PROVABLY CONVERGENT COORDINATE DESCENT IN STATISTICAL TOMOGRAPHIC RECONSTRUCTION

Suhail Saquib, Jun Zheng†, Charles A. Bouman*, and Ken D. Sauer†*

*School of Electrical Engineering, Purdue University
West Lafayette, IN 47907-0501, (317) 494-0340
†Department of Electrical Engineering, University of Notre Dame
Notre Dame, IN 46556, (219) 631-6999

ABSTRACT

Statistical tomographic reconstruction algorithms generally require the efficient optimization of a functional. A recent algorithm known as iterative coordinate descent with Newton-Raphson updates (ICD/NR) has been shown to be much more computationally efficient than indirect optimization approaches based on the EM algorithm. However, while the ICD/NR algorithm has experimentally been shown to converge stably, no theoretical proof of convergence is known.

In this paper, we prove that a modified algorithm, which we call ICD functional substitution (ICD/FS), has guaranteed global convergence in addition to the computational efficiency of ICD/NR. The ICD/FS method works by approximating the log likelihood at each pixel by an alternative quadratic functional. Experimental results show that the convergence speed of the globally convergent algorithm is nearly identical to that of ICD/NR.

1. INTRODUCTION

Statistical tomographic reconstruction methods such as maximum-likelihood (ML) and maximum *a posteriori* probability (MAP) estimation seek the solution that best matches the probabilistic behavior of the data. But these estimates may require excessive computation for the resulting large-scale, iterative optimizations. The expectation-maximization (EM) algorithm has been widely applied for finding ML estimation in the emission tomography problem [1]. Several other algorithms take an indirect approach based on EM to solve the MAP optimization problem[2, 3]. But these algorithms retain one main problem of EM, i.e. its slow convergence speed for tomographic reconstructions.

Recently, we have proposed an iterative coordinate descent (ICD) algorithm [4] for statistical image reconstruction based on the direct optimization of the MAP criterion. This algorithm is based on the sequential greedy optimization of pixel values in the reconstruction. It applies a truncated Taylor series expansion to derive a local quadratic approximation to the exact log likelihood function. We use the term iterative coordinate descent/Newton Raphson (ICD/NR) for this computationally advantageous version of the coordinate descent algorithm. It has been experimentally demonstrated to converge very rapidly compared to EM algorithms, but is thus far not guaranteed theoretically to converge to the unique global MAP solution.

In this paper, we present a modified ICD algorithm which we call ICD functional substitution (ICD/FS). The ICD/FS algorithm locally approximates the exact log likelihood function with an alternative quadratic function. We prove that the ICD/FS algorithm is globally convergent for both the emission and transmission reconstruction problem when the log prior distribution is convex. As with ICD/NR, the new method easily incorporates non-negativity constraints and non-Gaussian prior distributions. We also note that as a functional substitution approach, the new algorithm is mathematically similar to optimization methods allowing parallel updates of arbitrary sets of pixels within the sequential framework [5, 6].

We include experimental results on synthetic phantoms, using both ICD/NR and ICD/FS algorithms for the comparison under the emission and transmission reconstruction cases. These results indicate that ICD/FS retains the rapid convergence properties of ICD/NR, but with the desirable global convergence properties.

THIS WORK WAS SUPPORTED BY THE NATIONAL SCIENCE FOUNDATION UNDER GRANT NO. MIP93-00560.
TO APPEAR IN THE *1996 IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING*, LAUSANNE, SWITZERLAND

2. MODELING ASSUMPTIONS

The MAP tomographic reconstruction problem results in the numerical optimization problem of

$$x_{MAP} = \arg\max_x \{\log P(Y = y|X = x) + \log P(X = dx)\}$$

where x is the unknown image and y is the projection data. We refer to $\log P(Y = y|X = x)$ as the data term, and the density function $\log P(X = dx)$ as the prior term. The approach we propose is based on the ICD algorithm. In this work, we will assume that the function $\log P(X = x)$ is strictly concave with continuous derivatives on the set $\mathbb{R}^{N+} = \{x \in \mathbb{R}^N : x_j \geq 0\}$. For the emission case, we have

$$\begin{aligned} & \log P(Y = y|X = x) \qquad (1) \\ = & \sum_{i=1}^{M} (-A_{i*}x - r_i + y_i \log\{A_{i*}x + r_i\} - \log(y_i!)) \end{aligned}$$

where M is the number of projections, A_{i*} is the i^{th} row of the projection matrix A, y_i are the observed photon counts for projection i, and r_i are additive terms usually due to background noise or random coincidences in the case of PET. For the transmission case, we have

$$\begin{aligned} & \log P(Y = y|X = x) \qquad (2) \\ = & \sum_{i=1}^{M} \left(-y_T \exp^{-A_{i*}x} + y_i(\log y_T - A_{i*}x) - \log(y_i!)\right) \end{aligned}$$

where y_T is the photon dosage per ray [4].

We will prove global convergence by verifying that ICD/FS meets the conditions of the theorem presented by Fessler and Hero in [7] for convergence of space-alternating generalized EM. Since this proof requires continuity of the log likelihood on $\mathbb{R}^{N+}$, we must assume that the background noise is greater than zero, i.e. $r_i > 0$. We discuss alternative methods for the case $r_i = 0$ later in the paper.

3. THE FUNCTIONAL SUBSTITUTION APPROACH TO ICD

In this section, we will develop the ICD/FS algorithm.

3.1. ICD/FS Algorithm

The ICD method sequentially optimizes with respect to each pixel (i.e. coordinate of x). Let x^n be the image at the n^{th} iteration. Then the ICD update of the pixel x_j is computed by solving the equation

$$x_j^{n+1} = \arg\min_{x_j \geq 0} \{F_j(x_j) + P_j(x_j)\} \qquad (3)$$

where $F_j(x_j)$ and $P_j(x_j)$ represent the contribution of the data and prior terms, respectively, to the objective function expressed in terms of x_j. The particular form of $F_j(x_j)$ varies for the emission and transmission tomography cases of equations (1) and (2). But in both cases $F_j(x_j)$ is a convex function on $\mathbb{R}^{N+}$ and its derivative $f_j(x_j) = \frac{dF_j(x_j)}{dx_j}$ is strictly concave.

Unfortunately, direct optimization of (3) is not desirable because evaluation of the data term, $F_j(x_j)$, is computationally expensive. ICD/NR solves this problem by approximating the function with its second order Taylor series, replacing $F_j(x_j)$ with

$$F_j^{(nr)}(x_j) = \theta_1^{(nr)}(x_j - x_j^n) + \frac{1}{2}\theta_2^{(nr)}(x_j - x_j^n)^2,$$

where $\theta_1^{(nr)} = f_j(x_j^n)$, $\theta_2^{(nr)} = \frac{df_j(x_j^n)}{dx_j}$. Although this approximation has been shown experimentally to be quite good, a theoretical proof of convergence for the resulting iterations has not been found.

The new algorithm ICD/FS results from using a slightly different value for θ_2. This new choice of θ_2 is slightly more conservative and will allow us to prove the desired global convergence property. The update equations for ICD/FS are given as follows.

$$\theta_1^{(fs)} = f_j(x_j^n), \qquad (4)$$

$$\theta_2^{(fs)} = \begin{cases} \frac{f_j(x_j^n) - f_j(0)}{x_j^n} & \text{if } x_j^n > 0 \\ \frac{df_j(0)}{dx_j} & \text{if } x_j^n = 0 \end{cases} \qquad (5)$$

$$F_j^{(fs)}(x_j) = \theta_1^{(fs)}(x_j - x_j^n) + \frac{1}{2}\theta_2^{(fs)}(x_j - x_j^n)^2 \qquad (6)$$

$$x_j^{n+1} = \arg\min_{x_j \geq 0} \left\{F_j^{(fs)}(x_j) + P_j(x_j)\right\} \qquad (7)$$

The ICD/FS algorithm can be applied in both emission and transmission tomography problems. The only difference between these two cases is the specific computation of the values for $f_j(x_j^n)$, $f_j(0)$, and $\frac{df_j(0)}{dx_j}$ in (4) and (5). For the emission case, these values are given by

$$f_j(x_j^n) = \sum_{i=1}^{M} A_{ij}\left(1 - \frac{y_i}{p_i^n}\right) \qquad (8)$$

$$f_j(0) = \sum_{i=1}^{M} A_{ij}\left(1 - \frac{y_i}{p_i^n - A_{ij}x_j^n}\right) \qquad (9)$$

$$\frac{df_j(0)}{dx_j^n} = \sum_{i=1}^{M} y_i \left(\frac{A_{ij}}{p_i^n}\right)^2 \qquad (10)$$

where A_{ij} is the contribution of the j^{th} pixel to the i^{th} projection, and $p_i^n = A_{i*}x^n + r_i$ is the i^{th} projection of the reconstruction at iteration n. Note that p_i^n may be efficiently updated by $p_i^{n+1} = p_i^n + A_{ij}(x_j^{n+1} - x_j^n)$, with computation reduced by the sparse structure of A.

For the transmission case, the update values are given by

$$f_j(x_j^n) = \sum_{i=1}^{M} A_{ij}\left(y_i - y_T e^{-p_i^n}\right) \quad (11)$$

$$f_j(0) = \sum_{i=1}^{M} A_{ij}\left(y_i - y_T e^{-p_i^n} e^{A_{ij}x_j^n}\right) \quad (12)$$

$$\frac{df_j(0)}{dx_j^n} = \sum_{i=1}^{M} A_{ij}^2 y_T e^{-p_i^n} \quad (13)$$

where $p_i^n = A_{i*}x^n$.

ICD/FS has essentially the same computational requirements as ICD/NR since it generally requires the computation of two first derivatives in place of the first and second derivatives required for ICD/NR.

3.2. Global Convergence of ICD/FS

In order to prove the global convergence of this new algorithm, we simply verify that it meets the two assumptions and six conditions of the global convergence proof presented in [7].

Most of these conditions are either the same as for [7], or are may be simply verified [1]. However, we will demonstrate the critical Condition 1, which states that the change in the substitute function is an upper bound on the change in the true functional to be minimized.

By the construction of function $f_j^{(f_s)}(x)$, we know that $f_j^{(fs)}(0) = f_j(0)$, and $f_j^{(fs)}(x_j^n) = f_j(x_j^n)$. Since for both the emission and transmission case, $f_j(x)$ is a concave function and $f_j^{(fs)}(x)$ is a linear function, it follows that

$$f_j(x) \begin{cases} \geq f_j^{(fs)}(x) & 0 \leq x < x_j^n \\ \leq f_j^{(fs)}(x) & x > x_j^n \end{cases}$$

Integration of $f_j(x)$ and $f_j^{(fs)}(x)$ result in the inequality

$$F_j(x) - F_j(x_j^n) \leq F_j^{(fs)}(x) - F_j^{(fs)}(x_j^n) \ .$$

Defining the functions $\Phi_j(x) = F_j(x) + P_j(x)$ and $\Phi_j^{(fs)}(x) = F_j^{(fs)}(x) + P_j(x)$ then results in the following lemma.

Lemma: Let $F_j(x) + P_j(x)$ be convex, and $F_j(x)$ be continuously differentiable on $\mathbb{R}^{N+}$. Furthermore, let $f_j(x) = \frac{dF_j(x)}{dx}$ be concave and continuous on $\mathbb{R}^{N+}$, and let

$$x_j^{n+1} = \arg\min_{x\geq 0}\{F_j^{(fs)}(x) + P(x)\}.$$

Then for all $x \in \mathbb{R}^{N+}$

$$\Phi_j(x) - \Phi_j(x_j^n) \leq \Phi_j^{(fs)}(x) - \Phi_j^{(fs)}(x_j^n) \ .$$

Based on this lemma and the conditions proved in [7], the global convergence of the ICD/FS algorithm follows.

3.3. ICD/FS with Zero Background Emission Noise

As mentioned previously, the emission case when $r_i = 0$ is special since in this case the log likelihood term may tend to $-\infty$ on $\mathbb{R}^{N+}$. This occurs in the unusual case in which x_j is the only nonzero pixel on a projection which has a nonzero photon count. In this case, $A_{i*}x = A_{ij}x_j$ and the log likelihood functions have terms of the form $\log x_j$ which tend to $-\infty$ as $x_j \to 0$. There are a number of possible strategies for handling this case numerically.

Strategy 1: A very simple method for handling this case is to set r_i to a very small number such as $r_i = \frac{1}{100M}$. This guarantees that the expected number of additional photons due to this adjustment summed over all projections is much less than 1. In practice, such a small perturbation to the model should not have a significant effect on the resulting reconstruction. This strategy also has the added benefit of making the algorithm more robust to floating point round-off error.

Strategy 2: Modify the algorithm so that in the case when $f_j(0) = -\infty$, the function is recomputed at $f_j(\epsilon)$ where ϵ is chosen to assure that $f_j(\epsilon) < 0$. One such choice is $\epsilon = \min_j\{\frac{1}{K_j}\}$, where $K_j = \sum_{i=1}^{M} A_{ij}$. In this case, θ_2 is given by

$$\theta_2^{(fs)} = \frac{f_j(x_j^n) - f_j(\epsilon)}{x_j^n(1-\epsilon)} = \sum_{i=1}^{M} \frac{A_{ij}^2 y_i}{p_i^{n2} - p_i^n A_{ij}(x_j^n - \epsilon)}$$

and the update equation is still given by

$$x_j^{n+1} = \arg\min_{x_j\geq 0}\{F_j^{(fs)}(x_j) + P(x_j)\}$$

We conjecture that this update strategy is globally convergent since it appears that the proof of [7] could be simply extended to handle this more general case.

4. NUMERICAL RESULTS

Our experimental results consist of trials using synthetic phantoms with functional values similar to human tissue in low dosage emission and transmission tomography simulations. The data are 128×128 projections, and the reconstruction is computed at a resolution of 128×128 pixels. We solve the MAP reconstructions consisting of two choices of q for the generalized Gaussian Markov random field (GGMRF) prior

[1] Continuity of $f_j^{(fs)}(t)$ as a function of (t, x) on $\mathbb{R}^{(N+1)+}$ also appears to be necessary in Condition 2.

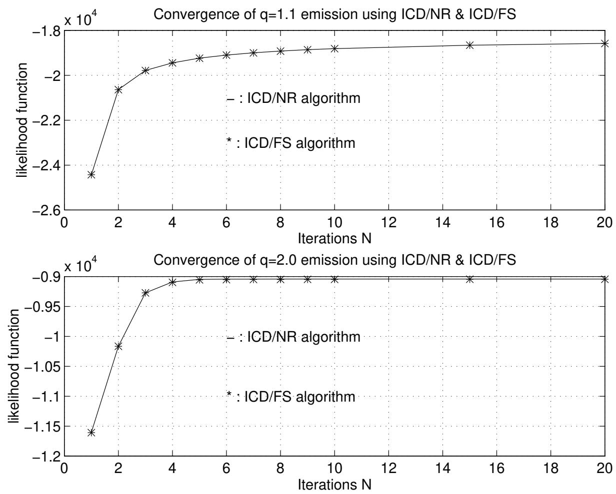

Figure 1: Convergence of likelihood vs. iterations for the emission case with $q = 1.1$ and $q = 2.0$ GGMRF prior models, employing ICD/FS and ICD/NR algorithms.

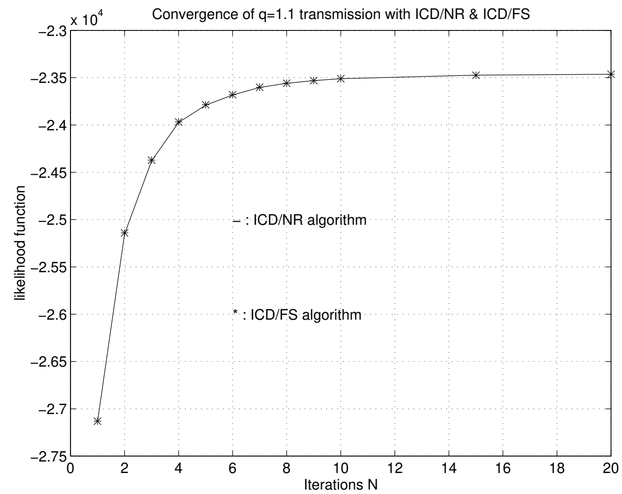

Figure 2: Convergence of likelihood vs. iterations for the transmission case with $q = 1.1$ GGMRF prior model, employing ICD/FS and ICD/NR algorithms.

model with the prior log density function of $P(x) = \sum_{i,j} \frac{b_{ij}}{q} \left(\frac{x_i - x_j}{\sigma} \right)^q$.

The experimental results of Figs. 1 and 2 show the convergence of the ICD/NR and ICD/FS for the emission case with a non-Gaussian prior, the emission case with a Gaussian prior, and the transmission case with a non-Gaussian prior. All trials employ sequential pixel updates in raster ordering. We note that with this particular non-Gaussian prior, the second derivative of the log likelihood is not bounded, so the technical conditions of the proof are not met. However, we conjecture that the proof of [7] may also be generalized for this case.

The plots of Figures 1 and 2 experimentally show that the convergence properties of ICD/NR and ICD/FS are virtually identical. This is not surprising, since the log likelihood function is generally close to quadratic and the values of $\theta_2^{(nr)}$ and $\theta_2^{(fs)}$ are therefore generally very close. While a proof for the global convergence of ICD/NR does not yet exist, its convergence appears consistently rapid.

5. REFERENCES

[1] L. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography," *IEEE Trans. Med. Im.,* vol. MI-1, no. 2, Oct. 1982.

[2] T. Hebert and R. Leahy, "A Generalized EM Algorithm for 3-D Bayesian Reconstruction from Poisson data Using Gibbs Priors," *IEEE Trans. Med. Im.*, vol. 8, no. 2, pp. 194-202, June 1989.

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

[4] C. Bouman, K. Sauer, "A Unified Approach to Statistical Tomography Using Coordinate Descent Optimization," *IEEE Trans. Image Proc.,* vol. 5, no. 3, pp. 480-492, March, 1996.

[5] K. Sauer, S. Borman and C. Bouman, "Parallel Computation of Sequential Pixel Updates in Statistical Tomographic Reconstruction," *1995 Proc. IEEE Intl. Conf. on Image Proc.,* Washington, DC, Oct. 22-25, 1995.

[6] J. Fessler, E. Ficaro, N. Clinthorne, and K. Lange "Fast Parallelizable Algorithm for Transmission Image Reconstruction," *1995 Proc. IEEE Nucl. Sci. Symp. & Med. Image Conf.,* San Francisco, CA. Oct. 21-28, 1995

[7] J. A. Fessler and A. O. Hero, "Penalized Maximum-Likelihood Image Reconstruction Using Space-Alternating Generalized EM Algorithms," *IEEE Trans. on Image Proc.*, vol. 4, no. 10, pp. 1417-1429, Oct. 1995.