PARTITIONED SEPARABLE PARABOLOIDAL SURROGATE COORDINATE ASCENT ALGORITHM FOR IMAGE RESTORATION

Saowapak Sotthivirat and Jeffrey A. Fessler

University of Michigan
Dept. of Electrical Engineering and Computer Science
Ann Arbor, MI 48109

ABSTRACT

We introduce a new fast converging parallelizable algorithm for image restoration. This algorithm is based on paraboloidal surrogate functions to simplify the optimization problem and a concavity technique developed by De Pierro to simultaneously update a set of pixels. To obtain large step sizes which affect the convergence rate, we choose the paraboloidal surrogate functions that have small curvatures. The concavity technique is applied to separate pixels into partitioned sets so that parallel processors can be assigned to each set. The partitioned separable paraboloidal surrogates are maximized by using coordinate ascent (CA) algorithms. Unlike other existing algorithms such EM and CA algorithms, the proposed algorithm not only requires less time per iteration to converge, but is guaranteed to monotonically increase the objective function and intrinsically accommodates nonnegativity constraints as well.

1. INTRODUCTION

Many imaging techniques have been applied to recover degraded images, such as maximum likelihood (ML), maximum *a posteriori* (MAP), and penalized maximum likelihood (PML) estimators. Since there is no closed form solution for these techniques, iterative algorithms are needed. However, there are some drawbacks of these existing algorithms such as convergence, computation time, parallelizability, etc. Fast converging algorithms are desirable to quickly recover degraded images. Simultaneously updating all parameters in the EM algorithms [1, 2] causes a very slow convergence rate. Therefore a number of algorithms have been proposed to increase the convergence rate such as the space-alternating generalized EM (SAGE) algorithm [3, 4]. The SAGE method converges quickly but it is inconvenient to implement. Due to slow convergence in simultaneous updates, algorithms based on sequential updates, such as a coordinate ascent algorithm with Newton-Raphson updates (CA-NR) [5], have become attractive. However, the CA-NR algorithm is not guaranteed to converge and is not parallelizable. The paraboloidal surrogate coordinate ascent (PSCA)

algorithm [6] was introduced to solve the convergence problem of the CA-NR algorithm but it is still not parallelizable. In this paper, we present a new algorithm called partitioned separable paraboloidal surrogate coordinate ascent (PPCA) algorithm to enable a fast converging parallelizable algorithm. Instead of simultaneously updating all unknown parameters, we update all pixel subsets in parallel and sequentially update pixels within each partitioned set. This approach provides rapid convergence rates while being parallelizable as well. Therefore parallel processors can be assigned to each set to reduce the computation time. The algorithms derived in [7–9] are closely related to this work; however, they simultaneously update pixels within each group whose pixels are not coupled and sequentially update pixels between groups. Furthermore, those algorithms converge more slowly and they are more difficult to implement than the PPCA algorithm. The PSCA algorithm is the special case of the PPCA algorithm when one subset is used. The proposed algorithm is guaranteed to monotonically increase the objective function and it intrinsically accommodates the nonnegativity constraints.

2. THE PROBLEM

In image restoration problems, an image is usually degraded by blur and noise. One approach to recover the degraded image is to use statistical characteristics of the measurements. In this paper, we consider the very broad class of objective functions having the following form:

$$\Phi(x) = \sum_{i=1}^{m} \psi_i([Bx]_i) \qquad (1)$$

where $x \in \Re^p$ represents a true image and B is an $m \times p$ matrix that typically includes both a $q \times p$ system matrix and an $r \times p$ coefficient matrix of a roughness penalty function, where $m = q + r$, q is the number of measurements, and r is roughly the number of neighbors of each pixel. A ψ_i function characterizes an agreement between the noisy measurement and the unknown image. Due to the ill-posed nature

of the restoration problem, the roughness penalty function is included in some of the ψ_i functions. We assume that the objective function has a unique global maximum. Thus our goal is to estimate x by maximizing the objective function:

$$\hat{x} \triangleq \arg\max_{x \geq 0} \Phi(x). \tag{2}$$

The ML, PML and MAP estimators are all special cases of maximizing objective functions of the general form (1).

3. PARTITIONED SEPARABLE PARABOLOIDAL SURROGATE COORDINATE ASCENT ALGORITHM (PPCA)

Many existing algorithms have been applied to obtain a maximizer of $\Phi(x)$ in (2); however, there is a tradeoff between convergence rate and parallelizability. Although the EM algorithm is generally guaranteed to converge to at least a local maximum, it converges very slowly. However, the EM algorithm is usually fully parallelizable. On the other extreme, the CA algorithm, which updates the unknown parameters sequentially, converges much faster than EM algorithms. However, the CA algorithm is not parallelizable. In this paper, we propose a new algorithm which not only converges quickly but is also well-suited to coarse-grain parallel processing. The partitioned separable paraboloidal surrogate coordinate ascent (PPCA) algorithm is based on a concavity technique developed by De Pierro [2] and uses tangent parabolas.

To derive the PPCA algorithm, we first find a paraboloidal surrogate function for the original objective function, and then partition pixels into K sets. Since the parabola is concave, we can form a partitioned separable surrogate function by using a concavity technique to separate pixels into partitioned sets. Finally, the CA algorithm is applied in parallel to each set of pixels. Here is our general idea for deriving the surrogates:

$$\Phi(x) \geq Q(x; x^n) \geq \phi(x; x^n) = \sum_{k=1}^{K} Q_k(x_{J_k}; x^n),$$

where $Q(x; x^n)$ denotes the paraboloidal surrogate function, $\phi(x; x^n)$ denotes the separable paraboloidal surrogate function, $Q_k(x_{J_k}; x^n)$ denotes the partitioned separable paraboloidal surrogate function, J_k denotes the kth set of pixels and x_{J_k} denotes the vector of length $|J_k|$. Thus, instead of directly maximizing Φ, we obtain the next estimate x in each set by maximizing $Q_k(x_{J_k}; x^n)$:

$$\hat{x}_j^{n+1} \triangleq \begin{cases} \arg\max_{x_j \geq 0} Q_k(x_{J_k}; x^n), & j \in J_k \\ \hat{x}_j^n, & j \notin J_k. \end{cases} \tag{3}$$

3.1. Paraboloidal Surrogates

The paraboloidal surrogate function for the original objective function can be expressed as follows:

$$Q(x; x^n) \triangleq \sum_{i=1}^{m} q_i([Bx]_i; t_i^n)$$

$$q_i(t; t_i^n) \triangleq \psi_i(t_i^n) + \dot{\psi}_i(t_i^n)(t - t_i^n) - \frac{1}{2}c_i(t_i^n)(t - t_i^n)^2,$$

where $t_i^n \triangleq [Bx^n]_i$, $\dot{\psi}_i$ is the first derivative of ψ_i and $c_i(t_i^n)$ is the curvature of the parabola $q_i(t; t_i^n)$. The parabola is derived such that $q_i(t_i^n; t_i^n) = \psi_i(t_i^n)$ and $\dot{q}_i(t_i^n; t_i^n) = \dot{\psi}_i(t_i^n)$ for differentiable surrogate and objective functions. The choice of $c_i(t_i^n)$ controls the parabola curvature which affects the algorithm convergence rate. In addition, monotonicity is satisfied if we choose the curvature such that the following inequality at each iteration holds:

$$\psi_i(t) \geq q_i(t; t_i^n), \qquad \text{for } t \geq 0.$$

3.2. Separable Surrogates

After obtaining the paraboloidal surrogate function Q, we apply the concavity technique developed by De Pierro [2] to separate pixels into partitioned sets. We can rewrite $[Bx]_i$ as follows:

$$[Bx]_i = \sum_{k=1}^{K} \rho_{ik} \left(\frac{1}{\rho_{ik}} [B_{J_k}(x_{J_k} - x_{J_k}^n)]_i + t_i^n \right),$$

where the $m \times |J_k|$ matrix B_{J_k} is formed by selecting the columns of B that are indexed by elements of J_k. Since q_i is concave, the following inequality holds:

$$q_i([Bx]_i; t_i^n) \geq \sum_{k=1}^{K} \rho_{ik} q_i \left(\frac{1}{\rho_{ik}} [B_{J_k}(\underline{x}_{J_k} - \underline{x}_{J_k}^n)]_i + t_i^n; t_i^n \right).$$

The constraint $\sum_{k=1}^{K} \rho_{ik} = 1$ must be satisfied to guarantee monotonicity of the algorithm. A simple choice of ρ_{ik} is $\frac{\sum_{j \in J_k} |b_{ij}|}{\sum_{j=1}^{p} |b_{ij}|}$. Thus we obtain the partitioned separable paraboloidal surrogate Q_k as follows:

$$Q_k(x_{J_k}; x^n) = \sum_{i=1}^{m} \rho_{ik} q_i \left(\frac{1}{\rho_{ik}} [B_{J_k}(\underline{x}_{J_k} - \underline{x}_{J_k}^n)]_i + t_i^n; t_i^n \right).$$

3.3. Coordinate Ascent Algorithm

To implement the maximization in (3), we apply the coordinate ascent algorithm over each pixel of x_j by using the most recent values of other pixels of x_j in that set. Define

$$\hat{Q}_{kj}^n(x_j) \triangleq Q_k([\ldots, \hat{x}_{j-1}, x_j, \hat{x}_{j+1}, \ldots, j \in J_K]; x^n),$$

where $\hat{x}$ is the current estimate of x. Then the update x_j for $j \in J_k$ is sequentially estimated by maximizing $\hat{Q}_{kj}^n(x_j)$ as follows:

$$\hat{x}_j^{n+1} = \arg\max_{x_j \geq 0} \hat{Q}_{kj}^n(x_j) = \left[\hat{x}_j + \frac{\dot{Q}_{kj}^n(\hat{x}_{J_k})}{d_{kj}^n} \right]_+ , \quad (4)$$

where $\dot{Q}_{kj}^n(\hat{x}_{J_k})$ is the first derivative of $\hat{Q}_{kj}^n(x_j)$ evaluated at $x_j = \hat{x}_j$ and d_{kj} is the curvature of the parabola $\hat{Q}_{kj}^n(x_j)$. For $j \notin J_k$, we set $\hat{x}_j^{n+1} = \hat{x}_j^n$.

4. RESULTS

The 512×512 pepper image was degraded by a 15×15 Gaussian PSF with FWHM of 11.7 pixels ($\sigma = 5.0$) and Poisson noise with PSNR of 25 dB as shown in Fig. 1b. The noisy measurement Y can be modeled as follows:

$$Y_i \sim \mathrm{Poisson}\{[Ax]_i + b_i\}, \qquad i = 1, \cdots, m$$

where A is a system matrix and b_i is a background noise which is constant in this simulation. Our objective function included the likelihood function based on the above Poisson model and the nonquadratic roughness penalty function having the following potential function [10]:

$$\psi_i^{\mathrm{penalty}}(t) = \delta^2 \left[\left|\frac{t}{\delta}\right| - \log\left(1 + \left|\frac{t}{\delta}\right|\right) \right],$$

where δ controls the degree of edge preservation. The restoration with 4-PPCA algorithm (using four parallel processors) is shown in Fig. 1c.

Table 1 compares the elapsed time of the CA-NR, PSCA, and PPCA algorithms with different numbers of partitioned sets. Convergence in this table is defined such that $\Phi(x^n) - \Phi(x^0) > 0.999(\Phi(x^*) - \Phi(x^0))$, where $\Phi(x^0)$ is the objective value of the initial image, and $\Phi(x^*)$ is the smallest objective value among all methods obtained in 50 iterations. The algorithms were tested on the IBM SP2 parallel computer. Our results confirm that the PPCA algorithm is well suited for parallel processing. Our patterns for different partition sets were designed as shown in Fig. 2.

As shown in Table 1, the PPCA algorithms require less elapsed times to converge than the PSCA and CA-NR algorithms due to their parallelizability. Fig. 3 shows that the PPCA algorithms increase the objective function essentially as much per iteration as the PSCA algorithm. Fig. 4 shows that the PPCA algorithms converge at faster rates (in terms of elapsed time) than the PSCA algorithm. (The CA-NR algorithm is not included in the plots because it is a non-monotonic algorithm.)

(a) Original Image

(b) Noisy Image

(c) Restored Image

Fig. 1. Simulated images and restoration using a 4-PPCA algorithm.

Convergence	CA-NR	PSCA	2-PPCA	4-PPCA	6-PPCA
#iters	24	32	32	33	34
time/proc	395.61s	269.19s	180.57s	135.10s	126.43s
time/iter/proc	16.48s	8.41s	5.64s	4.09s	3.72s
%speed up (wrt PSCA)	-	-	32.94%	51.37%	55.77%

Table 1. Comparison of elapsed times and number of iterations to converge for CA-NR, PSCA, and PPCA algorithms.

2 partitioned sets   4 partitioned sets   6 partitioned sets
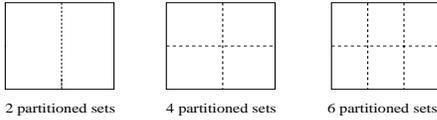
Fig. 2. Partitioned set patterns.
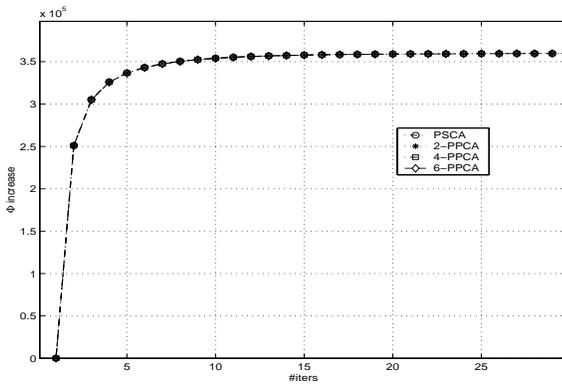
Fig. 3. Comparison of objective function increase versus iteration number of PSCA and PPCA algorithms.
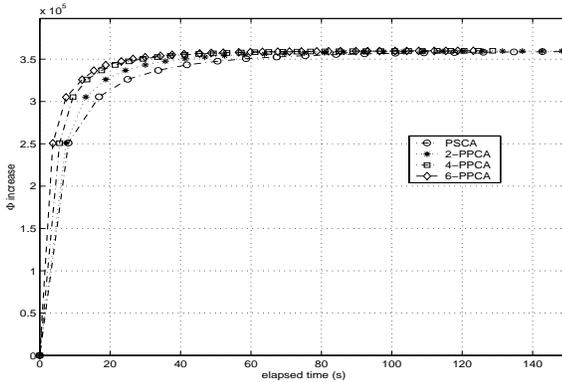
Fig. 4. Comparison of objective function increase versus elapsed time of PSCA and PPCA algorithms.

5. CONCLUSION

We have presented a new parallelizable converging algorithm that overcomes the drawbacks of the CA-NR algorithm, which is nonparallelizable and is not guaranteed to converge. Unlike EM algorithms that perform completely simultaneous updates, the PPCA algorithm has a faster con-

vergence rate due to larger updating step sizes obtained by using small parabola curvature. Furthermore, the parallel processors can be assigned to each set to reduce the computation time. Therefore the PPCA algorithm converges much faster than the CA-NR and PSCA algorithms in terms of elapsed time.

6. REFERENCES

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Roy. Stat. Soc. B*, vol. 39, pp. 1–38, 1977.

[2] A. R. De Pierro, "A Modified Expectation Maximization Algorithm for Penalized Likelihood Estimation in Emission Tomography," *IEEE Trans. Med. Imaging*, vol. 14, no. 1, pp. 132–137, March 1995.

[3] J. A. Fessler and A. O. Hero, "Space-Alternating Generalized Expectation-Maximization Algorithm," *IEEE Trans. Signal Processing*, vol. 42, no. 10, pp. 2664–2677, October 1994.

[4] 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, October 1995.

[5] C. A. Bouman and K. Sauer, "A unified approach statistical tomography using coordinate descent optimization," *IEEE Trans. Image Proc.*, vol. 5, no. 3, pp. 480–492, March 1996.

[6] H. Erdoğan and J. A. Fessler, "Monotonic Algorithms for Transmission Tomography," *IEEE Trans. Med. Imaging*, vol. 18, no. 9, pp. 801–814, September 1999.

[7] S. Saquib, J. Zheng, C. A. Bouman, and K. D. Sauer, "Parallel Computation of Sequential Pixel Updates in Statistical Tomographic Reconstruction," in *Proc. IEEE Int'l Conf. on Image Processing*, 1995, vol. 2, pp. 93–96.

[8] J. A. Fessler, "Grouped Coordinate Ascent Algorithms for Penalized-Likelihood Transmission Image Reconstruction," *IEEE Trans. Med. Imaging*, vol. 16, no. 2, pp. 166–75, April 1997.

[9] J. A. Fessler, "Grouped Coordinate Descent Algorithms for Robust Edge-Preserving Image Restoration," *SPIE 97*, vol. 3170, pp. 184–194, July 1997.

[10] K. Lange, "Convergence of EM Image Reconstruction Algorithms with Gibbs Smoothing," *IEEE Trans. Med. Imaging*, vol. 9, no. 4, pp. 439–446, December 1990.