ICASSP M6.6, pp. 2593-2596, April 14-17, 1991

M6.6

Bayesian Estimation from Projections with Low Photon Dosages

Ken Sauer
Department of Electrical Engineering
University of Notre Dame, Notre Dame, IN 46556

Charles Bouman
School of Electrical Engineering
Purdue University, West Lafayette, IN 47907

Abstract

We present a method for Bayesian reconstruction from projections which updates single pixel values, rather than the entire image, at each step. The technique is similar to Gauss-Seidel (GS) iteration for the solution of differential equations on finite grids. The computational cost per iteration of the GS approach is found to be approximately equal to that of gradient methods. For continuously valued images, GS is found to have significantly better convergence at modes representing high spatial frequencies. In addition, GS is well-suited to segmentation when the image is constrained to be discrete-valued.

1 Introduction

Although convolution backprojection (CBP)[1] is the most common technique for computed tomography, in many situations the quality and/or quantity of data is inadequate for conventional CBP reconstruction[2]. Bayesian estimation allows weightings or hard constraints on solutions which reflect knowledge concerning acceptable estimates. In the most previous work on the problem, the overall cost function specified by the log likelihood for either maximum likelihood (ML) or Bayesian techniques has been assumed quadratic for tomographic image reconstruction(see ref. in [2]).

In this paper, we develop a local update method to perform a Bayesian segmentation from projections. Our approach is similar to Gauss-Seidel(GS) iterations employed in the solution of differential equations on finite grids[3], and we will use the same terminology. Each step includes only the optimization with respect to a single pixel's value, making its application very simple if we follow a monotone increase in likelihood. Reconstruction of cross-sections which are known to consist of a few distinct densities is possible with very low SNR when formulated as a segmentation task for GS. The cost functions in this problem are generally nonconvex and nondifferentiable, and therefore not amenable to gradient approaches, which rely on partial derivatives of the cost.

Bayesian estimation by GS is also applicable to the reconstruction of continuously-valued functions. The computational cost of a single pass of local updates across the entire image is approximately the same as that of a single iteration of the gradient approach. In addition the GS approach has faster convergence in high spatial frequencies for continuously-valued images than either simple gradient ascent(GA), or conjugate gradient(CG).

2 Model of Physical System

In practice, reconstruction requires finite-dimensional representation of both the projection data, p, and the modeled image, f. The Radon transform equations may be written in the discrete form

$$p = \mathbf{A}f$$

where $\mathbf{A}$ is a sparse MxN matrix with A_{ji} equal to the length of the intersection of projection ray j and pixel i.

In transmission tomography the projections, p, are not measured directly. Instead, raw data are in the form of the number of photons(λ) detected after passing through an absorptive material. In order to simplify theoretical analysis, we will also introduce a quadratic approximation to the log likelihood of photon counts given the image f, resulting from a series expansion of the exact log likelihood:

$$L(\lambda|f) \approx -\frac{1}{2}(\hat{p} - \mathbf{A}f)^t\mathbf{D}(\hat{p} - \mathbf{A}f) + c(\lambda), \qquad (1)$$

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

$$\begin{aligned} \hat{p}_i &= \log(\lambda_T/\lambda_i) \\ \mathbf{D} &= \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_M\} \end{aligned}$$

for input photon count λ_T. The key to the behavior of (1) 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. In the limit of opaque projections where no photons pass through the material, the approximation simply applies no weight to the measurement.

The maximum *a posteriori* (MAP) estimate under prior distribution $g(f)$ is the value which maximizes the *a posteriori* density given the observations Λ:

$$\hat{F} = \arg\max_{f\in\Omega}\{L(\Lambda|f) + \log g(f)\} \qquad (2)$$

Here, Ω is the set of feasible solutions, and upper case letters denote random entities. The ML estimate may be derived as the special case where the prior distribution is constant.

When F is continuously valued, we will assume that it is a Gaussian random vector, yielding the MAP estimation criterion:

$$\hat{f} = \arg\max_{f\in\Omega}\left\{-(\hat{p} - \mathbf{A}f)^t\mathbf{D}(\hat{p} - \mathbf{A}f) - \gamma f^t\mathbf{R}f\right\} \qquad (3)$$

where $\mathbf{R}$ is symmetric, and $\frac{1}{\gamma}\mathbf{R}^{-1}$ is the covariance matrix for F. We will assume that, ignoring boundary effects, multiplication by $\mathbf{R}$ has the effect of convolution with a stationary kernel. When $\Omega = I\!R^N$, the solution to this optimization problem may be found by differentiating with respect to f. The resulting equation has the form

$$b = (\mathbf{H} + \gamma\mathbf{R})\hat{f} \qquad (4)$$

where $\mathbf{H} = \mathbf{A}^t\mathbf{D}\mathbf{A}$ and $b = \mathbf{A}^t\mathbf{D}\hat{p}$.

A second distinct case results when Ω is a proper subset of $I\!R^N$ due to hard constraints on $\hat{f}$. If Ω is a convex set, there

CH2977-7/91/0000-2593 $1.00 © 1991 IEEE

is a unique global maximum to (3). The problem of enforcing hard constraints becomes more difficult when the feasible set is not convex. This occurs if each pixel is assumed to have one of a fixed number of densities. Reconstruction then corresponds to labeling each pixel with the most appropriate of the discrete densities.

For this case, the discrete-valued Markov random field (MRF) has the advantage of being general and restricting interaction to be local. Discrete MRF's have been used to model the behavior of pixel labels in a variety of segmentation problems[5]. While the theory of MRF's is quite extensive, we will restrict our attention to a simple model based on a 8-point neighborhood. To define this model, we first define two simple functions: $t_1(f)$ is the number of horizontally and vertically neighboring points of different value in f, and $t_2(f)$ is the number of diagonally neighboring points of different value in f. The density function for f is then assumed to be of the form

$$\log g(f) = -\gamma\left(t_1(f) + \frac{t_2(f)}{\sqrt{2}}\right) + \log(Z) \qquad (5)$$

where Z is a constant.

3 Optimization Techniques

3.1 Gradient Ascent

We will first consider gradient ascent, an iterative scheme whose update equation may be written in the form of a standard discrete time system.

$$\begin{aligned} f^{(n+1)} &= \left[I - \alpha(\mathbf{A}^t\mathbf{D}\mathbf{A} + \gamma\mathbf{R})\right] f^{(n)} + \alpha\mathbf{A}^t\mathbf{D}\tilde{p} \qquad (6) \\ &= \left[I - \alpha(\mathbf{H} + \gamma\mathbf{R})\right] f^{(n)} + \alpha b \end{aligned}$$

Each iteration of (6) requires the computation of a projection, a backprojection and multiplication by the matrices $\mathbf{D}$ and $\mathbf{R}$.

Convergence analysis of (6) is difficult in the general case. We will gain insight into the problem by analyzing the convergence when $\mathbf{D} = \sigma^{-2}\mathbf{I}$ (i.e. the variance of each projection is equal). Because the matrix $\mathbf{H} = \sigma^{-2}\mathbf{A}^t\mathbf{A}$ is approximately Toeplitz-block-Toeplitz[4], each row of $\mathbf{H}$ represents a spatially shifted version of a single kernel. Therefore, multiplication of $f^{(n)}$ by $\mathbf{H}$ is approximately equivalent to convolving the function $f^{(n)}$ with a 2-D kernel, and may be approximated by multiplication in the spatial frequency domain. We take the 2-D DTFT of (6) to yield

$$f^{(n+1)}(\bar{\omega}) = \left[1 - \alpha(H(\bar{\omega}) + \gamma R(\bar{\omega}))\right] f^{(n)}(\bar{\omega}) + \alpha b(\bar{\omega}), \qquad (7)$$

where $H(\bar{\omega})$ is the 2-D DTFT of the spatially reversed kernel for $\mathbf{H}$, and $R(\bar{\omega})$ is that of the smoothing kernel of the prior. The spatial frequency components in f corresponding to frequency $\bar{\omega}$ will be attenuated or amplified by the factor $[1 - \alpha(H(\bar{\omega}) + \gamma R(\bar{\omega}))]$ with each iteration.

The kernel of $\mathbf{H}$ is approximately proportional to a filtered and sampled version of the function c/d where d is the distance of a pixel from the origin and c is a constant[4]. Since the Fourier transform of $1/d$ is $1/\|\bar{\omega}\|$, the rate of convergence at frequency $\bar{\omega}$ will be approximately

$$1 - \alpha\left(c/\|\bar{\omega}\| + \gamma R(\bar{\omega})\right).$$

In practice the value of $\mathbf{H}(\bar{0})$, and therefore the low frequency eigenvalues, are limited by the size of the reconstruction. To insure convergence, α must satisfy

$$0 < \alpha < 2\left(\max_{\bar{\omega}}\{H(\bar{\omega}) + \gamma R(\bar{\omega})\}\right)^{-1}$$

Since for large windows the value of $\max_{\bar{\omega}} H(\bar{\omega})$ is large, a small value of α will be needed to insure stability and convergence. But small values of α will cause the convergence to be slow for large values of $\|\bar{\omega}\|$. The conclusion is that the convergence of GA is poor at high spatial frequencies.

3.2 Gauss-Seidel

Gauss-Seidel iteration is a well-known technique for solving partial differential equations (PDE) using finite element methods[3]. The GS method can solve a wide variety of problems that incorporate hard constraints on Ω. This is particularly important in the segmentation problem since the gradient based algorithms implicitly require that Ω be a continuous convex set.

We pick the value of a single pixel, f_i, to minimize the posterior log likelihood function of (3). This can be done by setting the derivative with respect to f_i equal to zero.

$$0 = b_i - (\mathbf{H} + \gamma\mathbf{R})_{i*} f$$

The notation $(\mathbf{H} + \gamma\mathbf{R})_{i*}$ denotes the i^{th} row of the matrix. We may use this single equation to solve for the best value of f_i in terms of the remaining values of f. By separating the the matrix $\mathbf{H} + \gamma\mathbf{R}$ into its lower triangular, diagonal and upper triangular parts ($\mathbf{L}$, $\mathbf{K}$ and $\mathbf{U}$, respectively), we can derive a matrix expression for the sequential update. Assuming that the points are replaced in the order of their index i, the update equation for all the pixels has the form

$$f^{(n+1)} = -(\mathbf{K} + \mathbf{L})^{-1}\mathbf{U} f^{(n)} + (\mathbf{K} + \mathbf{L})^{-1} b\,. \qquad (8)$$

This update equation has the same form as (6) the update equation used in the GA. The convergence behavior of this method will depend on the eigenvalues of the update matrix, $(\mathbf{K}+\mathbf{L})^{-1}\mathbf{U}$.

Since $\mathbf{H} = \mathbf{H}^t$ and $\mathbf{R} = \mathbf{R}^t$, we also have that $\mathbf{L} = \mathbf{U}^t$. Using the Toeplitz-block-Toeplitz property of $\mathbf{L}$ and $\mathbf{K}$, we see that (8) has the interpretation of linear filtering, yielding the DTFT

$$f(\bar{\omega}) = \frac{-L^*(\bar{\omega})}{K_0 + L(\bar{\omega})} f(\bar{\omega}) + \frac{-1}{K_0 + L(\bar{\omega})} b(\bar{\omega}) \qquad (9)$$

where $L(\bar{\omega})$ is the DTFT of the spatially reversed kernel for $\mathbf{L}$, and $U(\bar{\omega}) = L^*(\bar{\omega})$ by the time reversing property of the DTFT.

We may draw conclusions about the properties of the convergence rate

$$P(\bar{\omega}) = \frac{-L^*(\bar{\omega})}{K_0 + L(\bar{\omega})}$$

by applying the local minimization interpretation. With each pixel replacement, (3) is reduced (unless f is at the global minimum). If the kernel of $\mathbf{H} + \gamma\mathbf{R}$ is positive definite, then the norm of the difference between the current and the MAP estimate may not increase indefinitely. This implies that $|P(\bar{\omega})| \leq 1$ for all $\bar{\omega}$. Since the solution of the MAP estimation problem is unique if and only if the kernel of $\mathbf{H} + \gamma\mathbf{R}$ is positive definite, we see that a positive definite kernel implies that $|P(\bar{\omega})| < 1$.

Fig. 1 shows the corresponding number of iterations of GS required to achieve 99% reduction of error energy at each frequency when $\gamma = 0$. These plots indicate that the convergence is rapid, especially for the high spatial frequencies.

Because the matrix update equation (8) involves the inversion of a large and relatively full matrix, $\mathbf{K} + \mathbf{L}$, this method would seem to be computationally expensive. However, GS has approximately equivalent computational cost per iteration to GA. By differentiating (3) with respect to Δf_i, the change in pixel i, we find its optimum value is

$$\Delta f_i = \frac{\mathbf{A}_{*i}^t\mathbf{D}(\tilde{p} - \mathbf{A}f) - \gamma\mathbf{R}_{i*}f}{\mathbf{A}_{*i}^t\mathbf{D}\mathbf{A}_{*i} + \gamma\mathbf{R}_{ii}}\,, \qquad (10)$$

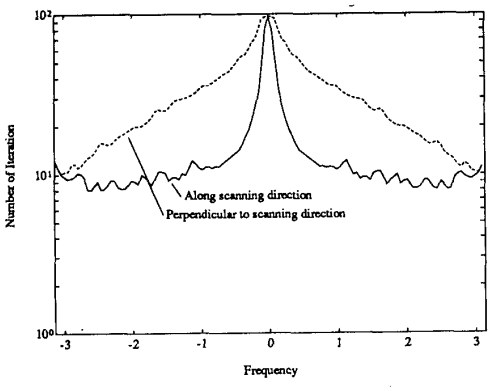

Figure 1: The maximum number of iterations required for 99% error reduction, as a function of spatial frequency, when no regularization is used ($\gamma = 0$).

Computation is dominated by the first term in the numerator, which for a single scan of the image(one iteration), has at most 50%more multiplies per pixel than GA. In practice, however, we have found that a single iteration of GS runs as fast as GA on a general purpose serial computer.

The GS method will also be used to solve the segmentation problem. For this discrete problem the gradient based methods of GA and CG are not applicable. However, the GS method may still be applied by choosing each pixel's value to minimize the *a posteriori* probability. The best choice for pixel i will depend only on the neighbors of f_i, and the updated value is chosen from the feasible set of densities Ω. Since Ω is not convex, determining a global maximum will generally not be possible. However, it is possible to guarantee convergence to a local minimum by changing the density of a pixel only when it strictly increases the log likelihood.

4 Experimental Results

Experimental data presented here result from very low photon dosages relative to most systems. The input photon count (λ_T) for each ray in our trials was 2000, orders of magnitude below dosages of commercial medical CT scanners[1]. Our goals in terms of reconstructed image quality, however, are also much more modest. CBP will serve as a good initial starting point for the iterations in most cases. The phantom is shown in Fig. 6(a), and contains only two non-zero densities. The background is of absorptivity 0.2cm^{-1}, while the four high-density regions are 0.48cm^{-1}. The physical diameter of the phantom is 20cm. At the given photon dosages, rays passing through the larger high-density regions are essentially blocked, making the given trials similar to hollow projection reconstruction. In each case, the number of raysums collected per angle was 128. We chose a single value for α in (6) for all applications of GA, based on the best experimental convergence across the set of simulations. Both methods were compared to conjugate gradient (CG), which has about the same per-iteration cost as GA, and generally significantly faster convergence.

Initial trials were used to verify the analysis of Sec. 3, and were therefore performed with $\gamma = 0$ in (3), corresponding to ML estimation. For results presented here, however, the diagonal entries of $\mathbf{D}$ are photon counts. Fig. 2 illustrates convergence in terms of log likelihood as expressed in (1). CG and GS are comparable in convergence rates, with GS showing an appreciable advantage. Both are much faster than GA. Our analysis predicted this behavior for high frequencies in the difference between the CBP starting point and the Bayesian estimate. As is illustrated in the periodogram spectral estimate of Fig. 3, the error image was of predominantly high frequency spectral content. Similar relative convergence behavior appeared when $\mathbf{D}$ was diagonal. GS appears more useful than the gradient methods for problems where the difference between the starting state and the optimal estimate is dominated by high frequencies. But ML estimates may have *excessive* high frequency content, in contrast to many images of practical interest. The ML estimate images

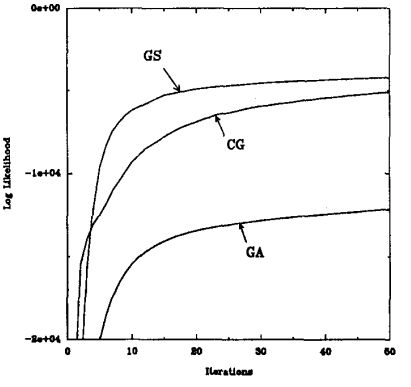

Figure 2: Convergence comparison for real projection weighting matrix D

whose convergence was discussed above, in fact, are of very poor visual quality.

Regularization both speeds convergence, and prevents excessive oscillation in the estimate. For the following results, we use an $\mathbf{R}$ with the form of a discrete 5-point Laplacian. Typical convergence rates are shown in Fig. 4 for MAP estimation with the same optimization methods and $\gamma = 100\text{cm}^2$. (This corresponds to a standard deviation of a pixel given its neighbors of 0.1cm^{-1}) The associated error spectrum has substantial energy at very low frequencies, plus an approximately flat spectral content across the higher frequencies. Here CG enjoys a slight advantage in convergence rate, and both CG and GS are essentially completely converged at fewer than 15 iterations. GA is much slower, as expected. Trials with larger γ yielded still faster convergence, but very similar relationships among the three techniques.

Fig. 5 shows the images resulting from the Bayesian enhancement of the CBP reconstruction. The artifacts of 5(a) are ameliorated substantially, with varying degrees of smoothing according to the choice of γ. Note that we have applied a particularly simple prior here (4-point neighborhood); improvement is possible with more accurate choices.

If an object is known to consist of only a few distinct densities, we can perform an approximate MAP segmentation as an enhancement of the CBP starting point. While gradient methods are not directly applicable, the GS algorithm is well-suited to this problem. Experimental results for a projection data set consisting of 128 rays at each of 16 equally spaced angles appear in the images of Fig. 6. The phantom and dosage per ray are the same as in the previous simulations. The CBP reconstruction is shown in 6(b), with the result of thresholding midway between the two known densities in 6(c). After less than 3 full iterations, GS pro-

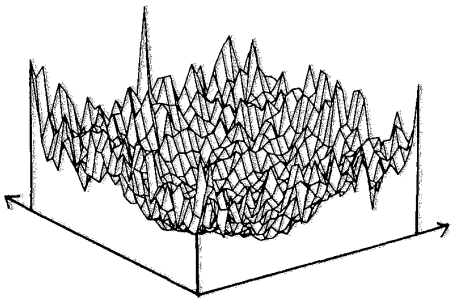

Figure 3: 32 x 32 periodogram of difference between CBP reconstruction and ML estimate. Center of plot is (0,0).

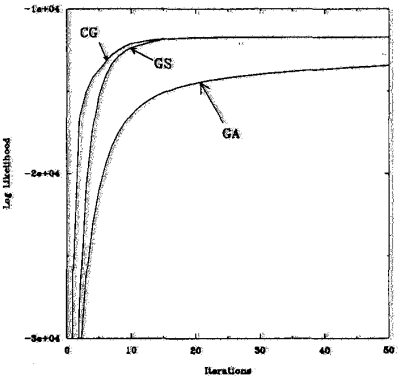

Figure 4: Convergence comparison for real projection weighting matrix D and regularization using a Gaussian prior.

duced the segmentation of 6(d). Nearly all the artifacts resulting from the radio-opaque regions of the object are removed.

5 Conclusion

The GS method shows very good convergence properties, and is applicable to a broad class of optimization problems. In the case of continuously valued reconstruction, these convergence advantages are due to its excellent suppression of high frequency error. The GS method did not show the fastest convergence in all cases, but an important augmentation of the process awaits investigation. Lower frequency components can be forced to converge much more quickly by *multi-grid* implementation. We envision a similar generalization of the discrete-valued version of GS. Performing greedy minimization at multiple resolutions has shown promise in other applications to image segmentation[7], in terms of both computational efficiency and quality of segmentation.

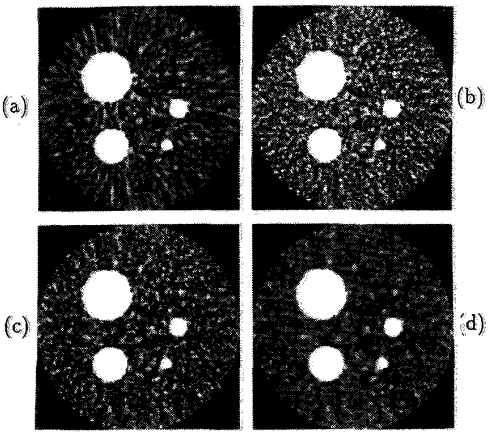

Figure 5: (a) CBP reconstruction from 128 x 128 noisy projections. (b) Bayesian estimate with $\gamma = 100$. (c) $\gamma = 200$. (d) $\gamma = 500$.

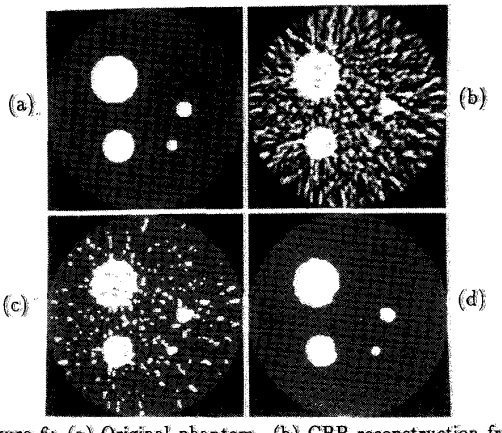

Figure 6: (a) Original phantom. (b) CBP reconstruction from 16 x 128 noisy projections. (c) Threshold segmentation of CBP image. (d) Gauss-Seidel approximation of MAP estimate.

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] D. M. Young, *Iterative Solution of Large Linear Systems*, Academic Press, New York, 1971.

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

[5] H. Derin and H. Elliott, "Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields," *IEEE Trans. Pat. An. Mach. Intell.*, vol. PAMI-9, pp. 39-55, Jan. 1987.

[6] 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.

[7] C. Bouman and B. Liu, "Multiple Resolution Segmentation of Textured Images," to appear in the *IEEE Trans. on Pat. An. Mach. Intell.*, Feb. 1990.