

A Local Update Strategy for Iterative Reconstruction from Projections ^{†‡}

Ken Sauer

Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556
(219) 239-6999

Charles Bouman

School of Electrical Engineering
Purdue University
West Lafayette, IN 47907-0501
(317) 494-0340

Abstract ¹

Iterative methods for statistically-based reconstruction from projections are computationally costly relative to convolution backprojection, but allow useful image reconstruction from sparse and noisy data. We present a method for Bayesian reconstruction which relies on updates of single pixel values, rather than the entire image, at each iteration. 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 discretely valued. We demonstrate that Bayesian segmentation using GS iteration produces useful estimates at much lower signal-to-noise ratios than required for continuously valued reconstruction. This paper includes analysis of the convergence properties of gradient ascent and GS for reconstruction from integral projections, and simulations of both maximum-likelihood and maximum *a posteriori* cases.

EDICS #6.2.2 Image Processing and Analysis; Reconstruction Restoration and Recovery.

[†]This work was supported by an NEC Faculty Fellowship.

[‡]*IEEE Trans. on Sig. Proc.*, vol. 41, no. 2, pp. 534-548, Feb. 1993.

¹Permission to publish this abstract separately is granted.

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].

Iterative methods have dominated techniques for the inversion problem in sparse and noisy data. Some deterministic iterative algorithms consist of alternately enforcing constraints imposed by consistency with data, and membership in sets of solutions possessing specific characteristics. A familiar example is the algebraic reconstruction technique (ART), which forces the candidate solution to satisfy one integral projection measurement at each iteration. ART is one of a general class of algorithms known as projection onto convex sets[7], with each set in ART described by the value of the given projection measurement. Other sets can be described by qualities such as non-negativity or upper limits on absorptive densities. The behavior of deterministic methods is often unsatisfactory if data is very noisy, and/or the solution is underspecified by the observations and set memberships. The introduction of regularization in the form of a cost function penalizing local irregularity of the reconstructed surface may improve performance, at the cost of resolution in the final image[1].

Statistical methods of image reconstruction and restoration may be quite similar to other iterative approaches in their implementation, but 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[8]. Bayesian 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 prior distribution weights the likelihood function and may even impose hard constraints on the solutions. In the preponderance of previous work on the problem, the overall cost function specified by the log likelihood for either ML or Bayesian techniques has been assumed quadratic for tomographic image reconstruction, allowing application of gradient methods. For example,

both the prior distribution of the cross-section and the distribution of the projection measurements conditioned on the underlying image function are assumed Gaussian in many reports[9, 10].

In this paper, we develop a local update method to perform Bayesian reconstruction from projections. Our approach is similar to Gauss-Seidel(GS) iterations employed in the solution of differential equations on finite grids[11], and we will use the same terminology. Each iteration of this method includes only the optimization with respect to a single pixel's value. Since this update strategy is local, it is well suited for use with typical prior models for the cross-section. 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, as is illustrated in the analysis of Sec. 3, and the experimental work of Sec. 5, the GS approach has faster convergence in high spatial frequencies for continuously-valued images than either simple gradient ascent, or conjugate gradient.

If a cross-section is known to consist of only a few distinct densities, then accurate reconstruction is possible with very low SNR. Image estimation then consists of assigning each pixel to one of the known densities. The resulting estimated cross-section is a segmentation which separates the image into discrete regions. The cost functions which arise in this problem are generally nonconvex and nondifferentiable, and therefore not amenable to gradient approaches, which rely on partial derivatives of the cost. However, the GS technique is well-suited to this segmentation problem, and finds an estimate very quickly.

We avoid using a Gaussian approximation for the error in the projection measurements by deriving a quadratic approximation to the log likelihood of the actual photon counts. This approximation is shown to be reasonably accurate even for small photon counts (≈ 10). The simplification is useful since it allows analysis of the convergence rate and reduces computation relative to the exact log likelihood. The resulting cost functional weights the residual error for each projection in proportion to the count of photons reaching the corresponding detector. In the limiting case of opaque projections, no photons are collected and the observation is given no weight. The algorithm is therefore applicable without modification to the canonical missing-data problems, such as limited-angle reconstruction, and truncated or hollow projections.

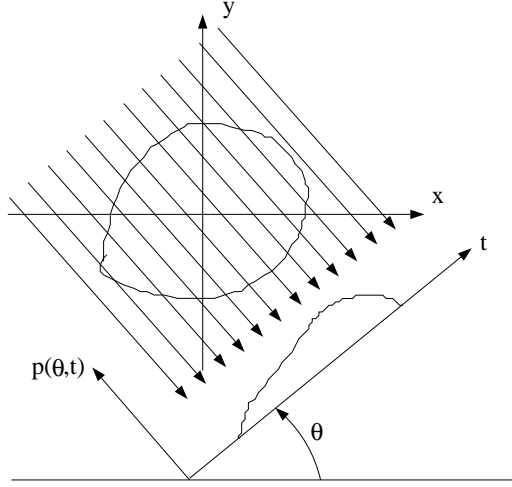


Figure 1: Projection data for angle θ , resulting in the one-dimensional function $p(\theta, t)$.

2 Model of Physical System

The 2-D Radon transform maps a function of two variables, which we denote by $f(x, y)$, into a function indexed by (θ, t) according to

$$p(\theta, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(t - x \cos \theta - y \sin \theta) dx dy \quad (1)$$

where $\delta()$ is an impulse function. Fig. 1 illustrates the collection of projection data for a single value of θ . The value of $p(\theta, t)$ represents the integral of some property of $f(x, y)$ along the ray at orientation $\theta + \frac{\pi}{2}$, at a displacement t from the center of the field. By the Fourier Slice Theorem[12], the Fourier transform of each 1-D set of projection data forms a sample $P(\theta, \omega)$ passing through the origin of the Fourier transform of $f(x, y)$ at orientation θ .

Given certain constraints on $f(x, y)$ [13], the 2-D function $p(\theta, t)$ can be inverted by

$$f(x, y) = \frac{1}{2\pi} \int_0^\pi \int_{-\infty}^{\infty} P(\theta, \omega) |\omega| e^{j\omega(x \cos \theta + y \sin \theta)} d\omega d\theta. \quad (2)$$

or

$$f(x, y) = \int_0^\pi \int_{-\infty}^{\infty} p(\theta, t) h(x \cos \theta + y \sin \theta - t) dt d\theta, \quad (3)$$

where $h(t)$ is a one-dimensional backprojection filter. The equality of (2) and (3) is due to the relationship of convolution to multiplication of Fourier transforms[12]. The filter $h(t)$ is in practice a lowpass version of the inverse Fourier transform of $|\omega|$, whose design is based on

an estimated bandlimit for $f(x, y)$ and the necessity for suppression of high-frequency noise in the projection data. The above inversion techniques are known as filtered backprojection (FBP) and convolution backprojection (CBP), respectively. “Convolution” refers to the inner integral in (3), and “backprojection” refers to the outer integral. Most commercial tomographic imaging systems employ CBP, with the choice of h the most important element in the design of the reconstruction algorithm[14, 15].

In practice, reconstruction requires finite-dimensional representation of both the observations, $p(\theta, t)$, and the modeled image, $f(x, y)$, as the finite-dimensional vectors p and f , respectively. In the case of CBP, a discrete approximation of (3) must be implemented. One approach is to formulate the problem directly in a discrete domain. To achieve this, we may represent $f(x, y)$ as a sum of N local components, $w_i(x, y)$, each modulated by the local pixel value $f_i[1]$.

$$f(x, y) = \sum_{i \in S} f_i w_i(x, y)$$

In this notation, S denotes the set of N pixel locations, and i is the pixel index. Generally, $w_i(x, y)$ will have the form

$$w_{(i_1, i_2)}(x, y) = w_{(0,0)}(x - i_1 T, y - i_2 T) \quad (4)$$

where T is the spatial sampling period, and $i = (i_1, i_2)$ is the position in a 2-D lattice. The projections may be discretized by computing them for only a finite set of M projection rays, $\{(\theta_j, t_j)\}_{j=0}^M$. The j^{th} projection is then written as $p_j = p(\theta_j, t_j)$. The Radon transform equations may now be written in the discrete form

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

where \mathbf{A} is a sparse $M \times N$ matrix of line integrals.

$$\mathbf{A}_{ji} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w_i(x, y) \delta(t_j - x \cos \theta_j - y \sin \theta_j) dx dy$$

If the $w_i(x, y)$ represent constant-valued pixels, \mathbf{A}_{ji} is simply the length of the intersection of projection ray j with pixel i .

2.1 Statistical Analysis

In transmission tomography the projections, p , are not measured directly. Instead, raw data are in the form of the number of photons detected after passing through an absorptive

material. This situation is illustrated in Fig. 2. Input photon counts are Poisson-distributed random variables. Calibration allows the input rate to be approximated well enough that we will assume it to be known, and denote it by λ_T . Given this assumption, the photon count, Λ_j , corresponding to projection j , is a Poisson distributed random variable with mean and variance $\lambda_T \exp\{-p(\theta_j, t_j)\}$. The random vector of photon counts at all angles and displacements is denoted by the upper case Greek letter Λ . We will denote particular realizations of this random vector by the lower case λ . Under the usual assumptions of independence of photon counts among angles and displacements, the conditional distribution of the photon counts Λ given f is

$$P(\Lambda = \lambda|f) = \prod_{j=1}^M (\lambda_j!)^{-1} \exp\left(-\lambda_T e^{-\mathbf{A}_{j*}f}\right) \left(\lambda_T e^{-\mathbf{A}_{j*}f}\right)^{\lambda_j} \quad (5)$$

where \mathbf{A}_{j*} is the j^{th} row of \mathbf{A} . The log likelihood function reduces to

$$L(\lambda|f) = -\sum_{j=1}^M [\lambda_T e^{-\mathbf{A}_{j*}f} + \lambda_j \mathbf{A}_{j*}f + \log(\lambda_j!)] . \quad (6)$$

Using this form, the maximum likelihood (ML) estimate may be computed by solving the following optimization problem over the feasible set of solutions, Ω :

$$\hat{f} = \arg \max_{f \in \Omega} L(\lambda|f) \quad (7)$$

The quality of the reconstruction can be improved by choosing the set Ω to constrain the solution based on prior knowledge of feasible solutions.

The ML estimate does not incorporate reasonable prior information about the behavior of the image. For instance, the ML estimator may contain excessive high frequency variation which we would not expect to exist in the original image. If the number of samples taken is small relative to the number of unknown pixel values, the solution may also be underdetermined.

These difficulties may be addressed by treating the original image as a random field, F , with prior distribution, $g(f)$. Again, we use a lower case f to denote a particular realization of the random vector F . The prior distribution regularizes the optimization problem so that a unique solution always exists[16]. The logarithm of the *a posteriori* distribution of F given Λ may be computed using Bayes' formula.

$$\begin{aligned} L_p(f|\lambda) &\triangleq \log P(F \in df|\Lambda = \lambda) \\ &= L(\lambda|f) + \log g(f) - \log P(\Lambda = \lambda) \end{aligned}$$

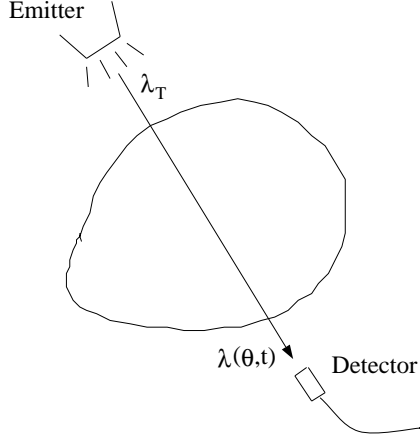


Figure 2: Collection of photon count $\Lambda(\theta, t)$. λ_T is the input photon rate, given by calibration.

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

$$\begin{aligned}
 \hat{f} &= \arg \max_{f \in \Omega} L_p(f|\lambda) \\
 &= \arg \max_{f \in \Omega} \{L(\lambda|f) + \log g(f)\} \\
 &= \arg \max_{f \in \Omega} \{L(\lambda, f)\}
 \end{aligned} \tag{8}$$

The last equation indicates that the MAP estimate also maximizes the log of the joint distribution, $L(\lambda, f) = \log P(F \in df, \Lambda = \lambda)$. MAP estimation is computationally direct and has experimentally been shown to work well in a variety of problems[17, 18, 19, 20, 21]. We will treat only the MAP estimation problem, since the ML estimate is the special case of a constant prior distribution.

In order to simplify and extend the theoretical analysis of this paper, we will also introduce a quadratic approximation to the log likelihood function given in (6). In Appendix A we derive the approximation

$$L(\lambda|f) \approx -\frac{1}{2}(\hat{p} - \mathbf{A}f)^t \mathbf{D}(\hat{p} - \mathbf{A}f) + c(\lambda) \tag{9}$$

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

$$\begin{aligned}
 \hat{p}_j &= \log(\lambda_T/\lambda_j) \\
 \mathbf{D} &= \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_M\} .
 \end{aligned}$$

Since $c(\lambda)$ is independent of f , it may be ignored in the ML and MAP estimation of f . Note that this does *not* constitute a Gaussian approximation of any of the densities involved. The image f is equivalent to a parameter vector in this expression, and the log likelihood is of interest simply as a function of f in our optimization. The entries of λ appear in both \mathbf{D} and \hat{p} of (9) as indicated above, making this approximation non-quadratic in λ .

The key to the behavior of (9) 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. All of the algorithms which will be derived in this paper may be implemented using the more accurate likelihood function, (6), without dramatically changing the computational requirements or analytical properties of the methods². In particular, (6) is convex so global convergence properties will be consistently retained. However, we will use (9) because it is a good approximation for most realistic cases, and it will be used as the basis of an analysis which gives significant insight into the problem. Figure 3 shows the comparison of the actual log-likelihood function and the approximation over two orders of magnitude in photon count.

2.2 Prior Distributions and Problem Constraints

As mentioned earlier, it is desirable and often necessary to apply prior distributions and hard constraints in recovering the estimate \hat{f} . This prior information will have one of two forms, depending on whether the goal is reconstruction of a continuously valued f or segmentation of f into a discrete set of known densities. In either case, estimates benefit from prior distributions which reflect the high correlation of spatially adjacent pixels. The Markov random field (MRF) model[22] has proven very useful in image estimation problems such as ours. The MRF is characterized by localization of pixel interactions, and consequent simple optimization operations.

Under some weak technical conditions, a random field is a MRF if and only if it has a probability distribution corresponding to a Gibbs distribution[22, 23]. This result, which is known as the Hammersley-Clifford theorem, may be used to express the likelihood function

²Some minor differences in the following material are that the conjugate gradient method will no longer converge in N iterations, and the GS update method will only strictly increase the posterior probability with each update instead of maximizing it with respect to a pixel's value.

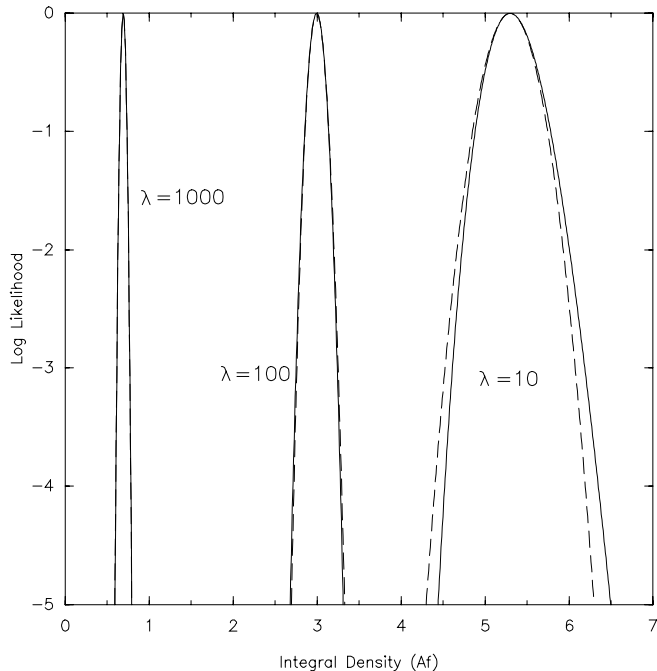


Figure 3: Plots of the log-likelihood as a function of a single projection across the reconstructed image. The exact Poisson model (dashed lines) and the quadratic approximation (solid lines) are each plotted for three values of λ . Each of these plots covers a confidence interval of 0.99 for a maximum-likelihood estimate of the projection value.

$\log g(f)$. While the theory of MRF's is quite extensive[24, 25, 26], we will restrict our attention to a simple model based on at most an 8-point neighborhood. Similar models have been successfully applied to image segmentation[18, 20, 21].

2.2.1 Gaussian Markov Priors

When F is continuously valued, we will generally assume that it is a Gaussian random vector, and therefore has a log likelihood function with the quadratic form

$$\log g(f) = -\frac{\gamma}{2} f^t \mathbf{R} f + c \quad (10)$$

where \mathbf{R} is symmetric, and $\frac{1}{\gamma} \mathbf{R}^{-1}$ is the covariance matrix for F . We will further assume that \mathbf{R} is equal to 1 on its diagonal, and that, ignoring boundary effects, multiplication by \mathbf{R} has the effect of convolution with a stationary kernel, $r_{(i,j)}$. The constant γ^{-1} is equal to the prediction variance of a pixel given its neighbors in the prior model. This is equivalent to

assuming that F is a Gaussian Markov random field (MRF) with 2-D power spectral density given by

$$G_F(\bar{\omega}) = \frac{1}{\gamma R(\bar{\omega})} \quad (11)$$

where $\bar{\omega} = (\omega_1, \omega_2)$, and $R(\bar{\omega})$ is the 2-D discrete time Fourier transform (DTFT) of $r_{(i,j)}$ [27, 22, 28]. In particular, we will make the common assumption that \mathbf{R} has the form of a discrete 5-point approximation to a Laplacian:

$$r_{(i,j)} = \begin{cases} 1 & \text{if } i = j = 0 \\ 1/4 & \text{if } \{i = 0 \text{ and } j = \pm 1\} \text{ or } \{i = \pm 1 \text{ and } j = 0\} \end{cases}$$

This results in the assumed power spectral density

$$G_F(\bar{\omega}) = \frac{1}{\gamma(1 - 1/2 \cos \omega_1 - 1/2 \cos \omega_2)} .$$

This prior distribution may be used in conjunction with the log likelihood of (9) to form the MAP estimation criterion:

$$\begin{aligned} \hat{f} &= \arg \max_{f \in \Omega} L(\lambda, f) \\ &= \arg \max_{f \in \Omega} \left\{ -(\hat{p} - \mathbf{A}f)^t \mathbf{D}(\hat{p} - \mathbf{A}f) - \gamma f^t \mathbf{R}f \right\} \end{aligned} \quad (12)$$

When $\Omega = \mathbb{R}^N$, the solution to this optimization problem may be found by differentiating with respect to f . The resulting equation has the form

$$b = (\mathbf{H} + \gamma \mathbf{R}) \hat{f} \quad (13)$$

where $\mathbf{H} = \mathbf{A}^t \mathbf{D} \mathbf{A}$ and $b = \mathbf{A}^t \mathbf{D} \hat{p}$. This equation will play a central role in our analysis.

2.2.2 Non-Gaussian Priors

A second distinct case results when Ω is a proper subset of \mathbb{R}^N . This may occur due to a variety of hard constraints on the solution's form. An important special case occurs when this set is convex. For example, it is often reasonable to assume that f must be a nonnegative function. In this case, Ω is a convex set and there is a unique global maximum to (12). The difficulty of enforcing such a convex constraint will vary with the type of optimization technique used. By projecting solutions onto the convex feasible set, many algorithms may be modified to enforce the constraint [7]. However, we will develop an algorithm in Section 3.1 which can more efficiently implement this type of convex constraint.

Enforcing hard constraints becomes more difficult when the feasible set is not convex, as is the case when each pixel is assumed to have one of a fixed number of densities. Reconstruction then corresponds to choosing each pixel from a discrete and finite set, O , of possible densities. This segmentation process can be performed more accurately than continuously valued reconstruction when the signal-to-noise ratio (SNR) is low.

We choose a discrete MRF model for F which is often used in segmentation problems[17, 20, 31]. This model encourages neighboring pixel to have the same densities. To define the model, we must first define two simple functions, $t_1(f)$ and $t_2(f)$. $t_1(f)$ is the number of horizontally and vertically neighboring pixel pairs with different densities in f , and $t_2(f)$ is the number of diagonally neighboring pixel pairs with different densities in f . The discrete density function for $f \in O^N$ is then assumed to be of the form

$$\log g(f) = -\gamma \left(t_1(f) + \frac{t_2(f)}{\sqrt{2}} \right) + \log(Z) \quad (14)$$

where Z is a constant. Substituting this form into (8), we obtain the optimization criterion

$$\hat{f} = \arg \max_{f \in O^N} \left\{ -(\hat{p} - \mathbf{A}f)^t \mathbf{D}(\hat{p} - \mathbf{A}f) - \gamma \left(t_1(f) + \frac{t_2(f)}{\sqrt{2}} \right) \right\}. \quad (15)$$

3 Optimization Techniques

The objective of this section is the development of a new iterative reconstruction technique, based on local Gauss-Seidel (GS) updates, for efficiently solving the optimization problems of (12) and (15). We show that this method can efficiently solve the quadratic problem of (12), and a wide variety of problems that incorporate hard constraints on Ω . This is particularly important in the segmentation problem of (15) since the gradient-based algorithms implicitly require that Ω be a continuous convex set. We analyze the convergence performance of GS for the case of the Gaussian prior, and compare it to two well-known numerical optimization techniques, gradient ascent (GA) and conjugate gradient (CG), which are directly applicable to quadratic optimization of (12). These comparisons also include computational expense per iteration of each method.

3.1 Gauss-Seidel

Gauss-Seidel iteration is a well-known technique for solving partial differential equations (PDE) using finite element methods[11]. The strategy of GS iterations in PDE's is to sequentially enforce the conditions of the differential equation at points in the solution. Each pixel update is chosen in sequence to minimize a global cost function. This is in contrast to methods such as gradient ascent(GA), which update all points simultaneously. We will use the local minimization interpretation as the basis for our development of GS.

Our objective in the case of the Gaussian prior is to pick the value of a single pixel, f_i , which minimizes the posterior log likelihood function of (12) as a 1-D function of the single entry, given the current state of the remainder of f . The updates are most simply viewed in terms of the Radon transform domain error. Changing the value of f at the point i by the amount Δf_i , changes the projection error $\hat{p} - \mathbf{A}f$ in the following way:

$$\hat{e} = e - \mathbf{A}_{*i}\Delta f_i \quad (16)$$

where \mathbf{A}_{*i} is the i^{th} column of \mathbf{A} and \hat{e} is the new value of the projected error. Therefore, the equation for the new log likelihood is

$$L(\lambda|\hat{f}) = L(\lambda|f) + \theta_1\Delta f_i - \frac{\theta_2}{2}(\Delta f_i)^2 \quad (17)$$

where $\hat{f}_i = f_i + \Delta f_i$ and the values θ_1, θ_2 are defined by

$$\theta_1 = \mathbf{A}_{*i}^t \mathbf{D}e \quad (18)$$

$$\theta_2 = \mathbf{A}_{*i}^t \mathbf{D}\mathbf{A}_{*i} . \quad (19)$$

We can now add the log likelihood for the prior distribution to the expression of (17). If we use a Gaussian prior distribution, the optimum value for Δf_i is given by

$$\Delta f_i = \frac{\theta_1 - \gamma \mathbf{R}_{i*} f}{\theta_2 + \gamma \mathbf{R}_{ii}} . \quad (20)$$

The complete update equations for the point at position i are given in Fig. 4.

The GS method will also be used to solve the segmentation problem of (15). For this discrete problem gradient-based methods are not applicable. However, the GS method may still be applied by choosing each pixel's value to minimize the *a posteriori* probability. Since the log of the prior distribution (14) depends only on the number of neighboring pixels with

$$\begin{array}{rcl}
\text{Precompute:} & \theta_1 & = \mathbf{A}_{*i}^t \mathbf{D} e \\
\theta_2(i) = \mathbf{A}_{*i}^t \mathbf{D} \mathbf{A}_{*i} & \Delta f_i & = \frac{\theta_1 - \gamma \mathbf{R}_{i*} f}{\theta_2(i) + \gamma \mathbf{R}_{ii}} \\
& \hat{f}_i & = f_i + \Delta f_i \\
& \hat{e} & = e - \mathbf{A}_{*i} \Delta f_i
\end{array}$$

Figure 4: The equations for the Gauss-Seidel update of the pixel at location i . The values θ_1 and Δf_i are temporary constants. \hat{f} and \hat{e} are the updated values of state vectors.

differing density, the best choice for pixel i will depend only on the neighbors of f_i , which we denote $f_{\partial i}$. Specifically, the updated value is chosen using the formula

$$\hat{f}_j = \arg \max_{x \in O} \left\{ \theta_1(x - f_j) - \frac{\theta_2}{2}(x - f_j)^2 - \gamma \left(v_1(x, f_{\partial j}) + \frac{v_2(x, f_{\partial j})}{\sqrt{2}} \right) \right\} \quad (21)$$

where O is the set of feasible densities for a pixel, $v_1(x, f_{\partial j})$ counts the number of horizontal and vertical neighbors of f_j which do not have density x , and $v_2(x, f_{\partial j})$ counts the number of diagonal neighbors of f_j which do not have density x . Since $\Omega = O^N$ 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 reduces the value of (21).

For the discrete problem, the GS algorithm is essentially the same as the iterated conditional modes (ICM) method proposed by Besag[20]. The only difference is that the ICM method uses the ML estimate as the initial starting point of the algorithm, while we use the pointwise segmentation of the thresholded CBP, since it generally gives a good initial starting point. Of course, the GS update algorithm of (21) is still essential, since it gives a specific method of implementing ICM (for the transmission tomography problem) which we will show is computationally efficient. Similarly, this same update algorithm can be used to implement stochastic optimization techniques such as simulated annealing. These methods can improve the quality of the reconstruction by determining a solution closer to the global optimum. Unfortunately, this improvement is at the cost of additional computation[19, 30]. Deterministic multiscale optimization methods may also be adaptable to this problem[21, 31] and have been shown to reduce computation and improve performance.

$$\begin{aligned}
f^{(1)} &= \text{CBP} & \nabla^{(n+1)} &= \nabla^{(n)} - \alpha^{(n)} \tilde{d}^{(n)} \\
\nabla^{(1)} &= \nabla_f L(\lambda, f^{(1)}) & \beta^{(n+1)} &= \frac{\|\nabla^{(n+1)}\|^2}{\|\nabla^{(n)}\|^2} \\
d^{(1)} &= \nabla^{(1)} & d^{(n+1)} &= \nabla^{(n+1)} + \beta^{(n+1)} d^{(n)} \\
\tilde{d}^{(1)} &= (\mathbf{H} + \gamma \mathbf{R}) d^{(1)} & \tilde{d}^{(n+1)} &= (\mathbf{H} + \gamma \mathbf{R}) d^{(n+1)} \\
\alpha^{(1)} &= \frac{\|\nabla^{(1)}\|^2}{(d^{(1)})^t \tilde{d}^{(1)}} & \alpha^{(n+1)} &= \frac{\|\nabla^{(n+1)}\|^2}{(d^{(n+1)})^t \tilde{d}^{(n+1)}} \\
& & f^{(n+1)} &= f^{(n)} + \alpha^{(n)} d^{(n)}
\end{aligned}$$

Figure 5: The equations for conjugate gradient search using the quadratic approximation.

3.2 Gradient Methods

Gradient ascent (GA) is an iterative scheme in which successive values of the reconstruction are given by

$$\begin{aligned}
f^{(n+1)} &= f^{(n)} + \alpha \nabla_f L_p(f^{(n)}, \lambda) \\
&= f^{(n)} - \alpha \left(\mathbf{A}^t \mathbf{D} (\mathbf{A} f^{(n)} - \hat{p}) + \gamma \mathbf{R} f^{(n)} \right).
\end{aligned} \tag{22}$$

This method is closely related to the numerical method of weighted Jacobi iterations (also known as the simultaneous overrelaxation method or JOR) used in solving finite element approximations to partial differential equations[11].

Conjugate gradient (CG) is a general method for maximizing differentiable functions. It is closely related to GA since it uses the computed gradient to determine the direction of change at each iteration. The critical difference is that CG requires $d^{(n+1)}$, the direction of minimization in iteration $n + 1$, to be conjugate to the previous direction of minimization under the $\mathbf{H} + \gamma \mathbf{R}$ inner product[32, 33].

$$(d^{(n+1)})^t (\mathbf{H} + \gamma \mathbf{R}) d^{(n)} = 0$$

CG generally converges much faster than GA because successive corrections do not interfere with one another. The update and initialization equations for CG are given in Fig. 5.

3.3 Computational Costs

At this point, it is not clear that individual GS iterations may be performed with reasonable computation. To compare relative computational cost of the methods, we will determine the number of multiplies required for one complete iteration of GS, GA and CG. Table 1 summarizes the results and indicates that one iteration of any of the three methods has computational complexity on the order of a single application of convolution backprojection.

The computational requirements of projection and backprojection (represented by $\mathbf{A}f$ and $\mathbf{A}^t p$, respectively), are equal since multiplication by \mathbf{A}^t and \mathbf{A} have equivalent complexity. 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.

If the length of the filter in CBP is denoted N_h , the number of multiplies required for filtering is $M N_h$. Backprojection of the filtered data in CBP normally involves averaging or interpolation, with cost approximately equal to that of the product $\mathbf{A}^t p$. This results in a total of $M_0 N + N_h M$ multiplies for CBP, with the first term usually the larger.

The basic operation of the GS method is the computation of Δf_i using (20). The computational cost of this operation is dominated by the evaluation of θ_1 and θ_2 . θ_2 does not depend on Δf_i so it may be precomputed for each pixel. Since an average column of \mathbf{A} contains M_0 nonzero elements, the number of multiplies required to compute θ_1 is either M_0 or $2M_0$, depending on whether we assume that $\mathbf{A}_{*i}^t \mathbf{D}$ is computed or stored in a table. The update of the projected error vector in (16) requires M_0 multiplies, and computation of Δf_i requires a single divide and N_R multiplies, where N_R is the number of points in each neighborhood of the MRF represented by \mathbf{R} . Therefore, the total number of multiplies and divides for a complete update of the GS method is $(2M_0 + N_R + 1)N$ if a lookup table is used and $(3M_0 + N_R + 1)N$ otherwise.

Each iteration of GA as given in (22) requires the computation of a projection, a backprojection and multiplication by the matrices \mathbf{D} and \mathbf{R} . This yields a total of $(2M_0 + N_R)N + M$ multiplies. The computational requirements of CG are comparable to those of GA. Operations are dominated by the projection and backprojection required in multiplication by \mathbf{H} . As in the case of GA, this requires $(2M_0 + N_R)N + M$ multiplies. In addition, $6N$ multiplies are needed for computing inner products and scaling vectors. Therefore, the total number

Convolution Backprojection	$M_0N + N_hM$
Gradient Ascent	$(2M_0 + N_R)N + M$
Conjugate Gradient	$(2M_0 + N_R)N + M$
Gauss-Seidel	$(3M_0 + N_R + 1)N$

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 and N_R is the number of points in the regularizing kernel. The iterative methods assume the use of a Gaussian prior distribution.

of multiplies per iteration of CG is given by $(2M_0 + N_R + 6)N + M$.

Table 1 summarizes the computational requirements of CBP and the three iterative reconstruction methods. Assuming that $M_0 \gg N_R$, the GS update method requires more multiplies than GA and CG for a single full update (assuming that a second lookup table is not available). However, we have found that the computational cost is largely governed by the communications and indexing associated with a single pass through the projected data set. In this less quantitative sense, the three methods are equivalent since the gradient based methods require a projection and then back projection operation and the GS method requires the computation of θ_1 and then an update of the projection error. In practice, we have found that a single iteration of GS runs as fast as GA and CG on a general purpose serial computer.

4 Convergence Analysis

To make the comparison of computational expenses meaningful, we must ascertain good convergence behavior for GS. Both GS and GA are amenable to analysis, and we compare their convergence performance in terms of spatial frequencies. Analysis is difficult in the general case, but we will gain insight into the problem by studying convergence when $\mathbf{D} = \sigma^{-2}\mathbf{I}$ (i.e. the variances of all projections are equal) and the projections are uniformly spaced in both θ and t . CG is not so tractable, but is compared to GS and GA experimentally.

4.1 Gauss-Seidel

In Section 3.1, we addressed the choice of each pixel f_i for GS in the Radon transform domain in terms of errors in projection estimates, for the sake of simple update operations. This can also be done by setting the derivative of the log-likelihood with respect to f_i equal to zero.

$$\begin{aligned} 0 &= \frac{dL(\lambda, f)}{df_i} \\ &= (\nabla_f L(\lambda, f))_i \\ &= b_i - (\mathbf{H} + \gamma\mathbf{R})_{i*}f \end{aligned}$$

The notation $(\mathbf{H} + \gamma\mathbf{R})_{i*}$ denotes the i^{th} row of the matrix. We may use this single equation to specify the best value of f_i in terms of the remaining values of f .

A sequential update of all the pixels can be formulated in a matrix expression. For this purpose, the matrix $\mathbf{H} + \gamma\mathbf{R}$ must be separated into its lower triangular, diagonal and upper triangular parts which we denote by \mathbf{L} , \mathbf{K} and \mathbf{U} respectively. Assuming that the points are replaced in the order of their index i , the update equation for the pixel f_i has the form

$$b_i = \mathbf{L}_{i*}f^{(n+1)} + \mathbf{K}_{i*}f^{(n+1)} + \mathbf{U}_{i*}f^{(n)} \quad (23)$$

where $f^{(n)}$ is the n^{th} complete update. This equation holds for all values of i , so the complete update equation may be written as

$$f^{(n+1)} = -(\mathbf{K} + \mathbf{L})^{-1}\mathbf{U}f^{(n)} + (\mathbf{K} + \mathbf{L})^{-1}b. \quad (24)$$

The update equation (24) has the form of a discrete time system where n indexes time. It is well known that the convergence behavior of such a system is determined by the eigenvalues of the update matrix, $(\mathbf{K} + \mathbf{L})^{-1}\mathbf{U}$.

To extend this analysis, we must first make some observations about the matrix $\mathbf{H} = \sigma^{-2}\mathbf{A}^t\mathbf{A}$ from which the update matrix is derived. In Appendix B, we show that \mathbf{H} is approximately Toeplitz-block-Toeplitz (if f is indexed in raster order, and both θ and t are sampled uniformly). This means that each row of \mathbf{H} represents a spatially shifted version of a single kernel $h_{(i,j)}$. In general, we will refer to the function which specifies the rows of a Toeplitz-block-Toeplitz matrix as the kernel of the matrix. Therefore, multiplication of $f^{(n)}$ by \mathbf{H} is equivalent to convolving the function $f^{(n)}$ with the 2-D kernel $h_{(-i,-j)}$ and retaining the result in the finite spatial window corresponding to the pixel lattice S . Multiplication by a Toeplitz-block-Toeplitz matrix is then approximately equivalent to convolution with the corresponding spatially reversed kernel.

Figure 6: The causal kernel of the Toeplitz-block-Toeplitz matrix \mathbf{L} . The kernel was measured from projection and backprojection of a point.

The form of \mathbf{H} indicates it is symmetric and positive semidefinite. This implies that the function $h_{(i,j)}$ will be symmetric and, ignoring boundary effects, will be real and positive semidefinite (i.e. have a nonnegative Fourier transform). In fact, we can experimentally measure the function $h_{(i,j)}$ by computing the projections of a single point and then backprojecting the result. Since we can only do this over a finite window, we apply a Barlett weighting to our measured kernel to retain the positive semidefinite property.

If \mathbf{R} is chosen to be Toeplitz-block-Toeplitz, \mathbf{L} , \mathbf{K} and \mathbf{U} will have the same property. Since \mathbf{K} is diagonal, $\mathbf{K} = K_0\mathbf{I}$. The interpretation for \mathbf{L} is that its kernel will be the causal part of the kernel of $\mathbf{H} + \gamma\mathbf{R}$ minus the origin. Fig. 6 shows the kernel of the experimentally measured matrix \mathbf{L} when $\gamma = 0$. Notice that it is nonzero only at points occurring before the origin in raster scan order. Similarly, the kernel of \mathbf{U} will include only points after the origin.

Since the matrices in (24) are Toeplitz-block-Toeplitz, they are approximately equivalent to linear spatially invariant filters. Therefore, we may apply frequency analysis methods by

taking the 2-D DTFT of the matrix update equation,

$$f^{(n+1)}(\bar{\omega}) = \frac{-L^*(\bar{\omega})}{K_0 + L(\bar{\omega})} f^{(n)}(\bar{\omega}) + \frac{-1}{K_0 + L(\bar{\omega})} b(\bar{\omega}) \quad (25)$$

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 some immediate 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, (12) 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 $|P(\bar{\omega})| \leq 1$.

Fig. 7 shows the two dimensional function $|P(\bar{\omega})|$ when $\gamma = 0$. (The convergence is generally better for larger values of γ .) It is not surprising that the function is anisotropic since the order of the pixel updates gives an orientation to the convergence behavior. The function has a wide but flat main lobe. This means that a narrow band of spatial frequencies will not converge rapidly. In practice this is not a serious problem, since complete updates may alternate between horizontal and vertical raster ordering. Fig. 8 shows two one-dimensional plots of the maximum of $|P(\bar{\omega})|$ along the ω_1 and ω_2 dimensions. Fig. 9 shows the corresponding maximum number of iterations required to achieve 99% reduction of error energy at each frequency. Fig. 10 show experimental results and theoretical predictions for the convergence of 2-D sinusoids (with a Bartlett window to localize frequency content). This was done by initializing the photon counts to be $\lambda_j = \lambda_T$, and measuring number of iterations required for the sinusoid to decay. In the scanning direction, the experimental and theoretical results agree well. The direction perpendicular to the scan is, as mentioned above, less critical. These plots indicate that the convergence is rapid for the high spatial frequencies.

Figure 7: The two dimensional convergence rate $|P(\bar{\omega})|$ when no regularization is used ($\gamma = 0$). The raster scanning is assumed to occur along rows in the ω_1 direction.

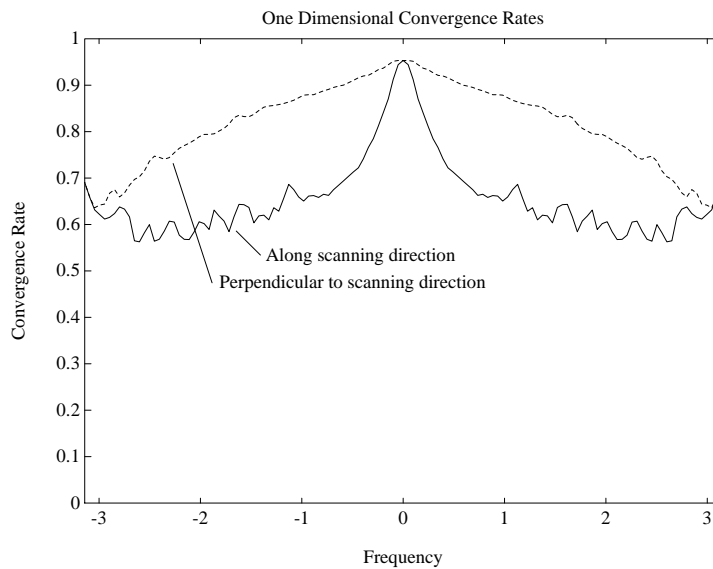


Figure 8: The two one-dimensional convergence rate functions $\max_{\omega_1} |P(\omega_1, \omega_2)|$ and $\max_{\omega_2} |P(\omega_1, \omega_2)|$ when no regularization is used ($\gamma = 0$). The best convergence occurs along the ω_1 direction since this corresponds to the direction of the updates. The large values in the ω_2 direction are caused by the thin main lobe.

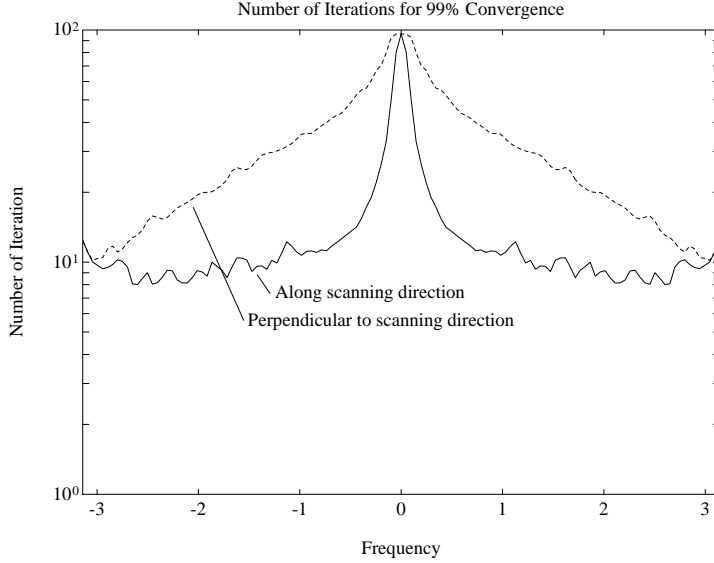


Figure 9: The maximum number of iterations required for 99% error reduction when no regularization is used ($\gamma = 0$). The best convergence occurs along the ω_1 direction since this corresponds to the direction of the updates. The large values in the ω_2 direction are caused by the thin main lobe. This poorer convergence rate may easily be corrected by alternating the direction of updates.

4.2 Gradient Ascent

In order to understand the convergence properties of GA, we rewrite the update equation (22) in the form of a discrete time system.

$$\begin{aligned} f^{(n+1)} &= [I - \alpha(\mathbf{A}^t \mathbf{D} \mathbf{A} + \gamma \mathbf{R})] f^{(n)} + \alpha \mathbf{A}^t \mathbf{D} \hat{p} \\ &= [I - \alpha(\mathbf{H} + \gamma \mathbf{R})] f^{(n)} + \alpha b. \end{aligned} \quad (26)$$

The convergence of this equation depends on the eigenvalues of the matrix multiplying $f^{(n)}$, which in turn depend on the eigenvalues of the matrix $\mathbf{H} + \gamma \mathbf{R}$.

As noted above, multiplication by \mathbf{H} is approximately equivalent to convolution when $\mathbf{D} = \sigma^{-2} \mathbf{I}$. If the pixel lattice is large, the effect of convolution may be approximated by multiplication in the spatial frequency domain. The 2-D DTFT of (26) yields the frequency domain equation

$$f^{(n+1)}(\bar{\omega}) = [1 - \alpha(H(\bar{\omega}) + \gamma R(\bar{\omega}))] f^{(n)}(\bar{\omega}) + \alpha b(\bar{\omega}), \quad (27)$$

where $H(\bar{\omega})$ is the 2-D DTFT of the spatially kernel for \mathbf{H} , and $R(\bar{\omega})$ is as in (11). The convergence properties of (27) are clear. Error in f corresponding to spatial frequency $\bar{\omega}$ will

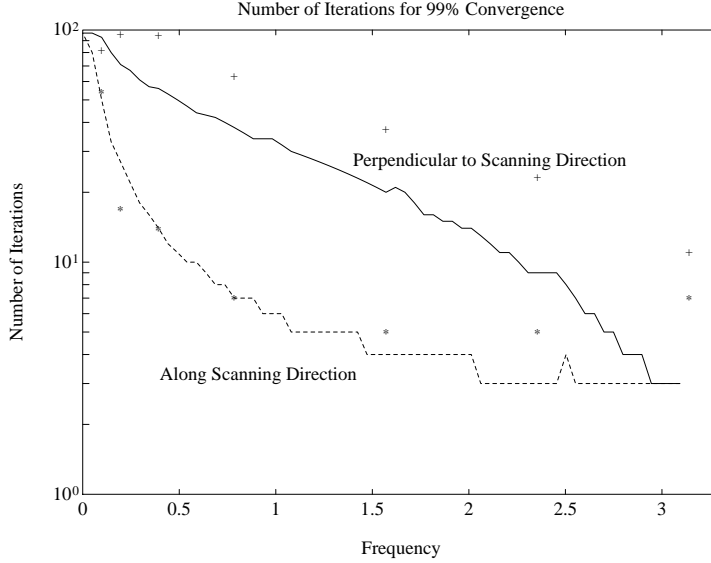


Figure 10: Experimental measurements of the number of iterations required for 99% error reduction when no regularization is used ($\gamma = 0$). Symbols indicate experimental convergence along (*) and perpendicular to (+) scanning direction. Continuous plots are theoretically predicted convergence.

be attenuated or amplified by the factor

$$P(\bar{\omega}) = 1 - \alpha(H(\bar{\omega}) + \gamma R(\bar{\omega}))$$

with each iteration. Again, $H(\bar{\omega})$ and $R(\bar{\omega})$ must be positive functions of $\bar{\omega}$ since they correspond to positive semidefinite matrices.

We may derive an approximate form for $P(\bar{\omega})$ by using a second result of Appendix B. The kernel of \mathbf{H} is approximately proportional to a filtered and sampled version of the function β/d where d is the distance of a pixel from the origin and β is a constant defined in Appendix B. Since the Fourier transform of $1/d$ is $1/||\bar{\omega}||$, the DTFT of the kernel of \mathbf{H} will also be approximately $\beta/||\bar{\omega}||$. This implies that rate of convergence at frequency $\bar{\omega}$ will be approximately given by

$$P(\bar{\omega}) \approx 1 - \alpha \left(\frac{\beta}{||\bar{\omega}||} + \gamma(1 - 1/2 \cos \omega_1 - 1/2 \cos \omega_2) \right). \quad (28)$$

Fig. 11 shows the DTFT of the experimentally measured kernel. Its form corresponds well to the predicted $1/||\bar{\omega}||$ shape. The value at $\bar{\omega} = 0$ is theoretically infinite, but is limited by the size of the window (in this case 128x128).

From (27), we see that stable convergence requires that $|P(\bar{\omega})| \leq 1$ for all $\bar{\omega}$. This

Figure 11: Fourier transform of Bartlett weighted kernel resulting from projection and back-projection of a point. The shape corresponds to the theoretically expected shape of $1/||\bar{\omega}||$.

implies that

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

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. However, small values of α will cause the convergence to be slow for large values of $||\bar{\omega}||$ since the corresponding values of $P(\bar{\omega})$ will be close to 1. The conclusion is that the convergence of GA is poor at high spatial frequencies.

5 Experimental Results

We have applied the methods discussed in Sec. 3 to synthetic phantoms with parameters corresponding to sizes and densities within the range encountered in medical X-ray tomographic imaging. This choice is intended only to attach some physical significance to the function $f(x, y)$, not to suggest primary applications are likely to be medical diagnostics. Experimental data presented here result from very low photon dosages relative to most systems. The input photon rate (λ_T) for each ray in our trials was 2000, orders of magnitude

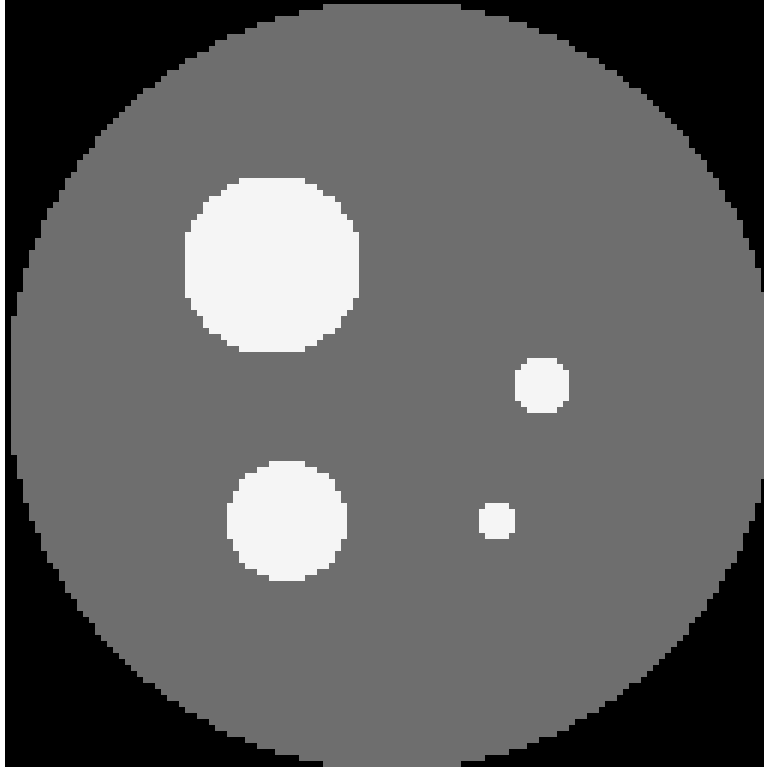


Figure 12: Original phantom at 128×128 resolution. Increasing brightness corresponds to higher X-ray absorptivity.

below dosages of commercial medical CT scanners[1]. Our goals in terms of reconstructed image quality, however, are also much more modest.

We are principally concerned with reconstruction/segmentation from a set of projections uniformly spaced in both θ and t . A CBP reconstruction will serve as a good initial starting point in most cases. Since all the algorithms we will analyze involve iterative computation of projections and/or backprojections, the relative computational expense of CBP is small. The residual error left by the CBP algorithm often has substantial high frequency content. This is a good match for the GS update method, which is shown above to be particularly effective at suppressing high frequency error components.

The phantom is shown in Fig. 12, 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, making Δt equal to the dimension of a pixel. The lowpass filter incorporated into CBP

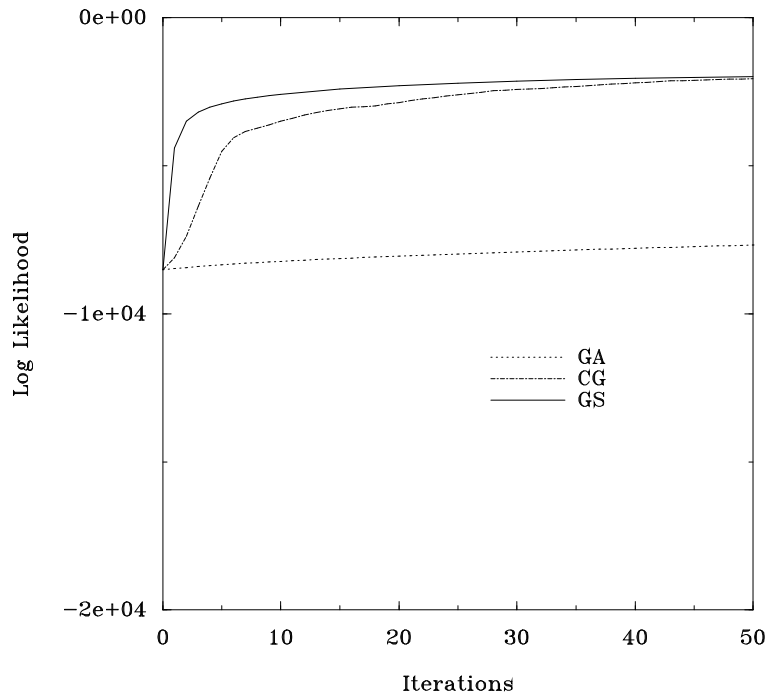


Figure 13: Convergence comparison for constant weighting matrix D

was a raised cosine rolloff in frequency response, with the cutoff frequency chosen for the best visual quality. We chose a single value for α in (22) for all applications of GA, based on the best experimental convergence across the set of simulations. All reconstructions will be presented at a resolution of 128×128 pixels.

Initial trials were used to verify the analysis of Sec. 3, and were therefore performed with $\gamma = 0\text{cm}^2$ in (12), corresponding to ML estimation. Fig. 13 illustrates convergence in terms of log likelihood as expressed in (9), with D diagonal, corresponding to an assumption of stationarity in photon counting noise. In order to avoid slow convergence of the frequencies represented by the ridge in Fig. 7, the orientation of the GS update scan alternated between vertical and horizontal patterns. CG and GS are comparable in convergence rates, with GS showing an appreciable advantage. Nearly 50 iterations are required for the two to reach the same log likelihood value. 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. The error image was of predominantly high frequency spectral content in this case.

When the diagonal of D is filled with photon counts, similar behavior is observed, with

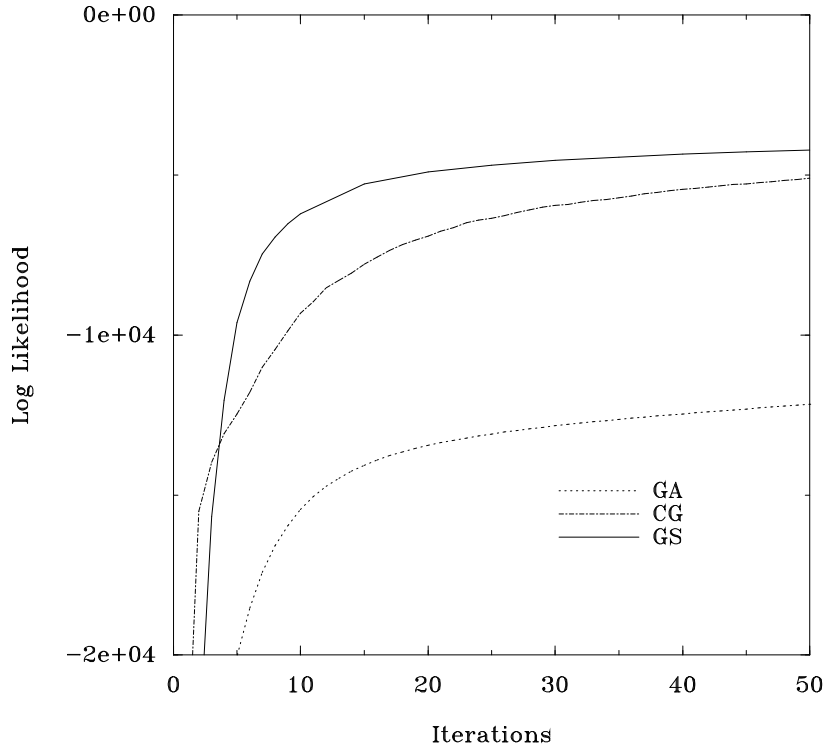


Figure 14: Convergence comparison for real projection weighting matrix D

convergence taking place significantly more slowly (Fig. 14). The slower convergence is likely due to the fact that with wide variation in the entries of D , eigenvalues of $A^t D A$ have similarly greater range, and large differences in the convergence rates among various modes. Here, GS shows an advantage over CG which is not erased after 50 iterations, and both again perform better than GA. This behavior agrees qualitatively with expectations from the analytical results in light of the error, which was again relatively high frequency.

The above simulations support the analysis of the comparative properties of the three iterative approaches. 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. Detail enhancement at higher SNR[34], for example, is a candidate application. But it is the high frequency characteristic in the evolution of the ML estimate which makes it questionable in some settings. ML estimates often have *excessive* high frequency content, in contrast to the images typical of most ensembles of practical interest. The ML estimate images whose convergence was discussed above, in fact, are of very poor visual quality.

Regularization both speeds convergence, and prevents excessive oscillation in the estimate. Typical convergence rates are shown in Fig. 15 for MAP estimation with the same

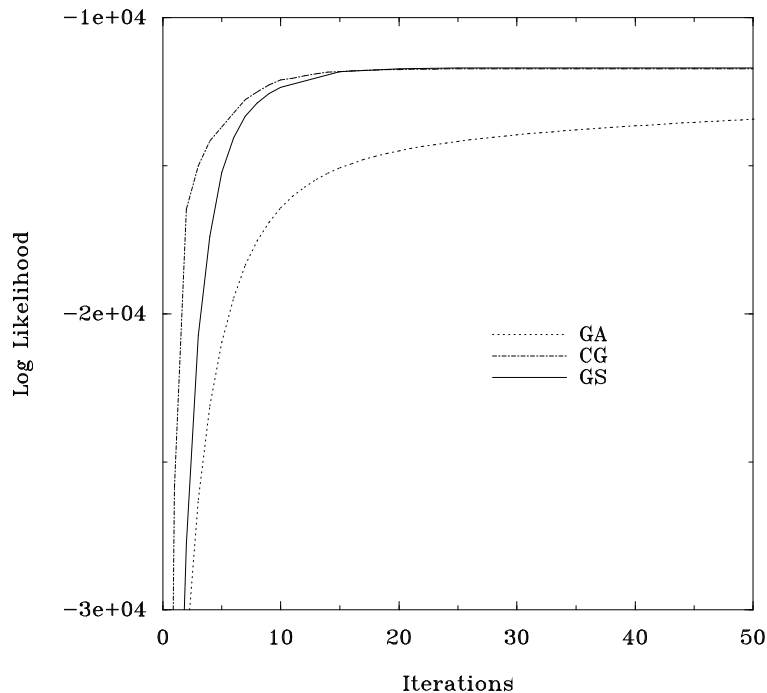


Figure 15: Convergence comparison for real projection weighting matrix D and regularization using a Gaussian prior.

optimization methods and $\gamma = 100\text{cm}^2$. (This corresponds to a standard deviation of a pixel given its neighbors of 0.1cm^{-1} , which is consistent with densities ranging from 0.2cm^{-1} to 0.48cm^{-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. 16 shows the images resulting from the Bayesian enhancement of the CBP reconstruction. The artifacts of 16(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.

Useful tomographic estimates can be achieved from still lower SNR if we can use more restrictive priors on $f(x, y)$. 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. As discussed earlier, this case presents the problem of non-convexity in the log likeli-

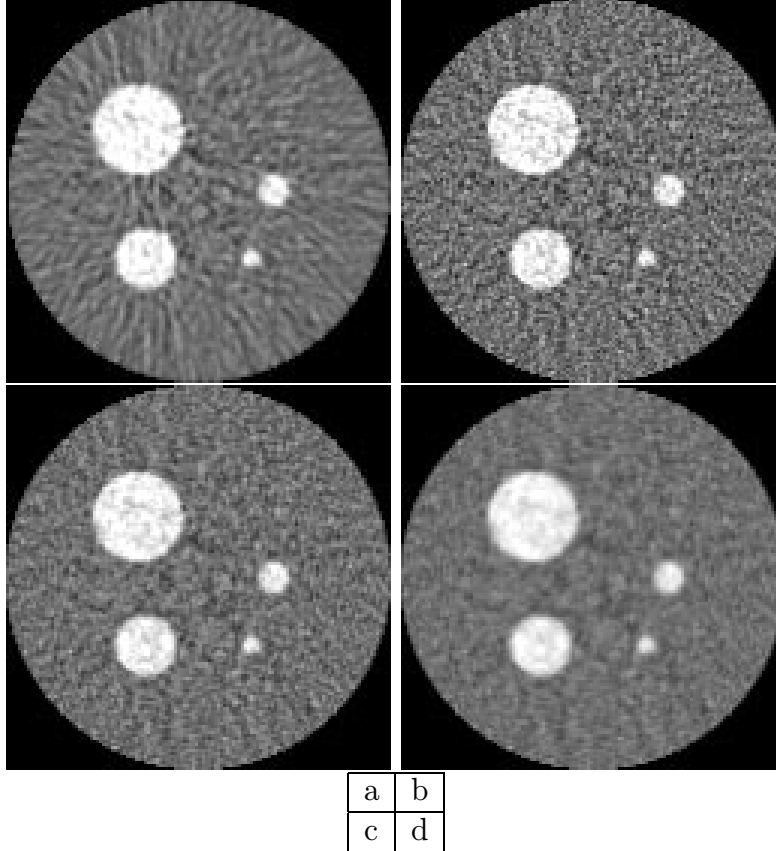


Figure 16: (a) CBP reconstruction from 128×128 noisy projections. (b) Bayesian estimate with $\gamma = 100$. (c) $\gamma = 200$. (d) $\gamma = 500$.

hood function, making true MAP estimation by known methods impractical. The gradient methods are not directly applicable to this problem.

The GS algorithm is well-suited to this problem. With a binary-valued prior for $f(x, y)$, evaluation of (21) is particularly simple. Because lexicographic updates produce directional dependencies, we implement GS in the discrete-valued reconstruction case with a scan consisting of four interlaced 2×2 decimated patterns. Experimental results for a projection data set consisting of 128 rays at each of 16 equally spaced angles appear in the images of Fig. 17. The phantom and dosage per ray are the same as in the previous simulations. The CBP reconstruction using a raised cosine filter with a roll-off factor of 1.0 is shown in 17(b), with the result of thresholding midway between the two known densities in 17(c). After less than 3 full iterations, GS produced the segmentation of 17(d). Nearly all the artifacts resulting from the radio-opaque regions of the object are removed.

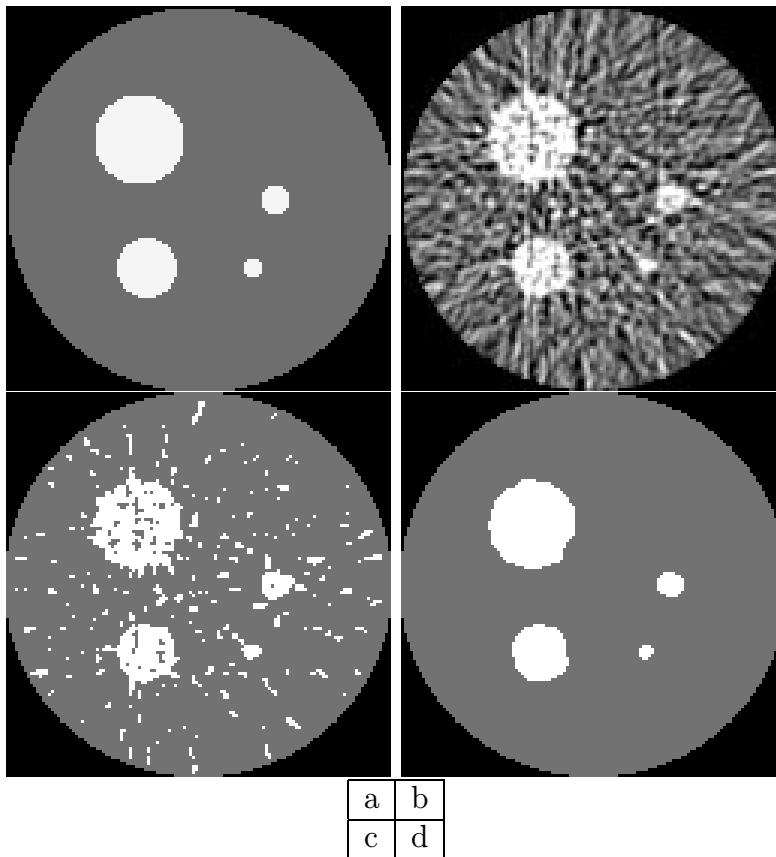


Figure 17: (a) Original phantom. (b) CBP reconstruction from 16×128 noisy projections. (c) Threshold segmentation of CBP image. (d) Gauss-Seidel approximation of MAP estimate.

6 Conclusion

We have proposed, analyzed, and simulated the application of a Gauss-Seidel algorithm for the iterative image reconstruction from projections using Bayesian criteria. We applied the GS method to the problems of reconstructing both continuously valued images and images which were constrained to have discrete values. This latter problem is equivalent to the direct segmentation of the image using the projected data. This paper also includes results, both analytic and experimental, concerning the gradient methods for CT reconstruction which to our knowledge were not previously available.

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[35]. Solving the same optimization problem on a coarser resolution image will effectively shift lower frequencies upward. The interpolated lower resolution image can then serve as a starting point on a finer grid. 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[21, 31], in terms of both computational efficiency and quality of segmentation. We are also currently investigating the use of a more general classes of statistical models which offer possible improvements in reconstruction quality[36].

A Appendix

The quadratic approximation to (6) may be obtained by expanding each term of the sum in a second order Taylor series. This is done using the relation $p_j = \mathbf{A}_{j*}f$, and expanding each term about the value $\hat{p}_j = \log(\lambda_T/\lambda_j)$.

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

$$\begin{aligned}
&\approx -\sum_{j=1}^M [\lambda_j(1 + \hat{p}_j) + \log(\lambda_j!) + \frac{\lambda_j}{2}(p_j - \hat{p}_j)^2] \\
&= -\sum_{j=1}^M [\lambda_j(1 + \hat{p}_j) + \log(\lambda_j!) + \frac{\lambda_j}{2}(\hat{p}_j - \mathbf{A}_{j*}f)^2] \\
&= -\frac{1}{2}\|\hat{\mathbf{p}} - \mathbf{A}f\|_D^2 - \sum_{j=1}^M [\lambda_j(1 + \hat{p}_j) + \log(\lambda_j!)]
\end{aligned}$$

where $D = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_M\}$.

B Appendix

The continuous-variable mapping from Radon transform space into image space analogous to the matrix operator \mathbf{A}^t , is the backprojection operation given by

$$\begin{aligned}
g(u, v) &= \int_0^\pi \int_{-\infty}^\infty p(\theta, t) \delta(t - u \cos \theta - v \sin \theta) dt d\theta \\
&= \int_0^\pi \int_{-\infty}^\infty p(\theta, t) h(\theta, t, u, v) dt d\theta .
\end{aligned}$$

The function $h(\theta, t, u, v)$ is the kernel of both the projection and backprojection transformations. The kernel for the combined projection, backprojection operation may be computed as

$$\begin{aligned}
k(x, y, u, v) &= \int_0^\pi \int_{-\infty}^\infty h(\theta, t, x, y) h(\theta, t, u, v) dt d\theta & (29) \\
&= \int_0^\pi \int_{-\infty}^\infty \delta(t - x \cos \theta - y \sin \theta) \delta(t - u \cos \theta - v \sin \theta) dt d\theta \\
&= \int_0^\pi \int_{-\infty}^\infty \delta(t - (x-u) \cos \theta - (y-v) \sin \theta) \delta(t) dt d\theta \\
&= \int_0^\pi \delta((x-u) \cos \theta + (y-v) \sin \theta) d\theta \\
&= \int_0^\pi \delta(d \cos(\phi - \theta)) d\theta \\
&= \frac{1}{d}
\end{aligned}$$

where $(x-u) = d \cos \phi$ and $(y-v) = d \sin \phi$. Gilbert[37] derived a similar result with frequency-domain methods.

Given this result, we would expect that the matrix $\mathbf{A}^t \mathbf{A}$ should have an approximately Toeplitz-block-Toeplitz form with a kernel proportional to $1/d$. Specifically, if $\{\theta_i\}_{i=1}^{M_1}$ and $\{t_j\}_{j=1}^{M_2}$ are the values of angle and displacement for which the projections are computed,

then the components of \mathbf{A} may be written $\mathbf{A}(i, j, m, n)$ where i and j are the angle and displacement index for the projections of pixel (m, n) . (Note the difference between the present use of subscripts i and j , and that in the main text.) We will also assume that the sampling in both θ and t is uniform with periods $\Delta\theta$ and Δt respectively, and that the spatial sampling period is T .

$$\begin{aligned}
\mathbf{A}^t \mathbf{A} &= \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} \mathbf{A}(i, j, m, n) \mathbf{A}(i, j, k, l) \\
&= \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} \int_{\mathbb{R}^2} w_{(m,n)}(x, y) h(\theta_i, t_j, x, y) dx dy \int_{\mathbb{R}^2} w_{(k,l)}(s, u) h(\theta_i, t_j, s, u) ds du \\
&\approx \frac{1}{\Delta\theta \Delta t} \int_0^\pi \int_{\mathbb{R}} \int_{\mathbb{R}^2} w_{(m,n)}(x, y) h(\theta, t, x, y) dx dy \int_{\mathbb{R}^2} w_{(k,l)}(s, u) h(\theta, t, s, u) ds du dt d\theta \\
&= \frac{1}{\Delta\theta \Delta t} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} w_{(m,n)}(x, y) w_{(k,l)}(s, u) \int_0^\pi \int_{\mathbb{R}} h(\theta, t, x, y) h(\theta, t, s, u) dt d\theta dx dy ds du \\
&= \frac{1}{\Delta\theta \Delta t} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} w_{(m,n)}(x, y) w_{(k,l)}(s, u) \frac{1}{\sqrt{(x-s)^2 + (y-u)^2}} dx dy ds du
\end{aligned}$$

The third equation approximates the sums by integrals, and the last equation uses the result of (29).

To establish that this form of $\mathbf{A}^t \mathbf{A}$ is Toeplitz-block-Toeplitz, we must use the fact of (4) that the functions $w_{(m,n)}(x, y)$ are simply shifted versions of a single kernel. Then writing $\sqrt{(x-s)^2 + (y-u)^2}$ as d we have

$$\begin{aligned}
\mathbf{A}^t \mathbf{A} &\approx \frac{1}{\Delta\theta \Delta t} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} w_{(0,0)}(x-mT, y-nT) w_{(0,0)}(s-kT, u-lT) \frac{1}{d} dx dy ds du \\
&= \frac{1}{\Delta\theta \Delta t} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} w_{(0,0)}(x, y) w_{(0,0)}(s-(k-m)T, u-(l-n)T) \frac{1}{d} dx dy ds du
\end{aligned}$$

The last integral has the form of two 2-D convolutions. This may be written formally as

$$\mathbf{A}^t \mathbf{A} \approx \frac{1}{\Delta\theta \Delta t} (w_{(0,0)} * w_{(0,0)} * \frac{1}{d})((k-m)T, (l-n)T). \quad (30)$$

So the kernel of $\mathbf{A}^t \mathbf{A}$ is formed by filtering the continuous function $1/d$ and then sampling the result. Since the components of the matrix $\mathbf{A}^t \mathbf{A}$ only depend on the differences, $k-m$ and $l-n$, the matrix must be approximately Toeplitz-block-Toeplitz. The accuracy of this approximation is affected only by Δt , the period of the sampling in t .

If the pixels (m, n) and (k, l) are distant, then the integrand will be approximately constant, and the following approximation holds

$$\mathbf{A}^t \mathbf{A}(m, n, k, l) \approx \frac{\beta}{\sqrt{(m-k)^2 + (n-l)^2}}$$

where

$$\beta = \frac{\int_{\mathbb{R}^2} w_{(m,n)}(x,y) dx dy}{\Delta\theta \Delta t T}.$$

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] P. Oskoui-Fard and H. Stark, "Tomographic Image Reconstruction Using the Theory of Convex Projections," *IEEE Trans. Med. Imaging*, vol. MI-7, no.1 , pp. 45-58, March 1988.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] D. M. Young, *Iterative Solution of Large Linear Systems*, Academic Press, New York, 1971.
- [12] A.C. Kak, "Computerized Tomography with X-Ray, Emission, and Ultrasound Sources," *Proc. of the IEEE*, vol 67, pp. 1245-1272, 1979.
- [13] S.R. Deans, *The Radon Transform and Some of Its Applications*, John Wiley and Sons, New York, 1983.
- [14] L.A. Shepp and B.F. Logan, "The Fourier Reconstruction of a Head Section," *IEEE Trans. Nucl. Sci.*, vol. NS-21, pp. 21-43, June 1974.
- [15] E. Tanaka and T.A. Iinuma, "Correction Functions for Optimizing the Reconstructed Image in Transverse Section Scan," *Phys. Med. Biol.*, Vol. 20, No. 5, pp. 789-798, 1975.

- [16] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*, Winston and Sons, New York, 1977.
- [17] H. Derin, H. Elliot, R. Cristi, and D. Geman, "Bayes Smoothing Algorithms for Segmentation of Binary Images Modeled by Markov Random Fields," *IEEE Trans. Trans. Pattern Anal. and Mach. Intell.*, vol. PAMI-6, no.6 , pp. 707-720, Nov. 1984.
- [18] 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.
- [19] 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.
- [20] J. Besag, "On the Statistical Analysis of Dirty Pictures," *J. Roy. Statist. Soc. B*, vol. 48, no. 3, pp. 259-302, 1986.
- [21] C. Bouman and B. Liu, "Multiple Resolution Segmentation of Textured Images," *IEEE Trans. on Pat. An. Mach. Intell.*, vol. 13, no. 2, pp. 99-113, Feb. 1991.
- [22] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *J. Royal Stat. Soc. B*, vol. 36, pp. 192-326, 1974.
- [23] R. Kindermann and J. L. Snell, *Markov Random Fields and their Applications*. Providence: American Mathematical Society, 1980.
- [24] R. Kashyap and R. Chellappa, "Estimation and Choice of Neighbors in Spatial-Interaction Models of Images," *IEEE Trans. Info. Theory*, vol. IT-29, no. 1, pp. 60-72, Jan. 1983.
- [25] D. Pickard, "Inference for Discrete Markov Fields: The Simplest Nontrivial Case," *Journal of the Amer. Stat. Assoc.*, vol. 82, pp. 90-96, March 1987.
- [26] R. Dubes and A. Jain, "Random Field Models in Image Analysis," *Journal of Applied Statistics*, vol. 16, no. 2, pp. 131-164, 1989.
- [27] J. Besag, "Nearest-neighbor Systems and the Auto-logistic Model for Binary Data," *J. Roy. Statist. Soc. B*, vol. 34, no. 1, pp. 75-83, 1972.
- [28] A. Jain, "Advances in Mathematical Models for Image Processing," *Proc. of the IEEE*, vol. 69, pp. 502-528, May 1981.
- [29] R. Kashyap, R. Chellappa and A. Khotanzad, "Texture Classification Using Features Derived from Random Field Models," *Pattern Recogn. Let.*, vol. 1, no. 1, pp. 43-50, Oct. 1982.
- [30] N. Metropolis, A.W. Rosenbluth, A.H. Teller, and E. Teller, "Equations of State Calculations by Fast Computing Machines," *J. Chem. Phys.*, vol. 21, pp. 1087-1091, 1953.
- [31] C. Bouman and B. Liu, "Segmentation of Textured Images Using a Multiple Resolution Approach," *Proc. IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, pp. 1124-1127, New York, NY, April 11-14, 1988.

- [32] F. Beckman, "The Solution of Linear Equations by the Conjugate Gradient Method," in eds. A. Ralston, H. Wilf and K. Enslein, *Mathematical Methods for Digital Computers*, Wiley, 1960.
- [33] W. Press, B. Flannery, S. Teukolsky and Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1988.
- [34] K. Sauer and B. Liu, "Statistical Refinement of Transmission Computed Tomograms in High Photon Counting Noise," *Proc. SPIE Conf. on Vis. Comm. and Image Proc.*, Philadelphia, PA, Nov. 8-10, 1989, pp. 944-951.
- [35] W.L. Briggs, *A Multigrid Tutorial*, Society for Industrial and Applied Mathematics, Philadelphia, 1987.
- [36] C.A. Bouman and K.D. Sauer, "An Edge Preserving Method for Image Reconstruction from Integral Projections," *Proc. Conf. on Inf. Sci. and Sys.* Johns Hopkins University, Baltimore, MD, Mar. 20-22, 1991, pp. 382-387.
- [37] P.F.C. Gilbert, "The Reconstruction of a Three-Dimensional Structure from Projections and Its Application to Electron Microscopy," *Proc. Royal Soc. London B*, vol. 182, pp. 89-102, 1972.