Bayesian Estimation of Transmission Tomograms Using Segmentation Based Optimization †

Ken Sauer
Laboratory for Image and Signal Analysis
Department of Electrical Engineering, University of Notre Dame
Charles Bouman
School of Electrical Engineering, Purdue University

Abstract

We present a method for nondifferentiable optimization in MAP estimation of computed transmission tomograms. This problem arises in the application of a Markov random field image model with absolute value potential functions. Even though the required optimization is on a convex function, local optimization methods, which iteratively update pixel values, become trapped on the nondifferentiable edges of the function. We propose an algorithm which circumvents this problem, by updating connected groups of pixels formed in an intermediate segmentation step. Experimental results show that this approach substantially increases the rate of converge and the quality of the reconstruction.

1 Introduction

Image reconstruction from integral projections has been approached in a variety of ways, with the method of choice dependent on the quality and character of the data set[1, 2]. Although convolution backprojection (CBP) is the technique most frequently found in commercial applications of X-ray computed tomography with relatively high signal-to-noise ratio and complete data sets, special cases benefit from alternative algorithms. In many situations the quality and/or quantity of data is inadequate for conventional CBP reconstruction. These cases arise, for example, in low dosage medical imaging[3], nondestructive testing of materials with widely varying densities[4], and applications with limited angle projections[5] or hollow projections[6].

Statistical methods of image reconstruction and restoration seek the solution which best matches the probabilistic behavior of the data. Maximum-likelihood (ML) estimation selects the reconstruction which most closely matches the data available, but may yield solutions which do not have many of the properties expected in the original function[7]. Maximum *a posteriori* (MAP) estimation allows the introduction of a prior distribution which reflects knowledge or beliefs concerning the types of images acceptable as estimates of the original cross-section.

†This work was supported by an NEC Faculty Fellowship.

The Markov random field (MRF) has proven a useful image model in the tomographic setting, with relatively simple parameterization. MRF's have been applied extensively in emission tomography[8, 9, 10, 11], and to a lesser extent in the transmission problem[12]. The differences between the MRF models often centers on the choice of the *potential function*, which penalizes differences among neighboring pixels in the estimate. The central idea of these generalizations is the avoidance of the excessive penalties extracted by the Gaussian's quadratic potential function, which tends to blur edges due to the high cost of abrupt transitions.

The generalized Gaussian MRF (GGMRF)[12] uses a potential function similar to the log of the generalized Gaussian noise density found commonly in robust detection and estimation[13]. It renders edges accurately without prior knowledge of their size, and it results in a convex optimization problem with no local minima.

A particular case of the GGMRF, with the absolute value of local inter-pixel differences as the potential function, appears to produce superior reconstructions, but at the cost of greater computational difficulty in finding optimal solutions. When applied to the transmission tomography problem, this approach presents a nondifferentiable cost function for optimization. Since the gradient of nondifferentiable functions cannot be computed, methods such as gradient ascent and conjugate gradient can not be directly applied. Modifications of standard gradients have been successfully applied in some nondifferentiable optimizations[14], but would require very large amounts of computation in a high-dimensional case such as image estimation. Alternatively, Gauss-Seidel(GS) type iterations move toward the MAP estimate by sequentially updating individual pixels[15]. The GS method does not require the computation of gradients. However, it may become "stuck" on nondifferentiable edges at points far from the optimal estimate.

We present an adaptation of the GS method of MAP optimization which overcomes the convergence problems in the case of the absolute-value potential function. Our method allows the image estimate to move along hyperplanes where its cost function is not differentiable. This is accomplished by updating connected *groups* of pixels with

0018-9499/92$03.00 © 1992 IEEE

similar value. These connected groups are first formed by an intermediate segmentation step, after which each group is shifted in value to minimize cost. The updating of a group is performed efficiently by a generalization of the GS algorithm, and the segmentation process only requires a moderate computational overhead. Convergence is deterministic, and significantly faster than that of an adapted gradient technique.

2 MAP Tomographic Estimation

2.1 Statistical Model for Data

In transmission tomography the raw projection data are in the form of the number of photons detected after passing through an absorptive material. We will assume that the input photon counts are Poisson-distributed random variables that are known through calibration to have fixed mean λ_T. We will denote the vector of measured photon counts for each angle and displacement as λ. Under the usual assumptions of independence of photon counts among angles and displacements, the log likelihood function reduces to

$$\begin{aligned} L(\lambda|f) &= \log p(\lambda|f) \qquad (1) \\ &= -\sum_{j=1}^{M}[\lambda_T e^{-\mathbf{A}_{j*}f} + \lambda_j \mathbf{A}_{j*} f + \log(\lambda_j!)] \,, \end{aligned}$$

where $\mathbf{A}_{j*}$ is the j^{th} row of the projection matrix $\mathbf{A}$, and f is the function to be estimated. Using this form, the maximum likelihood (ML) estimate may be computed by maximizing (1).

We next introduce a quadratic approximation to the log likelihood function given in (1) formed by a second order Taylor expansion as a function of f[15]. This approximation provides a much more intuitive view of the log likelihood, and permits analytic and computational simplifications.

$$L(\lambda|f) \approx -\frac{\lambda_T}{2}(\hat{p} - \mathbf{A}f)^t \mathbf{D}(\hat{p} - \mathbf{A}f) + c(\lambda) \qquad (2)$$

where $\hat{p}_j$ and $\mathbf{D}$ are defined by

$$\begin{aligned} \hat{p}_j &= \log(\lambda_T/\lambda_j) \\ \mathbf{D} &= \mathrm{diag}\left\{\frac{\lambda_1}{\lambda_T}, \frac{\lambda_2}{\lambda_T}, \cdots, \frac{\lambda_M}{\lambda_T}\right\} . \end{aligned}$$

Since $c(\lambda)$ is independent of f, it may be ignored in the ML and MAP estimation of f. Note that since this approximation is not a quadratic function of λ_i, it is *not* equivalent to a Gaussian approximation of the photon counts. Moreover, it is not a Gaussian approximation for f since the distribution for f is yet unspecified. Therefore, this method is distinct from previous work which has made Gaussian approximations for the projections[16, 17].

The key to the behavior of (2) is found in the matrix $\mathbf{D}$. The matrix more heavily weights errors corresponding to projections with large values of λ_i. These projections pass through less dense objects, and consequently have higher signal-to-noise ratio. Fig. 1 compares the true and approximate log likelihood values as a function of a single projection $\mathbf{A}_{j*}f$. This is done for observed photon counts ranging from 10 to 1000 and a confidence interval of 99%. The plots indicate that the approximation is reasonable even at 10 observed photons.

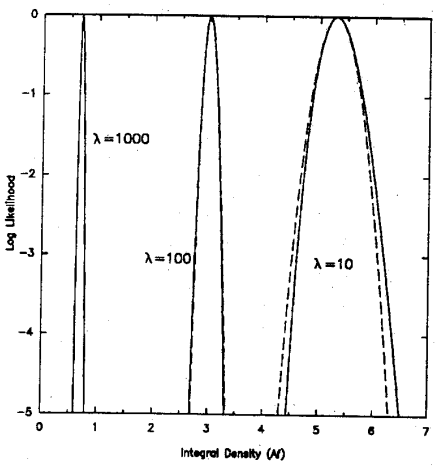


Figure 1: Plots of the log-likelihood as a function of a single projection across the reconstructed image. The exact Poisson model (dashed lines) and the quadratic approximation (solid lines) are each plotted for three values of λ. Each of these plots covers a confidence interval of 0.99 for a maximum-likelihood estimate of the projection value, and assumes $\lambda_T = 2000$.

In a practical settings, the input dosages, λ_T, may not be available. Often only the computed projection values, $\hat{p}$, are recorded, making the computation of the exact log likelihood impossible. However, the entries of the matrix $\mathbf{D}$ can be computed from the entries of $\hat{p}$. Thus if we are given a set of projection measurements $\hat{p}$, but do not know the input photon emission rate λ_T, the approximate log likelihood function can be computed within a constant. This makes the approximation essential for many real data sets.

2.2 Statistical Image Model

The maximum *a posteriori* (MAP) estimate is the value of $\hat{f}$ which maximizes the *a posteriori* density given the observations λ.

$$\hat{f} = \arg\max_{f \in \Omega} \{L(\lambda|f) + \log g(f)\} \,, \qquad (3)$$

where $g(f)$ is the prior distribution of the random field.

The Markov random field (MRF) model has proven very useful as a choice for $g(f)$ in image estimation

problems[18, 19]. The MRF is characterized by the localization of pixel interactions to a neighborhood. We will restrict our attention to simple MRF's based on at most a 8-point neighborhood.

A common choice for the prior model is a Gaussian Markov random field (GMRF). The log likelihood of the GMRF density function may always be written in the form

$$\log g(f) = -\sum_s a_s |f_s|^2 - \sum_{\{s,r\}\in U} b_{s,r}|f_s - f_r|^2 + c, \quad (4)$$

where a_s, $b_{s,r}$ and c are constants, s and r are pixel locations, and U is the set of all neighboring locations or cliques[20]. In addition, each squared difference of neighboring pixels is called a potential function of the distribution. Since (4) is a negative definite quadratic function, and $L(\lambda|f)$ is a concave function of f, the global maximum of (3) will be unique and feasible to compute. Unfortunately, the GMRF prior distribution results in estimates of f which are either excessively noisy or generally blurred[12]. This is because the squared difference potential functions in (4) excessively penalizes abrupt changes in f. The blurring effect is particularly undesirable along the edges that often occur in real tomograms.

In [12], we proposed a simple generalization of MRF's based on the concept of generalized Gaussian noise[13]. This model has the functional form

$$\log g(f) = -\gamma^q \left(\sum_s a_s |f_s|^q + \sum_{\{s,r\}\in U} b_{s,r}|f_s - f_r|^q \right) + c, \quad (5)$$

where $1 \leq q \leq 2$, and γ is a parameter which is inversely proportional to the scale of f. We call the class of random fields with this distribution generalized Gaussian Markov random fields (GGMRF) since this model is contained within the more general class of MRF's and includes all Gaussian MRF's when $q = 2$. In practice, we may choose $a_s = 0$, which results in an ill-defined density. However, this is not a practical problem since the function $L(\lambda|f)$ causes the MAP estimation problem to be unique.

The GGMRF is unique among proposed MRF's because it has three desirable properties. It allows the sharp reconstruction of edges; it results in a convex log likelihood, and it does not require the choice of a particular threshold corresponding to the expected edge size[12].

Larger values of q discourage abrupt discontinuities while smaller values of q allow them. The particular case of $q = 1$ will be the focus of this paper. To gain insight, consider the example shown in Fig. 2 of a one-dimensional function f with $q = 1$, $a_s = 0$ and $b_{s,r} = 1$. As long as f is a monotone (increasing or decreasing) function, then

$$\log g(x) = \sum_{i=1}^{N} |f_i - f_j| + c = |f_1 - f_N| + c\ .$$

Therefore, any monotone function has the same cost, and reconstructions with abrupt edges are not discouraged.

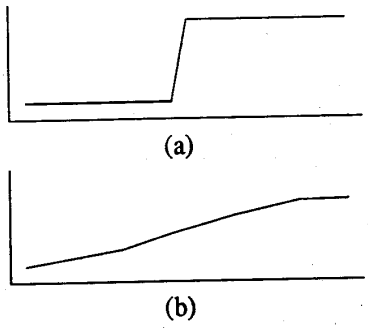


Figure 2: When $q = 1$, any monotone function which starts and ends at the same values has the same total cost. Therefore, both the sharp edge and the smooth edge have the same cost.

Since tomograms often contain abrupt transitions between densities, the GGMRF with $q = 1$ is an appropriate model. Using (2), the MAP estimate of f is then given by the minimum of the cost function

$$C(f) = (\hat{p} - \mathbf{A}f)^t \mathbf{D}(\hat{p} - \mathbf{A}f) + \frac{2\gamma}{\lambda_T} \sum_{\{s,r\}\in U} |f_s - f_r|. \quad (6)$$

The choice of the parameter γ may be made from ensembles of typical images. However, in practice the input rate λ_T may also be unknown. Since the MAP estimate is determined by the ratio $\frac{2\gamma}{\lambda_T}$, only a single constant must be determined experimentally.

2.3 Optimization

The potential function $C(f)$ of (6) is convex, and therefore contains no local minima. However, the derivative of $C(f)$ does not exist whenever $f_s = f_r$ where s and r are neighbors. This makes optimization difficult since gradient methods (i.e. gradient accent, conjugate gradient) are not directly applicable at these points where the gradient does not exist.

Nondifferentiable optimization problems have been widely studied[21, 22], and many algorithms hinge on the concept of a *subgradient*. The subgradient of the function $C(f)$ at f_0 is defined as any vector $\phi(f_0)$ with the property that

$$C(f) \geq C(f_0) + \phi(f_0)^T (f - f_0)\ \forall\ f.$$

Since $C(f)$ is formed by the sum of convex functions, its subgradient may be computed as the sum of its component subgradients[23]. In particular, we compute the subgradient by differentiating using the convention that

$$\frac{d|f_s - f_r|}{df_s} = \begin{cases} 1 & f_s > f_r \\ 0 & f_s = f_r \\ -1 & f_s < f_r \end{cases}.$$

Unfortunately, an incremental step in the direction opposite a subgradient may cause the functional $C(f)$ to

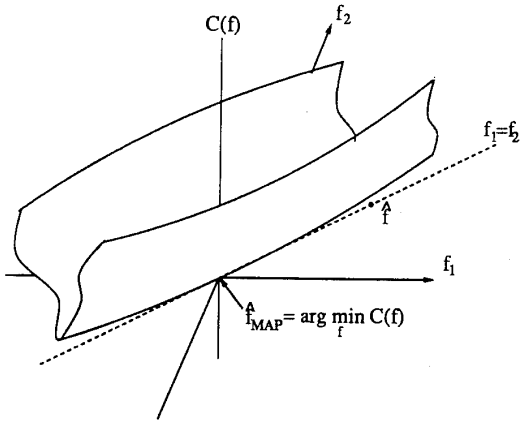


Figure 3: Negative *a posteriori* log density function of two pixels, f_1 and f_2, when $q = 1$ in the GGMRF, and quadratic approximation is used for the Poisson in f. The MAP estimate of f is at f_{MAP}, but the current estimate $\hat{f}$ is a local optimum with respect to either f_1 or f_2.

increase[23]. Using a simply computed $\phi(f)$ in an iterative MAP estimation process can lead to very slow convergence. These results can be improved in terms of iteration counts by choosing the subgradient yielding the direction of steepest descent[14]. But making this choice, rather than the simplest $\phi(f)$, has complexity $\mathcal{O}(n^2)$, which is unreasonably expensive for problems of the dimensions of interest here.

In [15], we introduced a Gauss-Seidel (GS) type approach to produce MAP estimates of transmission tomograms. The GS technique scans the current estimate of f, choosing each pixel's value to minimize $C(f)$, with the remainder of the image fixed. The computational expense of a single pass through the image is approximately the same as one iteration of gradient ascent or conjugate gradient. It has also been shown analytically that with a Gaussian prior distribution, GS iterations have relatively fast convergence[15].

GS can quite easily be applied to optimize nondifferentiable functions such as $C(f)$, since the computation of the current optimum with respect to a single pixel remains relatively simple. However, this algorithm is also limited in its power to find the MAP image. The problem faced by GS and others methods is illustrated in Fig. 3. In the "trough" of this cost function, which corresponds to the state with $f_1 = f_2$, we can see that moving toward f_{MAP} reduces $C(f)$. But optimization with respect to *either* f_1 *or* f_2 individually results in no change of state.

3 Segmentation-Based Optimization

The derivative of $C(f)$ in (6) does not exist whenever f contains neighboring pixels of equal value. When the prior distribution is strong, solutions tend to fall on these edges of the cost function, forming large groups of pixels with equal value. If the function to be estimated is known to have piecewise constant regions, as is often the case in nondestructive testing applications, this may be quite desirable. But when multiple pixels cluster at a single density value, the example in Fig. 3 is expanded to high-dimensional edges, which makes the cost function nondifferentiable with respect to many variables.

Our approach to this problem can be understood by viewing Fig. 3 and doing what seems obvious on inspection of the surface of $C(f)$. From $\hat{f}$, we wish to follow the valley $f_1 = f_2$ to the minimum along that line. This can be accomplished by changing the value $\alpha = f_1 = f_2$ until a minimum of $C(f)$ is achieved.

Higher-dimensional cases are similar, though not quite so simple. Edges of $C(f)$ correspond to groups of neighboring pixels with the same value, $f_{s_1} = f_{s_2} = \ldots = f_{s_K}$, where $S = \{s_1, s_2, \ldots, s_K\}$ is the set of pixel locations. We say that S is a connected set since each pixel is a neighbor of some other pixel in S. In fact, each group S is the largest connected set of pixels with equal value. If a multiple-pixel spatial cluster S has acquired a common value, the apparent generalization is to change all the pixels in S together. Specifically, we move along the hyper-plane $\alpha_S = f_s$ for all $s \in S$, seeking the minimum of $C(f)$ as a function of the scalar α_S.

The first step in implementing this segmentation-based optimization (SBO) is to segment f into connected regions of constant value. This can be done with only a few simple operations per pixel using well known algorithms[24]. Next, each segment is visited in sequence, and each segments value is adjusted to minimize $C(f)$ while fixing the remaining values in f.

While the segment-based approach accelerates motion down the edge of the cost function, it does not guarantee convergence to the global minimum. By locking the values of multiple pixels together, and looking only along resulting hyperplanes in image space, we sacrifice consideration of many alternative directions corresponding to other edges of the likelihood function. We ameliorate the problem by alternating the segmentation based optimization with the standard GS pixel updates. During the GS operations, each pixel is freed to seek its own conditionally optimal value.

The algorithm for efficiently updating the value of each segment is closely related to the algorithm used for basic GS updates. For computational efficiency, the following projection error must be stored at the k^{th} iteration.

$$e^k = \hat{p} - \mathbf{A}f^k$$

At update k, the value of segment S is changed by the amount Δ_S^k.

$$\alpha_S^{k+1} = \alpha_S^k + \Delta_S^k$$

If we define, ∂S to be all the neighboring pixels of S, then the cost function of (6) may be written to emphasize the

dependence on Δ_S^k.

$$C(f^{k+1}) = \sum_{i=1}^{M} \frac{\lambda_i}{\lambda_T} \left[e_i^k - \Delta_S^k \sum_{j \in S} A_{ij} \right]^2 + \frac{2\gamma}{\lambda_T} \sum_{r \in \partial S} |f_r - \alpha_S^k - \Delta_S^k| \,. \quad (7)$$

The the updated value for the segment, α_S^{k+1} may be found by minimizing the cost as a function of Δ_S^k. Suppose we put the P neighbors of segment S into order of increasing pixel value $s_{(1)}, s_{(2)}, \cdots, s_{(P)}$. If the segment's final value lies between the values of the $p-th$ and $(p+1)-th$ neighbors of S, then the derivative of the prior distribution term in (7) is $(2\gamma/\lambda_T)[P-2p]$. In this case, optimizing (7) with respect to Δ_S^k yields the following result.

$$\hat{\Delta} = \frac{\sum_{i=1}^{M} \frac{\lambda_i}{\lambda_T} \left[e_i^k \sum_{j \in S} A_{ij} \right] + \frac{\gamma}{\lambda_T}[P-2p]}{\sum_{i=1}^{M} \frac{\lambda_i}{\lambda_T} (\sum_{j \in S} A_{ij})^2} \quad (8)$$

Because $C(f)$ is convex, if $\hat{\Delta} + \alpha_S^k$ falls in the interval between $f_{s_{(p)}}$ and $f_{s_{(p+1)}}$, then $\Delta_S^k = \hat{\Delta}$ is in fact the optimum.

Therefore, to compute α_S^{k+1}, we scan through the ordered values of neighbors. For each p, if the value $\hat{\Delta} + \alpha_S^k$ is between $f_{s_{(p)}}$ and $f_{s_{(p+1)}}$, then the result is optimum. If $\hat{\Delta} + \alpha_S^k$ is greater than $f_{s_{(p+1)}}$, then increment p and continue. If $\hat{\Delta} + \alpha_S^k$ is less than $f_{s_{(p)}}$, the optimum is $\Delta_S^k = f_{s_{(p)}} - \alpha_S^k$.

Once the optimum value of α^{k+1} is computed, the projection error must be updated using the equation

$$e_i^{k+1} = e_i^k - \Delta_S^k \sum_{j \in S} \mathbf{A}_{i,j}$$

3.1 Computational Costs

To compare relative computational cost of several methods, we consider the number of multiplies required for one complete iteration. Table 1 summarizes the results. If we assume that no computation is required in evaluating the elements $\mathbf{A}_{ji}$ (look-up tables may be used), then the number of multiplies required to compute $\mathbf{A}f$ is $M_0 N$, where M_0 is the average number of projections which pass through a single pixel, and N is the number of pixels.

The computational cost of the segmentation based algorithm will depend on the size of the segments. Generally, smaller segments will require more computation. Therefore, we bound the computation using the assumption that each pixel forms its own segment. In practice, this is a conservative assumption. In [15] we computed the number of multiplies and divides required for one GS update. Given our assumption of single pixel segments, segmentation based optimization requires the same number of multiplies and divides as GS for updating the error and evaluating the numerator of (8). However, the denominator of (8) requires an additional $2NM_0$ multiplies since it cannot be precomputed.

Convolution Backprojection	$M_0 N + N_h M$
Gradient Ascent	$2M_0 N + M$
Gauss-Seidel	$(3M_0 + 2)N$
Segmentation-Based Opt.	$(5M_0 + 2)N$

Table 1: Number of multiplies and divides required for one iteration of each method. N is the number of points in the image, M is the number of projections, M_0 is the average number of projections passing through a pixel in the image, N_h is the number of points in the CBP filter.

Therefore, an iteration of segmentation based optimization followed by GS requires approximately $8NM_0$ multiplies, and an iteration of gradient ascent followed by GS requires approximately $5NM_0$ multiplies.

4 Experimental Results

We have applied the models and optimization methods discussed in this paper to problems of tomographic reconstruction under limited dosage X-rays. In these cases, high noise levels in those rays passing through very dense regions of the object under study yield artifacts which affect the entire reconstruction. The initial state for the Bayesian estimation iterations is a convolution backprojection (CBP) reconstruction, using a raised-cosine rolloff filter. The cutoff frequency was chosen for the best visually evaluated reconstruction. Because it is physically realistic to assume non-negative densities for our experiments, we include this constraint in our priors. Modification of the SBO and GS iterations for the strictly positive case is trivial, and involves no appreciable increase in complexity. The subgradient algorithms were modified by simply following each step of gradient ascent with a pixel-by-pixel limiting operation, which adds slightly to projection error update costs.

The first data set was taken from the phantom in Fig. 4(a), which consists of four discs of density $0.48cm^{-1}$ in a circular background of diameter 20 cm and density $0.2cm^{-1}$. The simulated dosage was a rate of 2000 photons per ray (λ_T). At these dosages, the rays passing through the most dense regions are essentially blocked, making this reconstruction similar to a hollow projections problem. The 128 angles were equally spaced, and 128 rays were collected at each angle.

Fig. 4(b) shows the CBP reconstruction from filtered backprojection, using a raised cosine rolloff filter. Though only the denser areas are too absorptive to render a good reconstruction, the entire image is of poor quality. We apply the GGRMF as a prior, with $q = 1$, which produces the nondifferentiable cost function. Fig. 4(c) is the result of applying 20 GS iterations directly to the MAP estimation. The reconstruction is much improved, and has essentially converged after only about 6 iterations. But

we cannot guarantee that GS can reach the MAP estimate, and in fact Fig. 4(c) represents the case of the estimate's begin "stuck" at a nondifferentiable point. Figure 4(d) illustrates the effects of using alternating segmentation with the GS iterations. This image is the result of 50 iterations, but was not appreciably changed after 6 iterations. Though the difference in convergence rates and final likelihood value between Fig. 4(c) and (d) is small, image quality is significantly improved in the latter.

Iterations using subgradients alternating with GS may also permit improved convergence relative to either of the two independently. Fig. 7 compares convergence rates for the SBO algorithm, subgradient descent, and alternating GS/subgradient. Unassisted subgradient is much slower than either of the other methods. We observed this behavior over a wide range of values for the step size in the gradient direction. Two plots appear for the GS/subgradient, representing two choices for the step size in the subgradient portion. With step size of 5×10^{-5}, the estimates oscillate for many iterations after nearing steady-state, and do not reach the level of log posteriori probability of the SBO, even after as many as 100 iterations. Convergence is improved somewhat in the other GS/subgradient plot, with step size 5×10^{-6}, but the image resulting from this setting is poorer than obtained by either SBO or GS/subgradient. The SBO algorithm demonstrates superior convergence and visual results in this case.

The projection data used as input for the reconstructions in Figures 5 and 6 were taken from a first-generation gamma-ray scanner at Lawrence Livermore National Laboratories. The object is an approximately cubic concrete block, 20 cm on each side, with four steel reinforcing rods perpendicular to the plane in which this data was collected. In the reconstruction of Fig. 5(a), the relatively low dosage/density ratio yields apparent photon counting noise artifacts throughout the image. Performance may be improved with higher photon counts, but this would require increased time in data collection.

Neither an original of this cross-section, nor the input dosage rates are available with this data. But, as discussed above, we need only determine the ratio of λ_T to γ for our estimation method. We present outcomes for three cases, spanning two orders of magnitude in this ratio. With $\lambda_T/\gamma = 2 \times 10^4$, the estimate is moderately smoothed, but some apparent noise artifacts remain, while in Fig. 6(b), the estimate is smoother than would be expected of material in this case. Fig. 6(a) results from $\lambda_T/\gamma = 2000$, and shows texture approximately appropriate for the material, with good preservation of detail, most of which can be seen with a bit more effort in Fig. 5(a). Note the greater smoothing which occurs near the center of the block, where the relative strength of the prior density's influence is greater. This is in keeping with the greater statistical uncertainty in this region, due to the lower signal-to-noise ratio of rays passing through these regions.

Convergence toward the reconstruction of Fig. 6(a) by the three optimization techniques appears in Fig. 8. Again, the SBO method converges most quickly, with GS/subgradient similar in terms of this quantitative measure. The appearance of the SBO reconstruction reaches a state visually equivalent to the final result far earlier than the latter approach, and is slightly better even after 50 iterations.

5 Conclusion

The segmentation-based optimization approach presented here to seek the minimum of a nondifferentiable cost function appears to perform well in Bayesian tomographic reconstruction. In addition to furnishing better results, its monotonic convergence behavior gives us greater certainty concerning the nearness to optimality of our results. Subgradient descent appears impractically slow, and the inclusion of the subgradient between Gauss-Seidel iterations may also fail to reach the state of which SBO is capable. SBO converges well, and has moderate computational costs.

Acknowledgment

The authors thank Dr. Steve Azevedo and Lawrence Livermore National Laboratories for supplying data for this research.

References

[1] G.T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography,* Academic Press, New York, 1980.

[2] R.M. Ranggayyan, A.T. Dhawan, and R. Gordon, "Algorithms for Limited-View Computed Tomography: An Annotated Bibliography and a Challenge," *Applied Optics,* vol. 24, no. 23, pp. 4000-4012, Dec. 1985.

[3] G.C. McKinnon and R.H.T. Bates, "Towards Imaging the Beating Heart Usefully with a Conventional CT Scanner," *IEEE Trans. Biomed. Eng.,* vol. BME-28, pp. 123-127, 1981.

[4] J.G. Sanderson, "Reconstruction of Fuel Pin Bundles by a Maximum Entropy Method," *IEEE Trans. Nucl. Sci.,* Vol. NS-26, pp. 2685-2688, 1979.

[5] T. Inouye, "Image Reconstruction with Limited Angle Projection Data," *IEEE Trans. Nucl. Sci.,* Vol. NS-26, pp. 2666-2684, 1979.

[6] G.H. Glover and N.J. Pelc, "An Algorithm for the Reduction of Metal Clip Artifacts in CT Reconstruction," *Med. Phys.,* vol. 8, pp. 799-807, 1981.

[7] D.L. Snyder and M.T. Miller, "The Use of Sieves to Stabilize Images Produced with the EM Algorithm for Emission Tomography," *IEEE Trans. Nucl. Sci.*, vol. NS-32, pp. 3864-3872, 1985.

[8] S. Geman and D. McClure, "Bayesian Image Analysis: An Application to Single Photon Emission Tomography," in *Proc. Statist. Comput. Sect. Amer. Stat. Assoc.*, Washington, DC, pp. 12-18, 1985.

[9] S. Geman and D. McClure, "Statistical Methods for Tomographic Image Reconstruction," *Bull. Int. Stat. Inst.*, vol. LII-4, pp. 5-21, 1987.

[10] 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.

[11] P. J. Green "Bayesian Reconstructions from Emission Tomography Data Using a Modified EM Algorithm,: *IEEE Trans. Med. Im.*, vol. 9, no. 1, pp. 84-93, March 1990.

[12] C. Bouman and K. Sauer, "An Edge-Preserving Method for Image Reconstruction from Integral Projections," *Proc. Conf. on Info. Sci. and Syst.*, The Johns Hopkins University, Baltimore, MD, March 20-22, 1991, pp. 382-387.

[13] S.A. Kassam, *Signal Detection in Non-Gaussian Noise*, Springer-Verlag, New York, 1988.

[14] P. Wolfe, "A Method of Conjugate Gradients for Minimizing Nondifferentiable Functions," in *Nondifferentiable Optimization*, M.L. Balanski and P. Wolfe, eds., North-Holland Publishing, Amsterdam.

[15] K. Sauer and C. Bouman, "A Local Update Strategy for Iterative Reconstruction from Projections," submitted to the *IEEE Trans. Sig. Proc.*.

[16] G.T. Herman, H. Hurwitz, A. Lent, and H-P. Lung, "On the Bayesian Approach to Image Reconstruction," *Info. and Cont.*, vol. 42, pp. 60-71, 1979.

[17] K.M. Hanson and G.W. Wecksung, "Bayesian Approach to Limited-Angle Reconstruction in Computed Tomography," *J. Opt. Soc. Am.*, vol. 73, no. 11, pp. 1501-1509, Nov. 1983.

[18] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. PAMI-6, no.6, pp. 721-741, Nov. 1984.

[19] J. Besag, "On the Statistical Analysis of Dirty Pictures," *J. Roy. Statist. Soc. B*, vol. 48, no. 3, pp. 259-302, 1986.

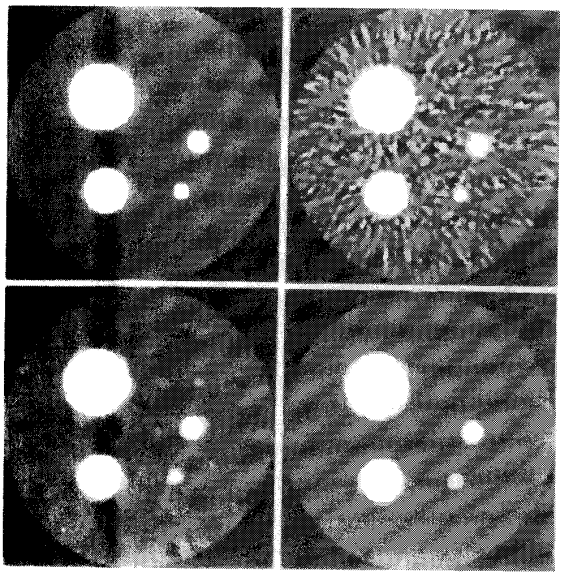
[20] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *J. Royal Stat. Soc. B*, vol. 36, pp. 192-326, 1974.

[21] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.

[22] D.P. Bertsekas and S.K.. Mitter, "A Descent Method for Optimization Problems with Nondifferentiable cost Functionals," *SIAM J. Control*, vol. 11, no. 4, pp. 637-652, Nov. 1973.

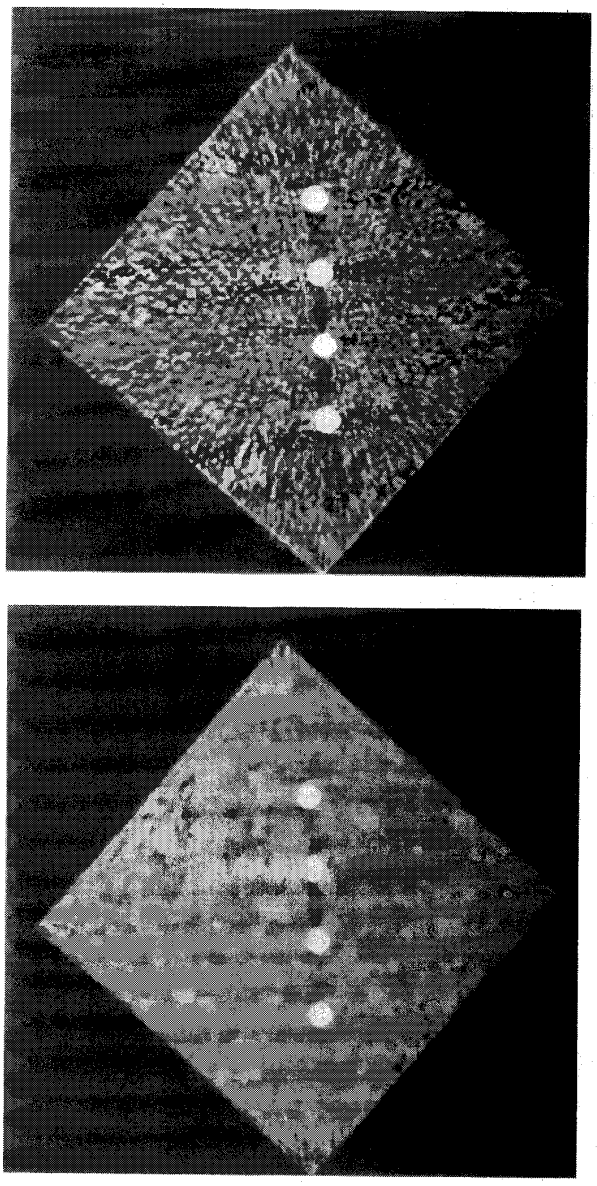
[23] S.P. Boyd and C.H. Barrett, *Linear Controller Design-Limits of Performance*, Prentice-Hall, Englewood Cliffs, New Jersey, 1991.

[24] A. Rosenfeld and A. Kak, *Digital Picture Processing*, vol. 2, Academic Press, San Diego, 1982.



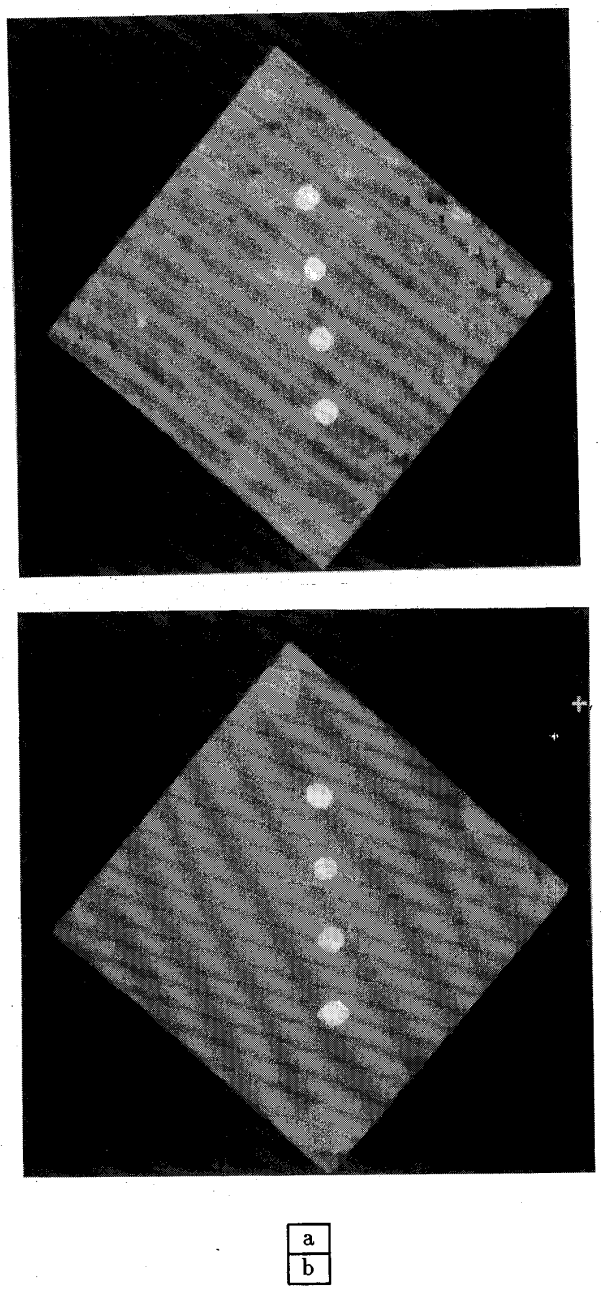
a	b
c	d

Figure 4: (a) Original phantom, diameter 20cm. Lighter areas are of density 0.48cm^{-1}, and background is 0.2cm^{-1}. (b) Convolution backprojection reconstruction from 128 raysums at each of 128 equally spaced angles. Emitter rate is 2000 photons/ray. (c) Result of Bayesian estimation using only local minimization with absolute value prior density. (d) MAP estimate computed by proposed method, with absolute value prior and $\gamma = 20$.



a
b

Figure 5: (a) Convolution backprojection reconstruction of concrete block with steel reinforcing rods from first-generation gamma-ray scanner. (Data courtesy of Lawrence Livermore National Laboratory.) (b) MAP estimate computed by 50 iterations of proposed algorithm with $\lambda_T/\gamma = 2 \times 10^4$.



a

b

Figure 6: MAP estimate from 50 iterations of proposed algorithm with $\lambda_T/\gamma = 2000$(a) and $\lambda_T/\gamma = 200$(b).

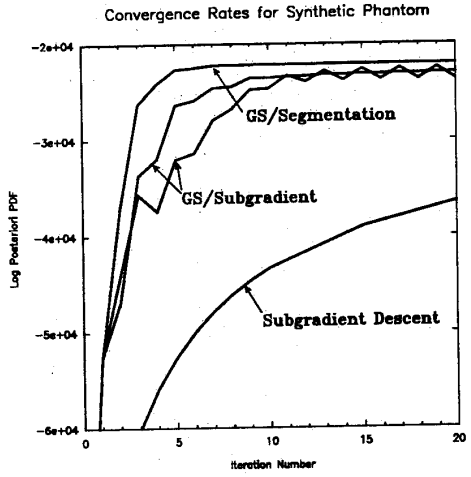


Figure 7: Convergence comparisons for synthetic phantom in Figure 4 with $\gamma = 20$. Two plots for GS alternating with subgradient interations have two values for step size for gradient: 5×10^{-6} and 5×10^{-5}. Oscillating plot corresponds to scaling of 5×10^{-5}.

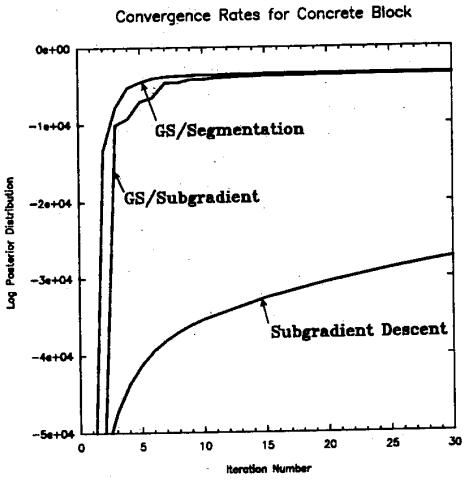


Figure 8: Convergence for concrete block data, with $\lambda_T/\gamma = 2000$.