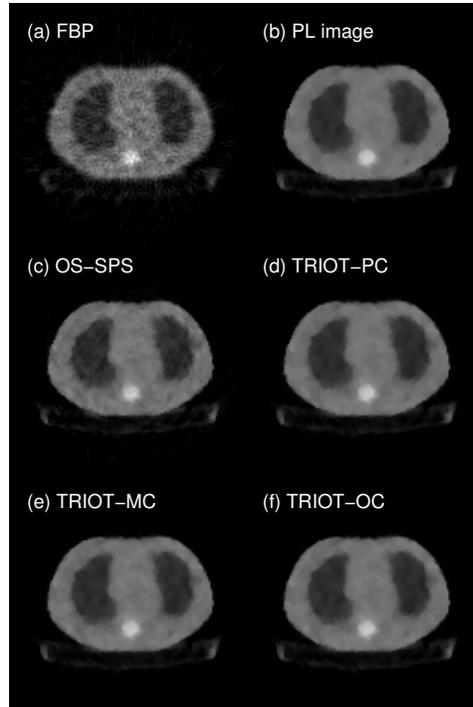
Fig. 4. Reconstructed attenuation maps. (a) FBP reconstruction. (b) PL estimate image $\hat{x}^{\mathrm{PL}}$ obtained using 30 iterations of OS-SPS with 16 subsets followed by 800 iterations of SPS-OC. (c) PL reconstruction using 20 iterations of OS-SPS with 64 subsets (an image that is one point of a limit cycle). (d) PL reconstruction using 2 iterations of OS-SPS and 18 iterations of TRIOT-PC with 64 subsets. (e) PL reconstruction using 2 iterations of OS-SPS and 18 iterations of TRIOT-MC with 64 subsets. (f) PL reconstruction using 2 iterations of OS-SPS and 18 iterations of TRIOT-OC with 64 subsets.
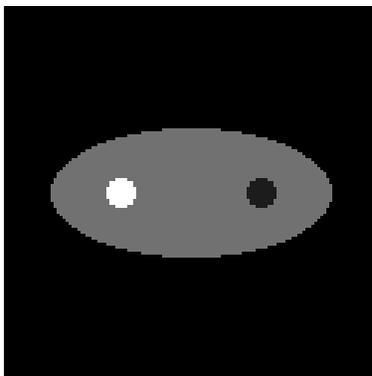
Fig. 5.   Digital phantom used in simulation. The background, left hot disc, and right cold disc have attenuation coefficients of 0.004mm^{-1}, 0.009mm^{-1}, and 0.001mm^{-1}, respectively.
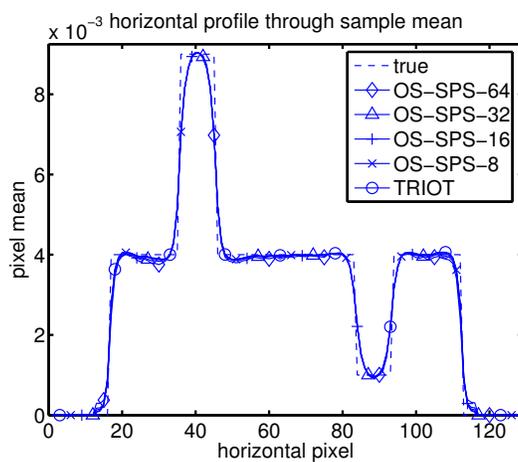

Fig. 6.   Horizontal profile 64 through sample mean of different ordered subsets methods.
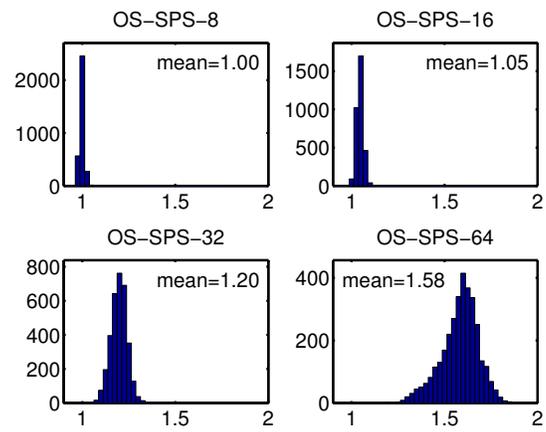
Fig. 7. Histogram of the ratio of sample standard deviation of different ordered subsets methods to sample standard deviation of TRIOT.