Optimal Resolution in Maximum Entropy Image Reconstruction from Projections with Multigrid Acceleration

Mark A. Limber
Tom A. Manteuffel
Stephen F. McCormick

Computational Mathematics Group
University of Colorado at Denver
Denver, CO.

David S. Sholl

Program in Applied Mathematics
University of Colorado at Boulder
Boulder, CO.

Abstract

We consider the problem of image reconstruction from a finite number of projections over the space $L^1(\Omega)$, where Ω is a compact subset of $\mathbb{R}^2$. We prove that, given a discretization of the projection space, the function that generates the correct projection data and maximizes the Boltzmann-Shannon entropy is piecewise constant on a certain discretization of Ω, which we call the "optimal grid". It is on this grid that one obtains the maximum resolution given the problem setup. The size of this grid grows very quickly as the number of projections and number of cells per projection grows, indicating fast computational methods are essential to make its use feasible.

We use a Fenchel duality formulation of the problem to keep the number of variables small while still using the optimal discretization, and propose a multilevel scheme to improve convergence of a simple cyclic maximization scheme applied to the dual problem.

1 Introduction

In computerized tomography (CT), one encounters the problem of reconstructing an image, or a density, defined by the function $\hat{x}(s,t)$, given only a finite number of projection data. General references include [1] and [2].

The projection data is typically of the form

$$b_k^m = \int_{\Omega_k^m} \hat{x}(s,t)ds\,dt \tag{1.1}$$

where the support of $\hat{x}$ is assumed to lie in the bounded region $\Omega \subset I\!R^2$ and Ω_k^m is the k^{th} strip orthogonal to the m^{th} projection (see Figure 1). We assume that there are M projections and that the m^{th} projection has K_m cells. Let b be the vector of projection data, N the length of b, and ψ_k^m the characteristic function of Ω_k^m. We then rewrite (1.1) as

$$b = A\hat{x}, \tag{1.2}$$

where $A : L^1(\Omega) \to I\!R^N$ via

$$(Ax)_{k,m} = \int_\Omega x(s,t)\psi_k^m(s,t)ds\,dt. \tag{1.3}$$

The reconstruction problem we study is: given the projection data b, find a density function x such that $Ax = b$. Since A has an infinite-dimensional kernel, solutions, if they exist, are not unique. The problem then becomes: find the "best" function x_0 such that $Ax_0 = b$. The concept of "best" is ambiguous to be sure, but some criteria have been gaining acceptance. In this paper we choose to study the solution with maximum entropy as defined by Shannon [3] in information theory. For an informal discussion of entropy and information theory, see for example [4]. For a discussion of maximum entropy in image reconstruction, see [5].

In our context, we wish to find the function $x_0 \in L^1(\Omega)$ such that x_0 attains

$$\sup\left\{-\int_\Omega x(s,t)\ln[x(s,t)]ds\,dt : Ax = b\right\}. \tag{1.4}$$

This is the maximum entropy solution to $Ax = b$. Simply because we would rather minimize a convex function than maximize a concave function, we rewrite this as a convex minimization program via

$$p = \inf\left\{\int_\Omega x(s,t)\ln[x(s,t)]ds\,dt : Ax = b\right\}. \tag{1.5}$$

This is called the *primal problem*. If we further define the function $\phi : I\!R \to (-\infty, +\infty]$ by

$$\phi(u) = \begin{cases} u\ln u & u > 0 \\ 0 & u = 0 \\ +\infty & u < 0 \end{cases}, \tag{1.6}$$

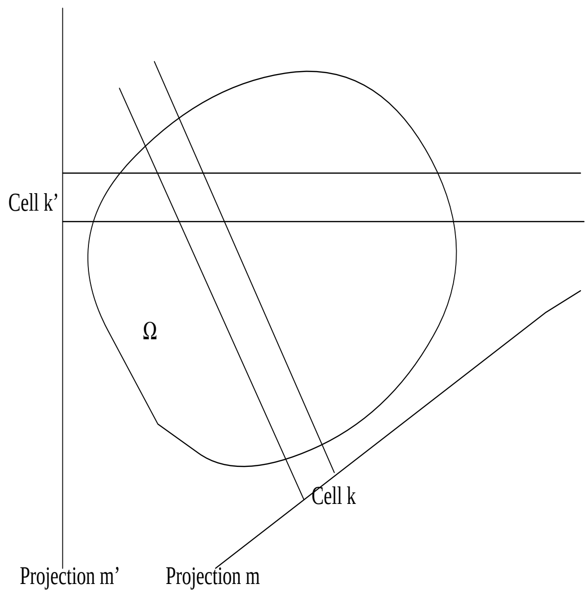

Figure 1: The geometry setup.

then we can rewrite (1.5) as

$$p = \inf\left\{ \int_\Omega \phi(x(s,t))ds\,dt : Ax = b \right\}, \tag{1.7}$$

which is in a form that has been receiving much attention in the optimization community of late. In particular, see Borwein and Lewis[6]. One of the features of this functional is that it forces feasible functions to be nonnegative, as a density or image function should be. We shall see that it has other properties that are computationally and theoretically attractive, one of the most important being that solutions exist in $L^1(\Omega)$ under rather mild conditions.

In this paper we shall characterize solutions to (1.5) given some reasonable conditions on the data, b, and show that the solution in $L^1(\Omega)$ for the CT case is piecewise constant, but usually not on a rectangular grid. While this result seems to be known in the tomography community, it is rare that one finds a mathematically sound derivation of the solution. The first half of this paper discusses the difficulties in addressing this problem and references the literature to outline a correct proof of our characterization. Note that we do *not* initially impose a discretization of $L^1(\Omega)$ or Ω, only of the data. The discretization we shall use arises as a consequence of the form of the functions ψ_k^m.

The second half of this paper discusses implementation details. It will turn out that the appropriate grid for optimal resolution (the "optimal grid") is very large compared to the amount of data, N, one has. We shall also see that finding the optimal function can be reduced to solving a problem in $I\!R^N$, but the intermediate calculations require use of the (large) optimal grid. We have found that a simple cyclic coordinate maximization scheme applied to the Fenchel dual of (1.5), while convergent, tends to stall after a few iterations. We describe a multigrid approach to accelerate convergence.

2 Maximum Entropy Solutions

2.1 The Entropy Functional

Using the entropy functional

$$f : L^1(\Omega) \to (-\infty, +\infty] : \quad x \mapsto -\int_\Omega \phi(x(s,t))ds\, dt \tag{2.1}$$

to pick a "best" density function x has been popularized by Shannon in information theory, but also arises in the context of thermodynamics with Boltzmann. For this reason we call this the Boltzmann-Shannon entropy.

Entropy is, in short, the expected amount of information present in a probability density x. In our context, one can think of an image (appropriately scaled) as a probability density function, and computing the feasible density with maximum entropy yields the density carrying the most information. The standard reference is Shannon and Weaver [3], but many basic probability books contain some discussion of information and entropy (see [4, 7] for example). References that deal specifically with entropies in image reconstruction are [5, 8].

We remark that the theory and methods developed here apply directly to other objective functionals, in particular, minimum L^2–norm solutions. In fact, with the minimum L^2–norm functional, our iterative method is essentially only changed by replacing $*$ by $+$ and $/$ by $-$.

2.2 Existence of Solutions

In this section we prove that solutions to (1.5) exist. Usually, this point is ignored, but is nevertheless an important issue.

Throughout, let X be a linear normed space with topology τ. We begin with some definitions.

Definition 2.1 *We say a set $K \subseteq X$ is* τ-sequentially compact *if every sequence from K has a τ-convergent subsequence.*

Definition 2.2 *Given a function $f : X \to (-\infty, +\infty]$, for $\alpha \in I\!R$, we define the* lower level sets L_α *of f to be*

$$L_\alpha = \{x : f(x) \leq \alpha\}. \tag{2.2}$$

The following is a general existence theorem for solutions to constrained optimization problems.

Theorem 2.3 *Let f be τ-lower semicontinuous (lsc) possessing τ-sequentially compact lower level sets, and let C be a τ-closed set. Then*

$$p = \inf \{f(x) : x \in C\} \tag{2.3}$$

is attained for some $x_0 \in C$.

Proof: Let $\{x_n\} \subset C$ be a sequence such that $f(x_n) \to p$. Letting $\alpha = p + 1$, eventually all $x_n \in L_\alpha$, for $n \geq N$. By τ-sequential compactness of L_α there is a subsequence x_{n_k} that converges to a point $x_0 \in L_\alpha$. Since C is τ-closed, then $x_0 \in C$; f is τ-lsc, so

$$p = \lim_{k\to\infty} f(x_{n_k}) \geq f(x_0) \geq p \tag{2.4}$$

and, thus, $p = f(x_0)$. ∎

We are interested in the case where the τ-topology is the weak topology on $L^1(\Omega)$.

Let $C = \{x : Ax = b\}$ for $A : X \to I\!R^N$ linear and continuous. Then $C = A^{-1}(\{b\})$ is closed (since A is continuous) and convex (since A is linear) and, hence, weakly closed. This is called Mazur's theorem; see [9, Corollary 4 Chapter 2], for example.

We now direct our attention to the functional

$$f(x) = \int_\Omega x(s,t) \ln[x(s,t)] ds\, dt. \tag{2.5}$$

From [10, Theorem 2.2], we see that f is weakly lsc with weakly compact lower level sets, provided Ω is of finite measure. To apply Theorem 2.3, we need weakly sequentially compact lower level sets. The Eberlein-Šmulian theorem [9] states that a subset of a Banach space is weakly compact if and only if it is weakly sequentially compact, and thus we see that the lower level sets of f are weakly sequentially compact, so we can apply Theorem 2.3 to obtain the following:

Theorem 2.4 *With f defined as in (2.5) and $C = \{x : Ax = b\}$ for $A : L^1(\Omega) \to I\!R^N$ linear and continuous, the infimum*

$$p = \inf \{f(x) : Ax = b\} \tag{2.6}$$

is attained for some $x_0 \in L^1(\Omega)$ such that $Ax_0 = b$.

Note that if $X = L^\infty$, then the level sets L_α of f are not bounded in the norm topology. Hence L_α is not compact for any of the norm, weak or weak-$*$ topologies, which is a consequence of the fact that a continuous function attains its maximum on a compact set, of the theory of dual pairs[11], and of the principle of uniform boundedness, respectively. Thus, although it is tempting to approach the image reconstruction problem in L^∞, the initial problem of existence is much more difficult. However we will see that L^1-optimal solutions are actually L^∞-optimal solutions as well. We will also see why one might want to pose the problem in L^∞.

2.3 Uniqueness of Solutions

In this section we show that solutions to (1.5) are unique.

Definition 2.5 *A function $g : X \to (-\infty, +\infty]$ is* convex *if, for all $x, y \in X$ and all $\lambda \in (0,1)$, we have*

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y). \tag{2.7}$$

Also, g is strictly convex *if, whenever $x \neq y$ and $g(x)$, $g(y) \in I\!R$, this inequality is strict. A set $C \subseteq X$ is* convex *if, whenever $x, y \in C$ and $\lambda \in (0,1)$, then*

$$\lambda x + (1-\lambda)y \in C. \tag{2.8}$$

Lemma 2.6 *If $f(x) = \int \phi(x(u))du$, then f is strictly convex if and only if ϕ is strictly convex.*

Proof: Assume that f is strictly convex and that there exists a $u \neq v$ and a $\lambda \in (0,1)$ such that

$$\phi(\lambda u + (1-\lambda)v) \geq \lambda\phi(u) + (1-\lambda)\phi(v). \tag{2.9}$$

Then let $x(t) = u$ and $y(t) = v$, so that

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y), \tag{2.10}$$

contradicting the assumption that f is strictly convex.

Conversely, let $E = \{t : x(t) \neq y(t)\}$ and assume that $m(E) > 0$. Then

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \int_E \phi(\lambda x(t) + (1-\lambda)y(t)dt + \int_{E^c} \phi(x(t))dt & (2.11)\\ &< \int_E \lambda\phi(x(t)) + (1-\lambda)\phi(y(t))dt + \int_{E^c} \phi(x(t))dt & (2.12)\\ &= \lambda f(x) + (1-\lambda)f(y) & (2.13) \end{aligned}$$

where the strict inequality is due to the strict convexity of ϕ. ■

It is easy to check that $\phi(u)$ as defined in (1.6) is strictly convex, thus f from (2.5) is as well.

Theorem 2.7 *If f is strictly convex and C is a convex set, then solutions to*

$$p = \inf\{f(x) : x \in C\} \tag{2.14}$$

are unique, provided they exist.

Proof: Suppose $p = f(x) = f(y)$, where $x \neq y \in C$. Since C is convex, then $\frac{1}{2}x + \frac{1}{2}y \in C$ and

$$f(x/2 + y/2) < \frac{1}{2}f(x) + \frac{1}{2}f(y) = p \tag{2.15}$$

contradicting the definition of p. ■

Putting together these results we have the following.

Theorem 2.8 *If Ω is a set of finite measure, then the solution to (1.5) exists and is unique.*

2.4 Characterizing the Solution

In this section we characterize solutions to

$$p = \inf \left\{ \int \phi(x(s,t)) ds\, dt : \int x\psi_k^m = b_k^m, \quad m = 1, \ldots, M, \quad k = 1, \ldots, K_m \right\}, \tag{2.16}$$

where

$$\phi(u) = \begin{cases} u \ln u - u & u > 0 \\ 0 & u = 0 \\ +\infty & u < 0. \end{cases} \tag{2.17}$$

This is the image reconstruction problem we introduced in Section 1. We have included a linear factor in the objective functional, however if we assume that the projections cover the image, then this factor does not change the solution; it only simplifies certain formulæ. A typical approach is to attach a Lagrange multiplier λ to the constraints and differentiate the Lagrangian at the optimal x_0 (which we now know exists) to obtain

$$x_0(s,t) = (\phi')^{-1}\left(A^T\lambda(s,t)\right) = \exp\left(A^T\lambda(s,t)\right), \tag{2.18}$$

where

$$A^T : L^\infty(\Omega) \to I\!R^N \text{ via } \left(A^T\lambda\right)(s,t) = \sum_{m=1}^{M}\sum_{k=1}^{K_m} \lambda_k^m \psi_k^m(s,t). \tag{2.19}$$

However, the functional $f(x) = \int \phi(x(s,t)) ds\, dt$ is *not* differentiable. Indeed, $f = +\infty$ on a dense subset of $L^1(\Omega)$ and is therefore not even continuous. It is for this reason that some people have chosen to work in $C(\Omega)$ or $L^\infty(\Omega)$ [12, 13] where f is differentiable, but the question of existence and attainment is much more difficult there.

A correct approach is to use Fenchel duality with a constraint qualification (CQ). While the classical CQ fails to apply in our example, in [6] a CQ is developed that does apply to CT. Heuristically, we assume a multiplier $\overline{\lambda}$ exists so that

$$
\begin{aligned}
p &= \inf_x \left\{ f(x) + \langle b - Ax, \overline{\lambda} \rangle \right\} &\quad (2.20)\\
&= \langle b, \overline{\lambda} \rangle + \inf_x \left\{ f(x) - \langle x, A^T \overline{\lambda} \rangle \right\} &\quad (2.21)\\
&= \langle b, \overline{\lambda} \rangle - \sup_x \left\{ \langle x, A^T \overline{\lambda} \rangle - f(x) \right\}. &\quad (2.22)
\end{aligned}
$$

For a convex function $g : X \to (-\infty, +\infty]$ we define $g^* : X^* \to (-\infty, +\infty]$ by

$$g^*(y) = \sup_x \left\{ \langle x, y \rangle - g(x) \right\}. \tag{2.23}$$

This is called the Fenchel conjugate of g at y. Using this definition, we see from (2.22)

$$p = \langle b, \overline{\lambda} \rangle - f^* \left(A^T \overline{\lambda} \right). \tag{2.24}$$

Now, for any λ,

$$\inf_x \left\{ f(x) + \langle b - Ax, \lambda \rangle \right\} = \langle b, \lambda \rangle - f^* \left(A^T \lambda \right) \le \inf \left\{ f(x) : Ax = b \right\} = p. \tag{2.25}$$

Therefore, if $\overline{\lambda}$ exists, then

$$p = \max_{\lambda \in I\!R^N} \left\{ \langle b, \lambda \rangle - f^*(A^T \lambda) \right\}. \tag{2.26}$$

This is the Fenchel dual of (1.5). In our problem, it can be shown [14] that for $y \in L^\infty(\Omega)$

$$f^*(y) = \int \phi^*(y(s,t)) ds\, dt, \tag{2.27}$$

that is, the conjugate of the integral functional f is given as an integral function of ϕ^*. From [6], if $\exists \tilde{x} \in L^1(\Omega)$ where $\tilde{x} > 0$ a.e., $f(\tilde{x}) \in I\!R$ and $A\tilde{x} = b$ (the constraint qualification), then a Lagrange multiplier $\overline{\lambda}$ does exist. Also, if $\overline{\lambda}$ solves (2.26), then the optimal $x_0(s,t)$ for (1.5) is

$$x_0(s,t) = (\phi^*)' \left(A^T \overline{\lambda}(s,t) \right). \tag{2.28}$$

In the case ϕ is of the form (2.17), it is easy to show that $\phi^*(v) = e^v$, and we get

$$x_0(s,t) = \exp \left(A^T \overline{\lambda}(s,t) \right), \tag{2.29}$$

where $\overline{\lambda}$ solves (2.26). This matches the heuristic derivation, but this is no accident since in fair generality, $(\phi')^{-1} = (\phi^*)'$. Thus, solving the image reconstruction problem, (1.5), is equivalent to solving the dual problem (2.26), which is an unconstrained finite-dimensional differentiable concave maximization problem.

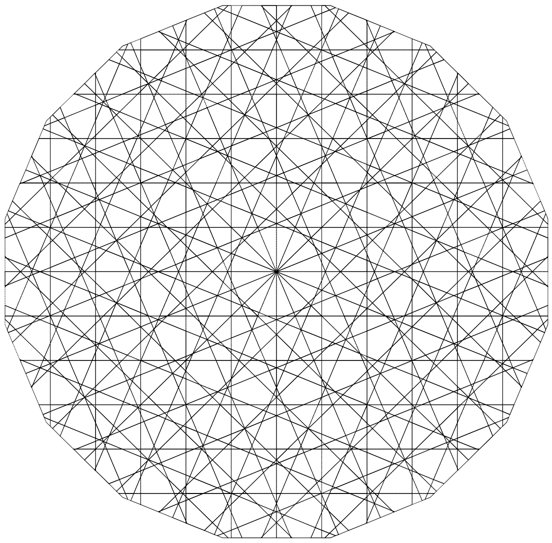

Figure 2: The optimal grid.

2.5 The Optimal Grid

Rewriting the solution to the image reconstruction problem (2.29) in a more explicit form, we have

$$x_0(s,t) = \exp\left(\sum_{m=1}^{M}\sum_{k=1}^{K_m}\overline{\lambda}_k^m\,\psi_k^m(s,t)\right), \tag{2.30}$$

see (2.19). It can be seen from the concavity of (2.26) that any function x of the form (2.30) that satisfies $Ax = b$ must be optimal.

Recall that ψ_k^m is the characteristic function of the k^{th} strip emanating from the m^{th} projection. Thus, the solution in (2.30) implies that the optimal function in $L^1(\Omega)$ is piecewise constant on the grid obtained by intersecting all of the strips. This grid has been observed for physical reasons, [8], but here we have shown that the best (from maximum entropy considerations) function from $L^1(\Omega)$ is this piecewise constant function. For this reason, we call this discretization of Ω the *optimal grid*.

A typical grid is shown in Figure 2. Here we have used 8 projections each divided into 12 cells. The central point of the theory presented above is that the exact solution of the image reconstruction problem is piecewise constant on this grid.

3 Implementation

3.1 Relaxation

Now we develop a simple iterative procedure to solve the image problem by solving the dual problem (2.26) for the optimal λ. Recall that to solve the dual problem we seek to maximize

$$G(\lambda) = \langle b, \lambda \rangle - \int \exp(A^T \lambda(s,t)) ds\, dt. \tag{3.1}$$

This is a simple matter once one observes that G is concave and differentiable, so maximizing G is just finding critical points. The derivative with respect to λ is

$$\nabla G(\lambda) = b - A \int \exp(A^T \lambda(s,t)) ds\, dt. \tag{3.2}$$

This gives N equations for the N unknown λ's.

An obvious iterative scheme is to cycle through the components, λ_k, of ∇G, correcting each λ_k in turn so that the k^{th} component of ∇G is zero, that is, so that G is maximal with respect each individual variable. We choose λ_k^m so that

$$b_k^m = \left(A \exp(A^T \lambda)\right)_{k,m} = \int_{\Omega_k^m} \exp(A^T \lambda(s,t)) ds\, dt. \tag{3.3}$$

After some simplifications, a single step of this scheme can be written as

$$\exp(\lambda_{k'}^{m'}) \leftarrow \frac{b_{k'}^{m'}}{\int_{\Omega_{k'}^{m'}} \exp\left(\sum' \lambda_k^m \psi_k^m\right)} = \frac{b_{k'}^{m'}}{\int_{\Omega_{k'}^{m'}} \prod' (\mu_k^m)^{\psi_k^m}}, \tag{3.4}$$

where $\sum'$ is the sum over all projections m and cells k except m' and k', $\prod'$ is similarly defined and $\mu_k^m = \exp(\lambda_k^m)$.

Now, because $A^T\lambda$ is piecewise constant on the optimal grid, the integral of $\exp(A^T\lambda)$ along any strip is just a sum over each polygon in that strip of the area of the polygon times the product of the μ's that correspond to the particular cell from each projection that makes up the polygon:

$$\int_{\Omega_{k'}^{m'}} \prod' (\mu_k^m)^{\psi_k^m} = \sum_{p \in \Omega_k^m} \left[\text{area}(p) \prod_{m \neq m'} \mu_{k_m}^m \right], \tag{3.5}$$

where the product is only over the cells k_m in projection m for polygon p.

The point of this discussion is twofold. First, we can see from (3.4) that we never need to exponentiate since we only need the μ's. Second, the areas of the polygons can be precomputed, meaning that these integrals can be calculated exactly; no numerical integration is needed.

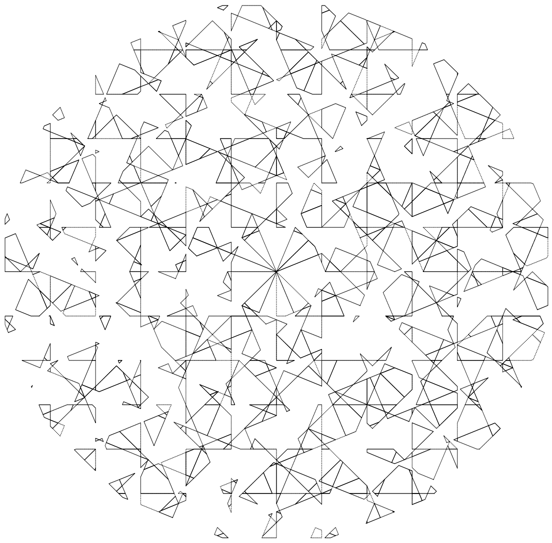

Figure 3: A subset of the optimal grid.

An important feature of this development is that there are no approximations or discretizations being made in the whole program, except for the initial discretization of the data into the vector b. We have characterized the $L^1(\Omega)$ solution as a piecewise constant function on the optimal grid, and the necessary integrals can be performed exactly on this grid. For this reason, in our implementation we calculate the areas of each of the polygons in the optimal grid; in fact, to obtain Figure 2, we in fact plotted the polygons themselves, not just intersecting lines. To demonstrate this fact, in Figure 3 we plotted 1000 of the 2096 polygons from Figure 2.

As can be observed in Figure 2, the number of polygons in the optimal grid can be very large compared to the number of projections and cells per projection. This number grows very quickly as a function of these two variables, for example, with 12 projections placed uniformly around the circle and 10 cells per projection, we generate 2904 polygons; with 20 cells per projection we generate 12,561 polygons. While the optimal grid generation is a time and storage intensive procedure, given a fixed geometry we need only run this part of the program once. To reconstruct images using such large data sets, fast methods are essential.

The iterative method described above tends to stall after a few iterations. This effect can be observed in the reported data from [8]. In Figure 4 the top two curves represent the rates of convergence using this scheme on a 3 projection, 4 cell per projection problem. Here we have graphed both the rate associated with the residual $||Ax - b||_\infty$ and the rate of convergence of the entropies computed as $G(\lambda^{\text{new}})$. Since we give as initial data a known function of given λ's, we can compute the true entropy to which the iterates should be converging; this is a useful debugging tool since we can monitor both the residual and

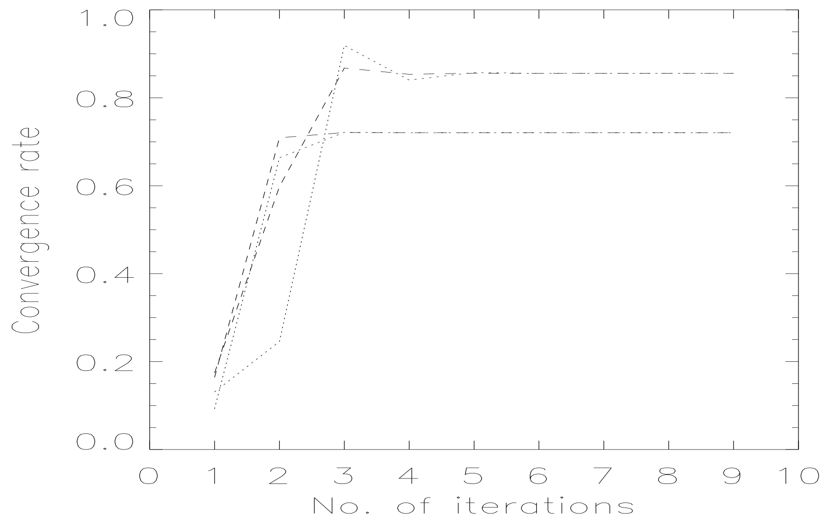

Figure 4: Residual and entropy rates, with and without MG.

the entropy. While convergent, note that the convergence is initially good, but the rate degrades rapidly to steady off at about 0.85.

3.2 Multilevel Methods

To improve the convergence rates observed, we have implemented a two-grid multilevel scheme with unigrid [15] corrections. In this section we describe our method and give some preliminary results.

Coarsening is achieved by pairing adjacent cells in the projections and updating the associated μ's with a single correction that makes their average residual zero. On the coarse grid, we iterate this relaxation process until the norm of the vector of average residuals is below a user supplied ϵ, typically 0.05. All of the calculations are done in a unigrid fashion, on the fine grid. This makes the process more expensive than necessary, but its performance is equivalent to the more efficient V-cycle multigrid scheme and it is much easier to implement and manipulate.

Using the same geometry and data as before, but with relaxation accelerated by coarse grid corrections, we obtain the rates given in the lower two curves in Figure 4. Note that with only a two-grid scheme, we have reduced the convergence rate from about 0.85 to about 0.72.

As a demonstration of the reconstructions we can obtain, we present Figures 5 and 6. Recall that the reconstruction is computed on the optimal grid, but for plotting purposes we essentially use a square grid. While there are several ways of translating from the optimal grid to a square grid, we have chosen simply to evaluate the optimal image

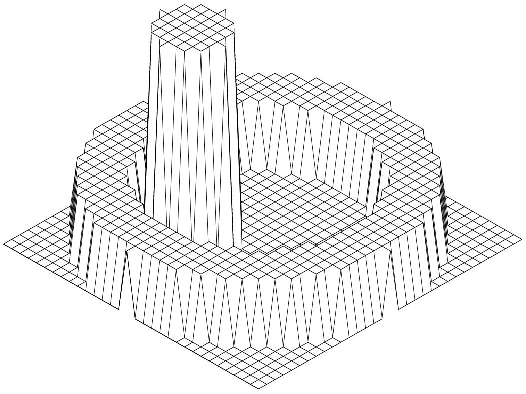

Figure 5: Original image.

function as reconstructed from the optimal λ via (2.30) at the lattice points of the square grid without performing interpolation. Improving this process could be a direction of future investigation.

Figure 5 is the original image. Note that this is *not* a piecewise constant function on the optimal grid, so our reconstructions cannot be exact. In fact, the function of maximum entropy with the data generated by this function is not the original image; as shown previously, it is piecewise constant on the optimal grid, as are our reconstructions.

To obtain Figure 6, we use a simple two-dimensional integrator, based on Simpson's rule, on the original function to make b using 5 projections each with 8 cells. We then iterated our multilevel scheme to convergence (so that the ℓ_2 norm of the average residuals was less than 10^{-4}) and plotted the result in Figure 6 as discussed above. This process produces a set of λ's that we then used as our initial data in the routine to obtain a reconstruction of the first reconstruction. This next reconstruction is virtually identical to the first, as the theory predicts.

As a final note, an extra benefit of this scheme is data compression. Given a data collection geometry, when we collect the N pieces of data, we need only solve the reconstruction problem once to get the N $\mu' s$. From these we can reconstruct the image to any level of resolution desired; indeed, we have shown that the piecewise constant function on the optimal grid is the most information one can extract from the data.

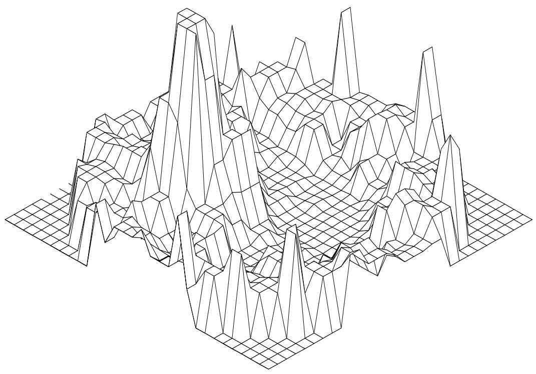

Figure 6: Reconstruction.

4 Conclusions

We have seen that the solution to the maximum entropy image reconstruction problem posed in $L^1(\Omega)$ is a piecewise constant function on the optimal grid. We have also seen that each iterate of the simple iterative scheme for solving the associated dual problem can be computed exactly; no numerical integration or approximations are needed. Finally, we observed that a unigrid scheme to accelerate convergence shows potential, though more testing and analysis is needed.

As a final comment, we note that that the mathematics used to derive the optimal grid and the iterative scheme can be applied to other objective functionals f and other geometries; for example minimum L^2-norm, fan beam projections and non-symmetric placement of the projections. These issues and a more complete discussion of the mathematics in optimization techniques for image reconstruction from projections will be covered in a future paper.

References

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

[2] F. Natterer. *The Mathematics of Computerized Tomography.* John Wiley & Sons, New York, New York, 1986.

[3] C. E. Shannon and W. Weaver. *The mathematical theory of communication.* University of Illinois Press, Urbana, Illinois, 1963.

[4] A. E. Scheerer. *Probability on Discrete Sample Spaces with Applications.* International Textbook Company, Scranton, Pennsylvania, 1969.

[5] J. Skilling and S. F. Gull. The entropy of an image. *SIAM–AMS Proceedings*, 14:167–189, 1984.

[6] J. M. Borwein and A. S. Lewis. Partially-finite convex programming, parts I and II. *Mathematical Programming*, pages 15–48,49–83, 1992.

[7] S. M. Ross. *A First Course in Probability.* Macmillan, New York, New York, 1976.

[8] M. L. Reis and N. C. Roberty. Maximum entropy algorithms for image reconstruction from projections. *Inverse Problems*, pages 623–644, 1992.

[9] J. Diestel. *Sequences and Series in Banach Spaces.* Springer-Verlag, New York, New York, 1984.

[10] A. Decarreau, D. Hilhorst, C. Lemaréchal, and J. Navaza. Dual methods in entropy maximization. Application to some problems in crystallography. *SIAM J. Optimization*, 2(2):173–197, May 1992.

[11] A. P. Robertson and W. Robertson. *Tolological Vector Spaces.* Cambridge University Press, Cambridge, England, 1973.

[12] C. Auyeung and R. M. Mersereau. A dual approach to signal restoration. *IEEE International Conference on Acoustics, Speech and Signal Processing, Glasgow*, pages 1326–1329, May 1989.

[13] J. H. McClellan and S. W. Lang. Duality for multidimensional MEM spectral analysis. *IEE Proceedings Part F*, pages 230–235, April 1983.

[14] R. T. Rockafellar. Integrals which are convex functionals. *Pacific J. of Math.*, 24(3):525–539, 1968.

[15] S. F. McCormick. *Multilevel Prjection Methods for Partial Differential Equations.* SIAM, Philadelphia, Pennsylvania, 1992.