Relaxed ordered-subset algorithm for penalized-likelihood image restoration

Saowapak Sotthivirat and Jeffrey A. Fessler

Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan 48109

Received May 20, 2002; revised manuscript received October 14, 2002; accepted October 15, 2002

The expectation-maximization (EM) algorithm for maximum-likelihood image recovery is guaranteed to converge, but it converges slowly. Its ordered-subset version (OS-EM) is used widely in tomographic image reconstruction because of its order-of-magnitude acceleration compared with the EM algorithm, but it does not guarantee convergence. Recently the ordered-subset, separable-paraboloidal-surrogate (OS-SPS) algorithm with relaxation has been shown to converge to the optimal point while providing fast convergence. We adapt the relaxed OS-SPS algorithm to the problem of image restoration. Because data acquisition in image restoration is different from that in tomography, we employ a different strategy for choosing subsets, using pixel locations rather than projection angles. Simulation results show that the relaxed OS-SPS algorithm can provide an order-of-magnitude acceleration over the EM algorithm for image restoration. This new algorithm now provides the speed and guaranteed convergence necessary for efficient image restoration.

OCIS codes: 100.0100, 100.1830, 100.2000, 100.3020, 100.3190, 180.1790.

1. INTRODUCTION

Statistical techniques have been shown to improve image quality in image restoration. They can incorporate physical models of imaging systems, thus improving restoration. Moreover, object constraints, such as nonnegativity, can be enforced easily. Since closed-form solutions for maximum-likelihood (ML) or penalized-likelihood (PL) estimates are usually unobtainable, iterative algorithms are needed.[1–5] Fast-converging algorithms are desirable to quickly recover an approximation of the original image. However, existing algorithms still lack one or more desirable properties of an ideal algorithm, such as the guarantee of convergence, rapid convergence, efficient computation, and parallelizability.

Expectation-maximization (EM) algorithms[6,7] and their ordered-subset (OS) versions[8–10] are among the most commonly used algorithms; however, they have limitations of either speed or convergence. Although EM algorithms are guaranteed to converge, they converge very slowly. The OS-EM algorithm[8] has become very attractive for image reconstruction in tomography owing to its fast convergence rate compared with the EM algorithms. It converges approximately M times faster than the EM algorithms, where M is the number of subsets. However, the OS-EM algorithm is not guaranteed to converge. After many iterations, the OS-EM algorithm appears to oscillate between solutions rather than converge to an ML solution. Several approaches have been proposed to solve the convergence problem of the OS-EM algorithm, such as the rescaled block-iterative EMML algorithm,[9,10] the row-action maximum likelihood algorithm[11] and its regularized version, the block-sequential regularized EM algorithm.[12] However, the rescaled block-iterative-EMML algorithm converges to a solution only in the consistent case. The same algorithm with a feedback approach[10] seems to be impractical for real applications, and it does not include the smoothness penalty function. The convergence proofs for the row-action ML and the block-sequential regularized EM algorithms invoked a strong *a posteriori* assumption that the objective sequence is convergent.

An alternative to the EM algorithms for image restoration and reconstruction is the separable-paraboloidal-surrogates (SPS) algorithm.[13,14] An OS version of the SPS algorithm[15] was first introduced for transmission tomography. Although the OS-SPS algorithm improves the objective function more rapidly than the SPS algorithm in the early iterations, convergence is not necessarily achieved. To overcome the convergence problem of the OS-SPS algorithm, the relaxed OS-SPS algorithm[16,17] was proposed recently by introducing the relaxation parameter into the algorithm. This algorithm not only retains the fast convergence rate of the OS-SPS algorithm but also is guaranteed to converge globally. In contrast, the relaxed version of the OS-EM algorithm is not guaranteed to converge to the optimal point.[11,16,17] Therefore in this paper we focus entirely on the relaxed OS-SPS algorithm for image restoration.[18]

Owing to the ill-posed nature of image restoration, our algorithm is based on PL estimation. Most existing OS methods have been applied to image reconstruction in tomography only rather than to image restoration. Effective use of OS methods in image restoration requires one to choose subsets appropriately to provide fast convergence rates. The global convergence property of relaxed OS-SPS holds for all choices of subsets. However, the convergence *rates* of most OS algorithms depend on the choice of subsets and scaling functions [the scaling functions are the diagonal entries of the scaling matrix D in Eq. (20) below]. Since the scaling functions in the ordi-

1084-7529/2003/030439-11$15.00

nary OS-SPS algorithm provide reasonably fast convergence, we focus here on finding subsets that provide fast initial convergence rates.

In tomography, the data are organized by projection angles, so the subsets used in tomography are unsuitable for pixel-based image restoration. Bertero and Boccacci applied the OS-EM method to the restoration of the large binocular telescope images.[19] However, the structure of that telescope's imaging is similar to that of computed tomography: multiple views of the same object are observed at different angles. Thus this technique cannot be applied to typical image restoration problems. In this paper we focus on the more traditional image restoration problem of recovering a scene from a single blurred, noisy measured image under the simplifying assumption that the point-spread function (PSF) is known. Instead of choosing subsets by downsampling projection angles as in tomography, for restoration we choose subsets by downsampling pixels rather than dividing pixels into subblocks. We show quantitatively that the downsampling approach satisfies the "subset-gradient-balance" conditions,[20] which are less restrictive than the subset-balance conditions defined in Ref. 8. These gradient-balance conditions are important for fast convergence.

This paper is organized as follows. Section 2 describes the measurement model and the objective function based on PL estimation. The derivation of the relaxed OS-SPS algorithm for image restoration starting from the basic idea of OS methods and the OS-SPS algorithm is presented in Section 3. Subset design for restoration problems is discussed in Section 4. In Section 5, we develop some efficient implementation strategies and quantify the computational complexity for the relaxed OS-SPS algorithm. Simulation results and the performance of subset designs are presented in Section 6. Conclusions are given in Section 7.

2. MEASUREMENT MODEL

In image restoration problems, the measurements are usually degraded by blur and noise. To recover an approximation of the original image, one can use the statistical characteristics of the measurement system to specify an objective function that is to be maximized. Since image restoration is an ill-posed problem, we focus on PL estimation, using an objective function of the following form:

$$\Phi(x) = L(x) - \beta R(x), \tag{1}$$

where x denotes the image parameter vector to be estimated, L denotes the log-likelihood function of the measurement, R denotes a roughness penalty function, and β denotes a parameter that controls the trade-off between resolution and noise in the restored image.

For photon-limited imaging (such as confocal microscopy), the noisy measurement $Y = [Y_1, \ldots, Y_N]$ can be modeled (approximately[21,22]) as follows:

$$Y_i \sim \text{Poission}\{[Ax]_i + b_i\}, \qquad i = 1,\ldots,N, \tag{2}$$

where A is the system matrix that is assumed to be known, b_i represents the mean number of the background noise and dark current, and N is the number of measurement pixels. The corresponding log-likelihood function is given by

$$L(x) = \sum_{i=1}^{N} \psi_i(l_i(x)), \tag{3}$$

where $\psi_i(l) = y_i \log(l + b_i) - (l + b_i)$, ignoring irrelevant constants independent of x, $l_i(x) = [Ax]_i = \Sigma_{j=1}^{P} a_{ij} x_j$, P is the number of pixels to be estimated, and the measured values y_i's are samples of independent Poisson random variables Y_i's.

To reduce noise, we penalize the differences between neighboring pixels using a roughness penalty function of the form

$$R(x) = \sum_{i=1}^{r} \psi^R([Cx]_i), \tag{4}$$

where ψ^R is the potential function, C is the penalty matrix, and r is the number of pairs of neighboring pixels. For the first-order neighborhood, the matrix C consists of horizontal and vertical cliques. For example, with a 2×2 image, the matrix C can be written as follows:

$$Cx = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_2 - x_1 \\ x_4 - x_3 \\ x_3 - x_1 \\ x_4 - x_2 \end{pmatrix} \tag{5}$$

We assume that each potential penalty function $\psi^R(t)$ satisfies the following conditions[14,23,24]:

- $\psi^R(t)$ is symmetric.
- $\psi^R(t)$ is everywhere differentiable (and therefore continuous).
- $\dot{\psi}^R(t) = \mathrm{d}/\mathrm{d}t\, \psi^R(t)$ is nondecreasing [and hence $\psi^R(t)$ is convex].
- $\omega(t) \triangleq \dfrac{\dot{\psi}^R(t)}{t}$ is nonincreasing for $t \geqslant 0$.
- $\omega(0) = \lim_{t \to 0} \dfrac{\dot{\psi}^R(t)}{t}$ is finite and nonzero.

With proper regularization, the objective function has a unique global maximum. Thus our goal is to estimate x by finding the maximizer of the objective function:

$$\hat{x} \triangleq \arg\max_{x \geqslant 0} \Phi(x). \tag{6}$$

Because closed-form solutions are unavailable for the maximizer, iterative algorithms are needed.

3. THE ALGORITHMS

This section summarizes the general principles of ordered-subset (also called block-iterative) methods and reviews the OS-SPS algorithm.

A. Ordered-Subset Technique

One can decompose the objective function in Eq. (1) into subobjective functions f_m as follows:

$$\Phi(x) = \sum_{m=1}^{M} f_m(x), \tag{7}$$

where M is the total number of subsets, chosen by the algorithm designer. Let $\{S_m\}_{m=1}^{M}$ be a disjoint partition of $\{1, \ldots, N\}$ such that $\cup_{m=1}^{M} S_m = \{1, \ldots, N\}$. Then f_m's are obtained by replacing a sum over all pixel indices in the likelihood function of Eq. (3) with a sum over a subset of data S_m and scaling the penalty term by M as follows:

$$f_m(x) \triangleq \sum_{i \in S_m} \psi_i(l_i(x)) - \frac{\beta}{M} R(x). \tag{8}$$

Instead of the restrictive subset-balance conditions originally defined in Ref. 8, we define the following less restrictive conditions called the subset-gradient-balance conditions, which use an approximate gradient computed from only a part of the data:

$$\nabla f_1(x) \cong \nabla f_2(x) \cong \ldots \cong \nabla f_m(x), \tag{9}$$

where $\nabla f_m(x)$ is the gradient of the subobjective function at the mth subset. When these conditions hold, the gradient of the objective function $\Phi(x)$ can be approximated as follows:

$$\nabla \Phi(x) \cong M \nabla f_m(x), \qquad m = 1, \ldots, M. \tag{10}$$

Using approximation (10), one can replace $\nabla \Phi(x)$ with $M \nabla f_m(x)$ in any gradient-based algorithm to construct an OS version of that algorithm. If proper subsets satisfying the above conditions are combined with suitable scaling functions [e.g., see Eq. (20) below], then the OS algorithms often exhibit acceleration by a factor of M in the early iterations. These conditions influence the rate of convergence of OS-SPS but are not required for global convergence of relaxed OS-SPS.

Figure 1 illustrates how ordinary OS algorithms work. For this figure we assumed that $\nabla \Phi(x) = \nabla f_1(x) + \nabla f_2(x)$. For x far from the solution, the condition $\nabla f_1(x) \cong \nabla f_2(x)$ holds and accelerated convergence speed is achieved. However, in the later iterations when x is close to the optimal solution, the subset-gradient-balance conditions are no longer valid, and a limit-cycle behavior

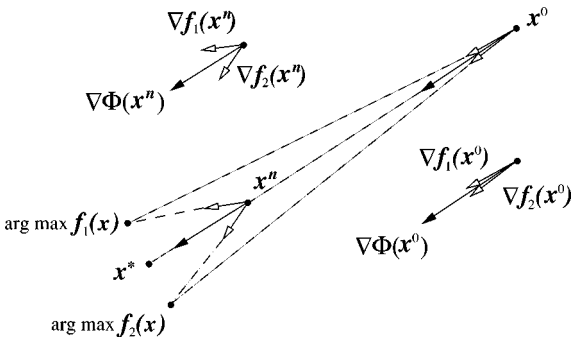


Fig. 1. Illustration of how the OS algorithms work. Assume that $\nabla \Phi(x) = \nabla f_1(x) + \nabla f_2(x)$. When x is far from the solution, the subset-gradient-balance conditions hold, and an order-of-magnitude acceleration can be achieved in the early iterations. However, for later iterations or when x is near the optimal solution, those conditions are no longer valid, and a limit-cycle behavior is observed.

around the optimal solution appears. Because ordinary OS algorithms use the same step size at each iteration, the limit cycle does not vanish. One way to suppress this limit cycle is to use a diminishing step size at each iteration (relaxation).

As previously discussed, OS-EM is not guaranteed to converge even in the relaxed version.[11,15,16] Therefore we focus on the convergent relaxed OS-SPS algorithm hereafter.

B. Ordered-Subset, Separable-Paraboloidal-Surrogate Algorithm

The SPS algorithm is based on paraboloidal surrogate functions[13,14,25] and the concavity technique developed by De Pierro.[7] The pixel update for the SPS algorithm can be summarized as follows:

$$x_j^{n+1} = \left[x_j^n + \frac{\nabla_j \Phi(x^n)}{\sum_{i=1}^{N} a_{ij} \gamma_i c_i^n + \beta p_j^n} \right]_+, \tag{11}$$

where the symbol $[x]_+$ represents x if $x > 0$ and 0 if $x \leq 0$. The gradient of the objective function at the jth pixel in the SPS algorithm is as follows:

$$\nabla_j \Phi(x) = \frac{\partial}{\partial x_j} \Phi(x) = \sum_{i=1}^{N} a_{ij} \dot{\psi}_i(l_i(x)) - \beta \sum_{i=1}^{r} c_{ij} \dot{\psi}^R([C_x]_i), \tag{12}$$

where c_{ij} is the ijth element of the matrix C. In Eq. (11), $\gamma_i = \Sigma_{j=1}^{P} a_{ij}$, and c_i^n is the following optimal curvature that guarantees convergence of SPS[13,25]:

$$c_i^n = \begin{cases} \left\{ \frac{2}{(l_i^n)^2} [\psi_i(l_i^n) - \psi_i(0) - l_i^n \dot{\psi}_i(l_i^n)] \right\}_+, & l_i^n > 0 \\ [-\ddot{\psi}_i(0)]_+, & l_i^n = 0 \end{cases}, \tag{13}$$

where $l_i^n = \Sigma_{i=1}^{P} a_{ij} x_j^n$. For the penalty function terms, the curvature p_j^n in Eq. (11) is given by

$$p_j^n = \sum_{i=1}^{r} c_{ij} \nu_i \omega([Cx^n]_i), \tag{14}$$

where $\nu_i = \Sigma_{j=1}^{p} c_{ij}$ and $\omega(t) = [\dot{\psi}^R(t)]/t$.

Erdoğan and Fessler[17] introduced the OS version of the SPS algorithm for transmission tomography. With use of approximation (10), the gradient of the objective function in Eq. (11) is replaced by the subobjective function multiplied by the number of subsets. We define $x_j^n \triangleq x_j^{(n,0)}$ and $x_j^{(n+1)} \triangleq x_j^{(n,M)}$. The first superscript refers to iterations, and the second superscript refers to subsets. Then the pixel update x_j for the OS-SPS algorithm becomes

$$x_i^{(n,m)} = \left[x_j^{(n,m-1)} + M \frac{\nabla_j f_m(x^{(n,m-1)})}{d_j + \beta p_j} \right]_+, \qquad m = 1, \ldots, M, \tag{15}$$

where

$$\nabla_j f_m(x) = \sum_{i \in S_m} a_{ij} \dot{\psi}_i(l_i(x)) - \frac{\beta}{M} \sum_{i=1}^{r} c_{ij} \dot{\psi}^R([Cx]_i). \tag{16}$$

Since the global convergence is not affected by the curvatures as long as they are positive, we precompute the curvatures d_j and p_j to save computation.[15] The curvature of the likelihood d_j in Eq. (15) is precomputed as follows:

$$d_j = \sum_{i=1}^{N} a_{ij} \gamma_i c_i, \tag{17}$$

where $c_i = -\ddot{\psi}_i(y_i - b_i)$. Likewise, the curvature of the penalty function p_j is precomputed as follows:

$$p_j = \sum_{i=1}^{r} c_{ij} \nu_i \omega(0). \tag{18}$$

Although the OS-SPS algorithm yields an order-of-magnitude acceleration over the SPS algorithm in the early iterations, it is not guaranteed to converge to the optimal solution.

C. Relaxed Ordered-Subset Separable-Paraboloidal-Surrogate Algorithm

To guarantee the convergence of the OS-SPS algorithm, Ahn and Fessler[15,16] modified the OS-SPS algorithm to include relaxation. Without relaxation, the OS-SPS algorithm has a constant step size, thus exhibiting a limit-cycle behavior. With modification of Eq. (15), the pixel update of the relaxed OS-SPS algorithm becomes

$$x_j^{(n,m)} = \left[x_j^{(n,m-1)} + \alpha_n M \frac{\nabla_j f_m(x^{(n,m-1)})}{d_j + \beta p_j} \right]_+, \quad m = 1, \ldots, M. \tag{19}$$

Equivalently, in the matrix-vector form,

$$x^{(n,m)} = \{x^{(n,m-1)} + \alpha_n D \nabla f_m[x^{(n,m-1)}]\}_+, \tag{20}$$

where $D = \mathrm{diag}\{M/(d_j + \beta p_j)\}$ is the diagonal scaling matrix. We use the same scaling functions as in the ordinary OS-SPS algorithm because they were shown to provide fairly fast initial convergence in the ordinary OS-SPS algorithm.[17] Finding optimal scaling functions for convergence speed is still an open question. A positive relaxation parameter α_n is chosen such that $\Sigma_n \alpha_n = \infty$ and $\Sigma_n \alpha_n^2 < \infty$. We use $\alpha_n = \xi/[(\xi - 1) + n]$, where ξ is a positive constant, a tuning parameter that affects the rate of convergence and is chosen empirically. The optimal choice of the relaxation parameter still remains an open question. With the diminishing step size, the relaxed OS-SPS algorithm is globally convergent.[15,16] The outline for the relaxed OS-SPS algorithm is shown below:

Precompute:

$$d_j = -\sum_{i=1}^{N} a_{ij} \gamma_i \ddot{\psi}_i(y_i - b_i),$$

$$p_j = \sum_{i=1}^{r} c_{ij} \nu_i \omega(0),$$

for $n = 1, \ldots,$ Niters

$$\alpha_n = \frac{\xi}{(\xi - 1) + n}$$

for $m = 1, \ldots, M$

$$\hat{l}_i = \sum_{j=1}^{P} a_{ij} x_j^{(n,m-1)}, \quad \forall_i \in S_m \tag{21}$$

$$\dot{\psi}_i = \frac{y_i}{\hat{l}_i + b_i} - 1, \quad \forall_i \in S_m$$

for $j = 1, \ldots, P$

$$\dot{L}_j = \sum_{i \in S_m} a_{ij} \dot{\psi}_i \tag{22}$$

$$\dot{R}_j = \sum_{i=1}^{r} c_{ij} \dot{\psi}^R([Cx^{(n,m-1)}]_i)$$

$$x_j^{(n,m)} = \left[x_j^{(n,m-1)} + \alpha_n M \frac{\dot{L}_j - \frac{\beta}{M} \dot{R}_j}{d_j + \beta p_j} \right]_+$$

end

end

end

D. Blind Restoration

Many blind restoration techniques have been applied to simultaneously restore the image and estimate the PSF.[26–29] The relaxed OS-SPS algorithm is applicable to blind restoration as well. For a blind restoration technique, the image can be updated with the relaxed OS-SPS algorithm, whereas the PSF can be updated with the ordinary SPS or EM algorithms because of the small number of parameters in the PSF.

4. SUBSET DESIGN

Since most OS algorithms have been used for image reconstruction to date, a different strategy for choosing subsets in image restoration needs to be considered because of differences in data acquisition. A good choice of subsets should satisfy the subset-gradient-balance conditions stated in Eqs. (9) and (10) to provide rapid convergence. In tomography, the subsets are chosen from downsampling projection angles. Since data in image restoration are based on pixel locations instead of projection angles as in tomography, one possible approach for choosing the subsets in the restoration problem is to downsample pixels in the image. Figure 2 shows possible choices of four subsets for a two-dimensional image. We define "4×1" OS-SPS for a downsampling approach with four subsets in each column and one subset in each row, as shown in Fig. 2(a). The downsampling approaches seem to satisfy the subset-gradient-balance conditions. To verify this, we compared the gradients of the original objective function and the subobjective functions using four subsets with a 2×2 configuration (Fig. 3). Specifically, we computed the gradients of the subobjective functions belonging to subsets 1 and 4 and then compared them with the gradient of the original objective function, as shown in the

1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4
1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4

(a) "4x1" OS–SPS

1	2	3	4	1	2	3	4
1	2	3	4	1	2	3	4
1	2	3	4	1	2	3	4
1	2	3	4	1	2	3	4
1	2	3	4	1	2	3	4
1	2	3	4	1	2	3	4
1	2	3	4	1	2	3	4
1	2	3	4	1	2	3	4

(b) "1x4" OS–SPS

1	3	1	3	1	3	1	3
2	4	2	4	2	4	2	4
1	3	1	3	1	3	1	3
2	4	2	4	2	4	2	4
1	3	1	3	1	3	1	3
2	4	2	4	2	4	2	4
1	3	1	3	1	3	1	3
2	4	2	4	2	4	2	4

(c) "2x2" OS–SPS

Fig. 2. Possible choices for four subsets with a downsampling approach. These subsets tend to satisfy the subset-gradient-balance conditions. The first number in quotation marks is the number of subsets in each column, and the second number is the number of subsets in each row. The total number of subsets is the product of these two numbers. The pixel label m belongs to the respective set S_m.

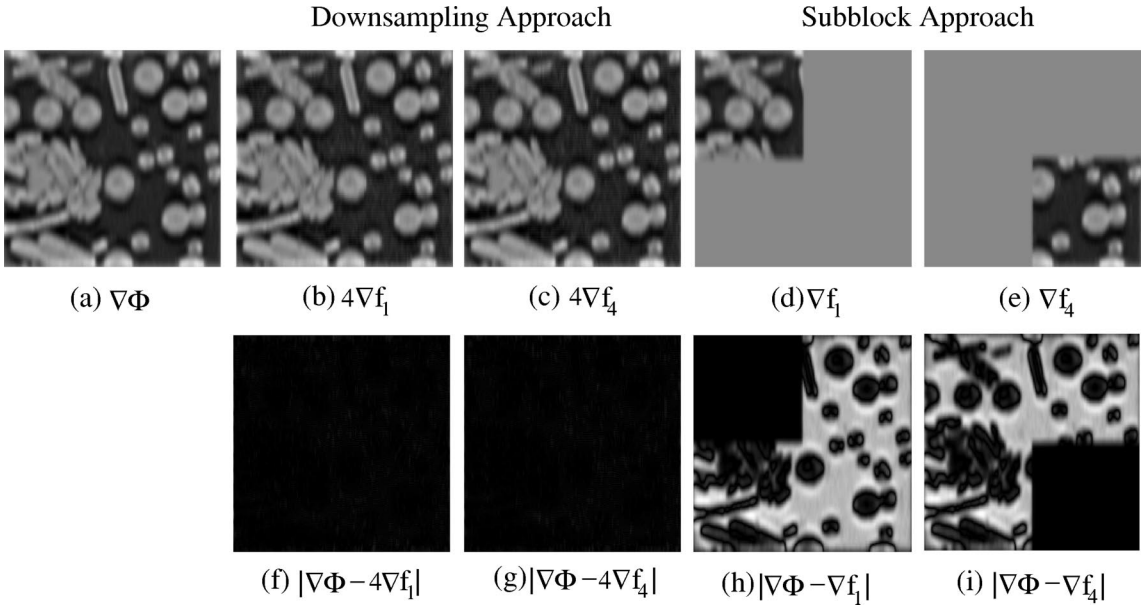


Fig. 3. Investigation of the subset-gradient-balance conditions in the OS-SPS algorithm. Four subsets with a 2×2 configuration were used. The second and third columns show the gradients of the subobjective functions from the downsampling approach with use of subset 1 and subset 4, respectively, and their differences compared with the gradient of the objective function. Similarly, the last two columns are from the subblock approach. The gradients of the subobjective functions in the downsampling approach were multiplied by 4 to compensate for the downsampled data. However, this scaling factor is not needed in the subblock approach, because a block of contiguous pixels is used.

second and third columns of Fig. 3. These differences are very small: the normalized rms error between the actual gradient $\nabla\phi$ and the subgradient $4\nabla f_m$ is less than 0.5%.

Another choice for choosing subsets is to divide the image into large contiguous blocks, called the subblock approach (Fig. 4). We define "4×1B" OS-SPS for a subblock approach with four subblocks in each column and one subblock in each row, as shown in Fig. 4(a). This approach tends to be a poor choice of subsets because it fails to satisfy the subset-gradient-balance conditions, as illustrated in the last two columns of Fig. 3. The differences between the gradients of the various subobjective functions using different subsets are large: the normalized rms is more than 65%. Section 6 reports empirical comparisons of how these possible choices of subsets affect the convergence speed.

5. IMPLEMENTATION TECHNIQUES AND COMPLEXITY

Most of the computation time in the OS-SPS algorithm takes place in Eqs. (21) and (22). In this section we discuss how to efficiently implement these two expressions for both space-variant and space-invariant systems.

A. Space-Variant Systems

A literal implementation of Eqs. (21) and (22) in the algorithm outline would be appropriate for a shift-variant imaging system whose collection of PSFs is tabulated as a sparse set of a_{ij} values. With this technique, the computational complexity of the OS-SPS algorithm is essentially the same as in the nonordered-subsets (non-OS) algorithm, except that the penalty-function gradient must

be evaluated m times per iteration. However, usually the likelihood-gradient computation dominates.

B. Space-Invariant Systems with Convolution

For shift-invariant systems, however, one would typically implement Eqs. (21) and (22) using convolution or fast Fourier transform(FFT)-based convolution in the conventional single-subset type of the gradient-based iteration. Since these operations dominate the algorithm, it is essential to formulate efficient implementations of these two expressions. Computing all values of $\hat{l}$ by using ordinary convolution would be inefficient when only some values of $\hat{l}$ will be used in a given subiteration. Therefore in this section we introduce the following technique for computing Eqs. (21) and (22) efficiently with convolution.

For a space-invariant system, we rewrite Eq. (21) in the convolution form as follows:

$$\hat{l}_i = \sum_{j=1}^{P} h_{i-j} x_j, \qquad \forall i \in S_m, \tag{23}$$

where h is the PSF. For illustration, we describe one-dimensional convolution. Extension to two and three dimensions is straightforward. To compute some values of $\hat{l}$ efficiently, we rewrite Eq. (23) using two summations:

$$\hat{l}_i = \sum_{m=1}^{M} \sum_{j \in S_m} h_{i-j} x_j, \qquad \forall i \in S_m. \tag{24}$$

Using this expression, we can compute $\hat{l}_i$ for $i \in S_m$ by convolving the downsampled image and the PSF belonging to subset S_m and then summing all the subsets (Fig. 5).

Similarly, to compute Eq. (22) efficiently by convolution, we can rewrite that expression as follows:

$$\dot{L}_j = \sum_{i \in S_m} h_{i-j} \dot{\psi}_i. \tag{25}$$

For each j, $\dot{L}_j$ can be computed by using $\dot{\psi}_i$ and a downsampled PSF. Different j's require a different downsampling of the PSF but use the same $\dot{\psi}_i$'s (Fig. 6). In this figure, the PSF is assumed to be symmetric. Otherwise, the indices of the PSF must be inverted before convolving.

If implemented carefully, computational complexity for this convolution technique does not increase as the number of subsets increases.

C. Space-Invariant Systems with Use of Fast Fourier Transforms

For simultaneous update methods, such as the EM algorithms for image restoration, one can use FFTs to reduce computation, especially for large 3D problems. Similarly, a strategy for using FFTs in the OS-SPS algorithm would be desirable to compute $\dot{L}_j$ and $\hat{l}_i$ efficiently. One possible solution is to implement the partial FFT algorithm,[30] where only a small number of frequencies are evaluated. Since there is a specific pattern for com-

1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4

(a) "4x1B" OS–SPS

1	1	2	2	3	3	4	4
1	1	2	2	3	3	4	4
1	1	2	2	3	3	4	4
1	1	2	2	3	3	4	4
1	1	2	2	3	3	4	4
1	1	2	2	3	3	4	4
1	1	2	2	3	3	4	4
1	1	2	2	3	3	4	4

(b) "1x4B" OS–SPS

1	1	1	1	3	3	3	3
1	1	1	1	3	3	3	3
1	1	1	1	3	3	3	3
1	1	1	1	3	3	3	3
2	2	2	2	4	4	4	4
2	2	2	2	4	4	4	4
2	2	2	2	4	4	4	4
2	2	2	2	4	4	4	4

(c) "2x2B" OS–SPS

Fig. 4. Possible choices for four subsets with a subblock approach. These subsets tend to violate the subset-gradient-balance conditions. The first number in quotation marks is the number of subblocks in each column, and the second number is the number of subblocks in each row.

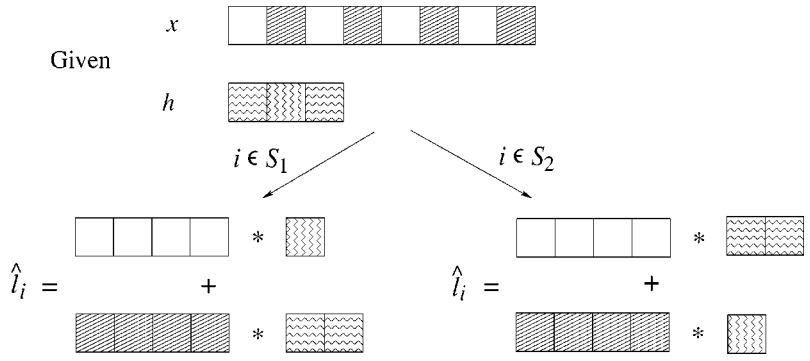


Fig. 5. Illustration of computing $\hat{l}_i$, $\forall i \in S_m$ ($M = 2$), using all the information of x and h. The asterisk represents convolution. The white blocks denote elements of x belonging to subset $m = 1$, and the striped blocks denote elements of x belonging to subset $m = 2$. The patterns in h indicate the symmetry of the PSF that has the center in the middle.

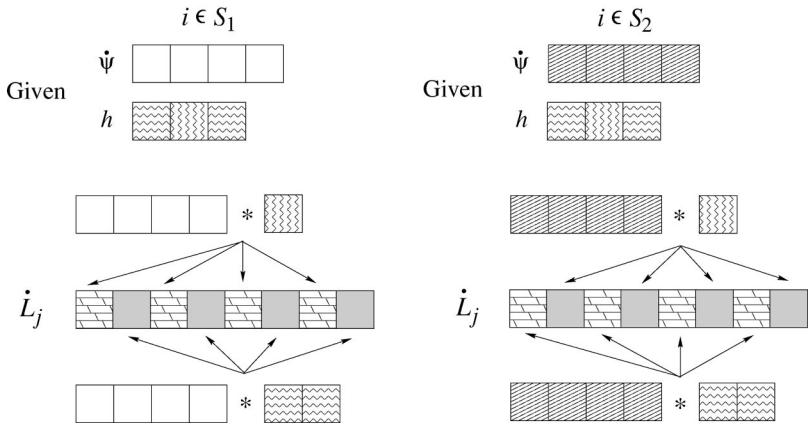


Fig. 6. Illustration of computing $\dot{L}_j$, $\forall_j$ $(M = 2)$, using some information of $\dot{\psi}_i$ but all the information of h.

puting $\hat{l}_i$ and $\dot{L}_j$ in each subset, rather than adapting and implementing this partial FFT technique into our algorithm we develop the following technique based on the ordinary FFT algorithm, which should yield the same complexity but avoids implementing new FFT code.

To describe our technique, we consider 1D data and two subsets $(M = 2)$. Let spatial indices be replaced by η to avoid confusion, $H(k)$ be an N-point FFT of h, and $X(k)$ be an N-point FFT of x. We assume that $P = N$ in this case. To compute $\hat{l}$ for 2 subsets using FFTs, we reformulate Eq. (23) into the following expression:

$$\hat{l}(\eta) = \frac{1}{N}\sum_{k=0}^{N-1} H(k)X(k)\exp\left(j\frac{2\pi\eta k}{N}\right), \qquad \forall \eta \in S_m. \tag{26}$$

Let $\eta = 0,\ldots,N/2-1$. Then the even indices of $\hat{l}$ belonging to subset 1 and the odd indices belonging to subset 2 are computed as follows:

$m = 1$:

$$\hat{l}(2\eta) = \frac{1}{N}\sum_{k=0}^{N/2-1} [H(k)X(k) + H(k+N/2)X(k+N/2)] \times \exp\left(j\frac{2\pi\eta k}{N/2}\right).$$

$m = 2$:

$$\hat{l}(2\eta+1) = \frac{1}{N}\sum_{k=0}^{N/2-1} [H(k)X(k) - H(k+N/2)X(k+N/2)]\exp\left(j\frac{2\pi k}{N}\right)\exp\left(j\frac{2\pi\eta k}{N/2}\right).$$

In this technique, a full N-point FFT is performed for h and x, but an N/M-point inverse FFT (IFFT) is performed on $\hat{l}$ for each subset. Given that the FFT of H is precomputed, the total number of complex multiplications required for computing $\hat{l}_i$ in one iteration of the OS-SPS algorithm with M subsets with use of FFT is given as follows:

$$\frac{MN}{2}\log_2 N + MN + \frac{N(M-1)}{M} + \frac{N}{2}\log_2\left(\frac{N}{M}\right). \tag{27}$$

The first term is for computing the FFT of x; the second term is for multiplying X and H; the third term is for multiplication by $\exp[j(2\pi k/N)]$; the fourth term is for the IFFT that yields $\hat{l}_i$, $i \in S_m$. For comparison, the number of complex multiplications for computing $\hat{l}_i$ in the non-OS algorithm is $N\log_2(2N)$. Table 1 compares the complexity of computing $\hat{l}_i$ in one iteration for the OS algorithm relative to the non-OS algorithm when FFTs are used. Although the number of complex multiplications increases as the number of subsets increases, it increases less rapidly than the number of subsets. Since the convergence rate increases roughly by a factor of number of subsets,[8,15,16,18] there is still an advantage in using FFTs in the OS-SPS algorithm, especially when N is large.

As in the case of $\hat{l}$, to compute $\dot{L}_j$ efficiently using FFT, we rewrite Eq. (25) in the following FFT form (assuming that h is symmetric):

$$\dot{L}(\eta) = \frac{1}{N}\sum_{k=0}^{N-1} H(k)\Psi(k)\exp\left(j\frac{2\pi\eta k}{N}\right), \qquad \forall \eta. \tag{28}$$

$\dot{L}$ is obtained by performing an N-point IFFT of the product of $H(k)$ and $\Psi(k)$; however, $H(k)$ and $\Psi(k)$ are computed from the reduced data given in each subset, i.e., even and odd sets of data for a two-subset case. Thus for $k = 0,\ldots,N/2-1$, we compute N-point $H(k)$ and N-point $\Psi(k)$ for both subsets as follows:

$m = 1$:

$$H(k) = \sum_{\eta=0}^{N/2-1} h(2\eta)\exp\left(-j\frac{2\pi\eta k}{N/2}\right) = H(k+N/2),$$

$$\Psi(k) = \sum_{\eta=0}^{N/2-1} \dot{\psi}(2\eta)\exp\left(-j\frac{2\pi\eta k}{N/2}\right) = \Psi(k+N/2).$$

$m = 2$:

$$H(k) = \exp\left(-j\frac{2\pi k}{N}\right)\sum_{\eta=0}^{N/2-1} h(2\eta+1)\exp\left(-j\frac{2\pi\eta k}{N/2}\right) = -H(k+N/2),$$

$$\Psi(k) = \exp\left(-j\frac{2\pi k}{N}\right)\sum_{\eta=0}^{N/2-1} \dot{\psi}(2\eta+1)\exp\left(-j\frac{2\pi\eta k}{N/2}\right) = -\Psi(k+N/2).$$

Thus $N/2$-point FFTs are performed to obtain the first halves of $H(k)$ and $\Psi(k)$. In this case, the multiplication complexity for computing $\dot{L}$ is the same as the complexity for computing $\hat{l}$.

In the FFT technique described above, we illustrate the techniques only for radix-2 FFT. If the data sizes are not powers of 2, then zero padding should be applied to avoid large prime factors.[31] Our technique can yield either circular or linear convolution depending on the amount of zero padding. However, we usually perform zero padding to obtain a linear convolution.

6. SIMULATION RESULTS

In this section we illustrate the proposed algorithm with 2D simulated data in comparison with existing algorithms. We also report the characteristics of various subset choices as discussed in Section 4.

Table 1. Multiplication Complexity Ratio for Computing $\hat{l}_i$ (with Use of FFTs) of OS-SPS and non-OS Algorithms with Different Numbers of Subsets

Number of Data Points	Number of Subsets	Complexity Ratio of OS and non-OS Algorithms
64	2	1.57
	4	2.68
	8	4.91
512	2	1.55
	4	2.62
	8	4.79

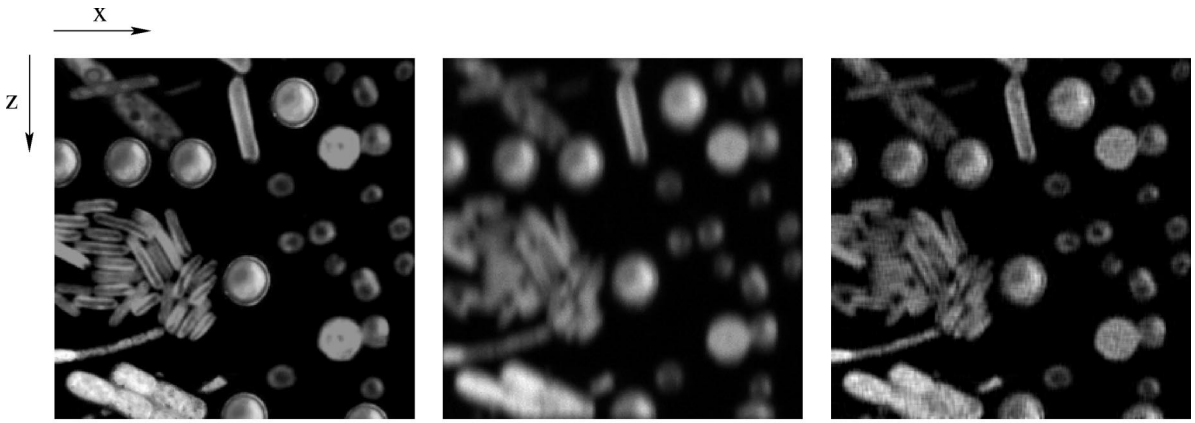
A. Two-Dimensional Results

A 256×256 cell image [Fig. 7(a)] was degraded by a 15×15 PSF, created from the XCOSM package,[32] and Poisson noise with peak signal-to-noise ratio (PSNR) of 40 dB, as shown in Fig. 7(b). The following parameters were used to create the confocal PSF from the XCOSM package[32]: pixel sizes of 0.15 μm (in all directions), $40 \times /1.0$ NA oil-immersion objective, and a fluorescent wavelength of 0.63 μm. However, we used only the central xz plane for the 2D simulation to clearly illustrate how elongation of the PSF in the z direction has been reduced after restoration. The PSNR for the data is defined as follows:

$$\mathrm{PSNR} = 10\log_{10}\left[\frac{\max_i(y_i - b_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(y_i - E[y_i])^2}\right]. \quad (29)$$

For the OS-SPS algorithm we used the relaxation parameter $\alpha_n = 11/(10+n)$, and for edge preservation we used the nonquadratic roughness penalty function[33] $\psi^R(t) = \delta^2[|t/\delta| - \log(1+|t/\delta|)]$, where δ controls the degree of edge preservation. Figure 7(c) shows the restoration from 50 iterations of the relaxed OS-SPS algorithm (eight subsets). The elongation in the z direction, very apparent in the degraded image, is greatly reduced in the restored image, thus improving the (axial) resolution.

Table 2 compares the elapsed time per iteration of different algorithms: De Pierro's modified EM (DPEM),[7] SPS (with optimal curvature), and relaxed OS-SPS (with precomputed curvature) algorithms. Theoretically, different subsets of the relaxed OS-SPS algorithm (with use of the convolution technique described in Subsection 5.B) should yield approximately the same computation time per iteration as the non-OS version. We were unable to



(a) Original image (b) Noisy blurry image (c) Restored image

Fig. 7. Simulated images and restoration using the relaxed OS-SPS algorithm with $\beta = 10^{-6}$ and $\delta = 100$. The PSF in the noisy blurry image was simulated from the 2D PSF of the confocal microscope only in the xz direction, where x is along the horizontal axis and z is along the vertical axis, to show elongation in the z direction. This elongation disappears in the restored image.

achieve that invariance, owing to MATLAB overhead, but nevertheless the computation time per iteration increases less rapidly than the number of subsets. Another way to compare the complexity of the OS-SPS algorithm with the non-OS version is by calculating the number of floating-point operations (FLOPs). Table 2 shows that the number of FLOPs required in the OS-SPS algorithms differs only slightly from the number of FLOPs required in the SPS algorithm. This agrees with our discussion given in Subsection 5.B.

Figure 8 shows the objective increase, $\Phi(x^n) - \Phi(x^0)$, at each iteration of DPEM, SPS, ordinary OS-SPS (eight subsets), and relaxed OS-SPS (eight subsets) algorithms. In this figure both ordinary OS-SPS and relaxed OS-SPS algorithms increase the objective function faster than the DPEM algorithm by roughly the number of subsets. However, the relaxed OS-SPS algorithm is guaranteed to eventually converge to the optimal point, unlike the ordinary OS-SPS algorithm. Figure 9 compares the convergence rates for different numbers of subsets. The relaxed OS-SPS-16 yields the fastest convergence rate, as expected.

Table 2. Comparison of Elapsed Times per Iteration and Number of FLOPs for DPEM, SPS, and OS-SPS Algorithms

Algorithm	Time/iter (s)	Time	Number of FLOPs	FLOPs Comparison
DPEM	1.03	0.92	84,937,142	0.92
SPS	1.12	1	92,406,026	1
OS-SPS-2	1.23	1.10	92,522,010	1.00
OS-SPS-4	1.86	1.66	95,944,812	1.04
OS-SPS-8	3.65	3.26	102,919,258	1.11
OS-SPS-16	6.83	6.10	116,976,572	1.27

B. Subset Design Analysis

Because one's choice of subsets can affect the convergence rate of the algorithm, we investigated the choices of subsets discussed in Section 4. Figure 10 shows the objective increase versus iteration for different choices of subsets (Figs. 2 and 4) with use of relaxed OS-SPS. The subsets with the subblock approach show a poor unpredictable behavior in the early iterations; however, owing to relaxation, the relaxed OS-SPS algorithm with use of these subsets will eventually converge to the optimal point. However, the subblock approach does not yield an order-of-magnitude acceleration in the early iterations. This unpredictable behavior is due to the violation of the subset-gradient-balance conditions.

Unlike the subblock approach, the downsampling approach provides an order-of-magnitude acceleration in the early iterations. Therefore the downsampling approach is preferable. With the downsampling approach, different designs of subsets provided almost the same convergence rate and a similar number of FLOPs. Thus the subset configuration does not affect the convergence rate much as long as the downsampling approach is used.

7. CONCLUSIONS

In this paper we demonstrated that the relaxed OS-SPS algorithm, conventionally used for tomography, can be adapted for use in image restoration by choosing appropriate subsets of (measured) pixels. As long as the subsets are chosen by downsampling the pixels, different choices of subsets hardly affect the convergence rate of the algorithm. Similar to tomography, our method is able to achieve an order-of-magnitude acceleration over the non-OS algorithm by combining subsets that approximately satisfy our subset-gradient-balance conditions

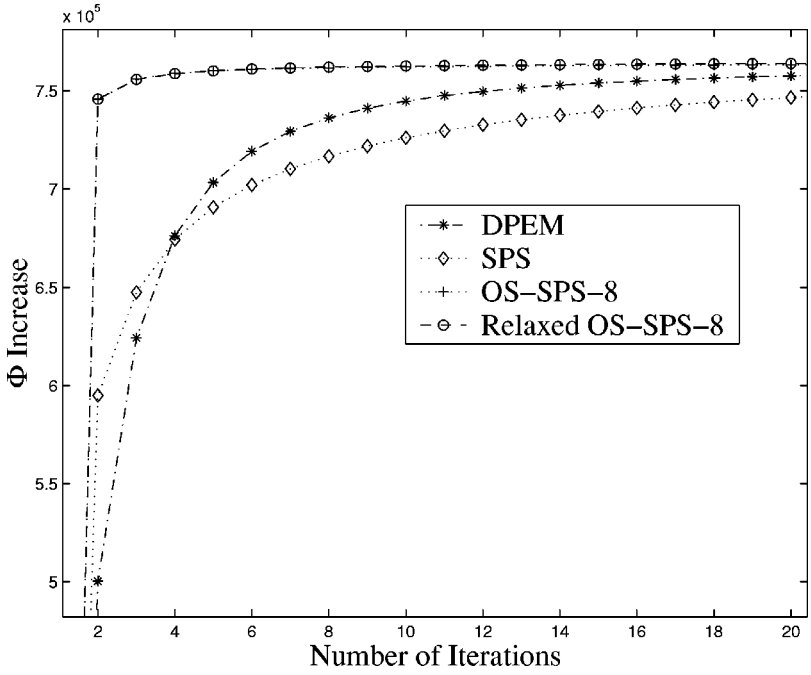


Fig. 8. Comparison of objective function increases of DPEM, SPS, OS-SPS, and relaxed OS-SPS algorithms. OS-SPS-8 stands for the OS-SPS algorithm with eight subsets. Both nonrelaxed and relaxed OS-SPS algorithms have order-of-magnitude acceleration over the DPEM and SPS algorithms.

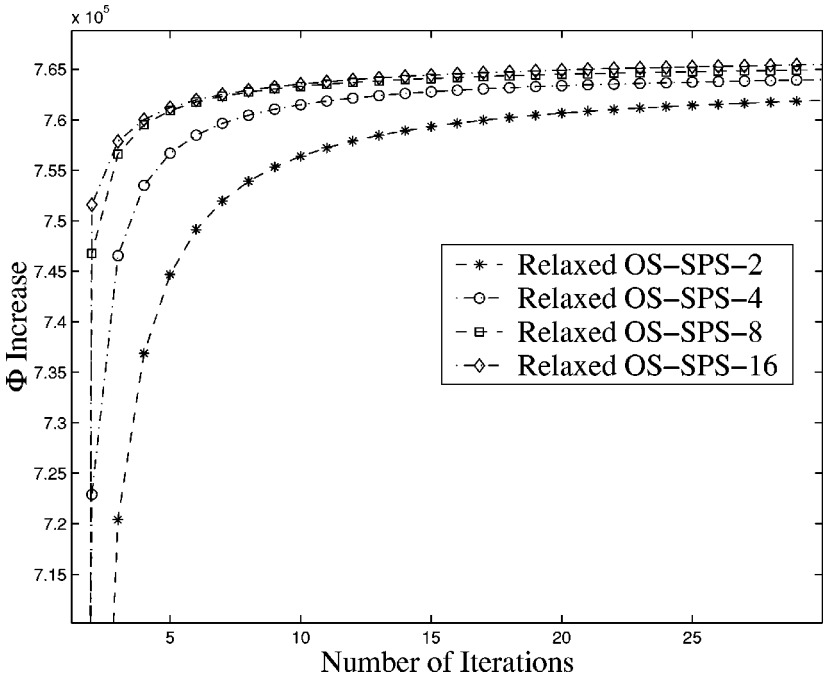


Fig. 9. Comparison of objective function increase versus elapsed time of relaxed OS-SPS with different numbers of subsets. The 16-subset relaxed OS-SPS algorithm yielded the fastest convergence rate.

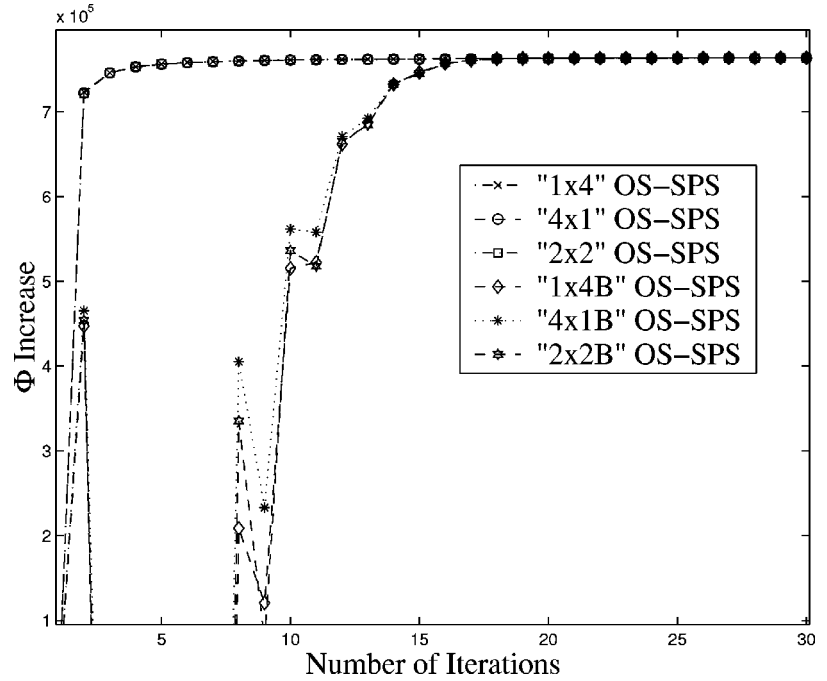


Fig. 10. Comparison of different choices of subsets with use of the relaxed OS-SPS algorithm. The subset unbalance of relaxed OS-SPS with the subblock approach causes an unpredictable behavior of the objective function increase at the beginning of iterations, but the algorithm eventually converges as a result of relaxation. The relaxed OS-SPS algorithms with the downsampling approach converge at almost the same rate for different choices of subsets.

with appropriate scaling functions in the iterative update, as shown in Eq. (20). The computational complexity of the OS-SPS algorithm with the convolution approach described in Subsection 5.B is theoretically the same for any number of subsets. Although the FFT approach described in Subsection 5.C increases the computational complexity of the algorithm when the number of subsets increases, the overall convergence rate is still faster than that of the non-OS algorithm.

The authors can be reached by e-mail at ssotthiv @umich.edu and fessler@umich.edu.

REFERENCES

1. R. Richardson, "Bayesian-based iterative method of image restoration," J. Opt. Soc. Am. **62**, 55–59 (1972).
2. T. J. Holmes, "Maximum-likelihood image restoration adapted for noncoherent optical imaging," J. Opt. Soc. Am. A **5**, 666–673 (1988).
3. S. Joshi and M. I. Miller, "Maximum *a posteriori* estimation with Good's roughness for three-dimensional optical-sectioning microscopy," J. Opt. Soc. Am. A **10**, 1078–1085 (1993).
4. R. G. Lane, "Methods for maximum-likelihood deconvolution," J. Opt. Soc. Am. A **13**, 1992–1998 (1996).
5. J. A. Conchello and J. G. McNally, "Fast regularization technique for expectation maximization algorithm for computational optical sectioning microscopy," in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing*, C. J. Cogswell, G. S. Kino, and T. Wilson, eds., Proc. SPIE **2655**, 199–208 (1996).
6. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. R. Stat. Soc. B **39**, 1–38 (1977).
7. A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," IEEE Trans. Med. Imaging **14**, 132–137 (1995).
8. H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," IEEE Trans. Med. Imaging **13**, 601–609 (1994).
9. C. L. Byrne, "Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative methods," IEEE Trans. Image Process. **7**, 100–109 (1998).
10. C. L. Byrne, "Convergent block-iterative algorithms for image reconstruction from inconsistent data," IEEE Trans. Image Process. **6**, 1296–1304 (1997).
11. J. A. Browne and A. R. D. Pierro, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography," IEEE Trans. Med. Imaging **15**, 687–699 (1996).
12. A. R. D. Pierro and M. E. B. Yamagishi, "Fast EM-like methods for maximum '*a posteriori*' estimates in emission tomography," IEEE Trans. Med. Imaging **20**, 280–288 (2001).
13. J. A. Fessler and H. Erdoğan, "A paraboloidal surrogates algorithm for convergent penalized-likelihood emission image reconstruction," in *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference* (Institute of Electrical and Electronics Engineers, New York, 1998), pp. 1132–1135.
14. H. Erdoğan and J. A. Fessler, "Monotonic algorithms for transmission tomography," IEEE Trans. Med. Imaging **18**, 801–814 (1999).
15. S. Ahn and J. A. Fessler, "Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms," IEEE Trans. Med. Imaging (to be published).
16. S. Ahn and J. A. Fessler, "Globally convergent ordered subsets algorithms: application to tomography," in *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference* (Institute of Electrical and Electronics Engineers, New York, 2001), pp. 1064–1068.
17. H. Erdoğan and J. A. Fessler, "Ordered subsets algorithm for transmission tomography," Phys. Med. Biol. **44**, 2835–2851 (1999).
18. S. Sotthivirat and J. A. Fessler, "Relaxed ordered subsets algorithm for image restoration of confocal microscopy," in *Proceedings of the IEEE International Symposium on Biomedical Imaging* (Institute of Electrical and Electronics Engineers, New York, 2002), pp. 1051–1054.
19. M. Bertero and P. Boccacci, "Application of the OS-EM method to the restoration of LBT images," Astron. Astrophys. Suppl. Ser. **144**, 181–186 (2000).
20. D. P. Bertsekas, "A new class of incremental gradient methods for least squares problems," SIAM J. Optim. **7**, 913–926 (1997).
21. D. L. Snyder, A. M. Hammoud, and R. L. White, "Image recovery from data acquired with a charge-coupled-device camera," J. Opt. Soc. Am. A **10**, 1014–1023 (1993).
22. D. L. Snyder, C. W. Helstrom, A. D. Lanterman, M. Faisal, and R. L. White, "Compensation for readout noise in CCD images," J. Opt. Soc. Am. A **12**, 272–283 (1995).
23. P. J. Huber, *Robust Statistics* (Wiley, New York, 1981).
24. J. A. Fessler, "Grouped coordinate descent algorithms for robust edge-preserving image restoration," in *Image Reconstruction and Restoration II*, T. J. Schulz, ed., Proc. SPIE **3170**, 184–194 (1997).
25. S. Sotthivirat and J. A. Fessler, "Image recovery using partitioned-separable paraboloidal surrogate coordinate ascent algorithms," IEEE Trans. Image Process. **11**, 159–173 (2002).
26. T. J. Holmes, "Blind deconvolution of quantum-limited incoherent imagery: maximum-likelihood approach," J. Opt. Soc. Am. A **9**, 1052–1061 (1992).
27. E. Thiébaut and J.-M. Conan, "Strict *a priori* constraints for maximum-likelihood blind deconvolution," J. Opt. Soc. Am. A **12**, 485–492 (1995).
28. J. Markham and J. A. Conchello, "Parametric blind deconvolution: a robust method for the simultaneous estimation of image and blur," J. Opt. Soc. Am. A **16**, 2377–2391 (1999).
29. E. Y. Lam and J. W. Goodman, "Iterative statistical approach to blind image deconvolution," J. Opt. Soc. Am. A **17**, 1177–1184 (2000).
30. J. C. D. de Melo, "Partial FFT Evaluation," in *International Conference on Signal Processing Application and Technology* (1996), Vol. 1, pp. 134–141, available at http://www.icspat.com/papers/234amfi.pdf.
31. S. K. Mitra, *Digital Signal Processing: A Computer-Based Approach*, 2nd ed. (McGraw-Hill, New York, 2001).
32. The XCOSM deconvolution package (online). Available at http://3dmicroscopy.wustl.edu/~xcosm/.
33. K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," IEEE Trans. Med. Imaging **9**, 439–446 (1990).