Mean Field Annealing: A Formalism for Constructing GNC-Like Algorithms

Griff L. Bilbro
Wesley E. Snyder
Stephen J. Garnier
James W. Gault

TR-91/3
February 1991

Mean Field Annealing: a formalism for constructing GNC-like algorithms[1]

Griff L. Bilbro[2], Wesley E. Snyder[3], Stephen J. Garnier[2], James W. Gault[4]

We approach optimization problems using Mean Field Annealing (MFA), which is a deterministic approximation, using mean field theory and based on Peierls' inequality, to simulated annealing. The MFA mathematics are applied to three different objective function examples. In each case, MFA produces a minimization algorithm that resembles Graduated Non-convexity (GNC). When applied to the "weak membrane" objective, MFA results in an algorithm qualitatively identical to the published GNC algorithm. One of the examples, MFA applied to a piecewise-constant objective function, is then compared experimentally to the corresponding GNC weak-membrane algorithm. The mathematics of MFA are shown to provide a powerful and general tool for deriving optimization algorithms.

Key Words: Simulated Annealing, Mean Field Approximation, Graduated Non-convexity

---

2 Center for Communications and Signal Processing, North Carolina State University, Raleigh, NC
3 Dept. of Radiology, Bowman Gray School of Medicine, Winston-Salem, NC
4 U.S. Army Research Office, Research Triangle Park, NC

1.0 Introduction

In this paper, we show that the mathematics of Mean Field Annealing (MFA) provides an approach which unifies several difficult optimization problems. We show that the application of MFA results in algorithms that, for the same problems, are qualitatively identical to Graduated Non-convexity (GNC) algorithms, as developed by Blake and Zisserman[6,7]. MFA is based on Simulated Annealing (SA) and derives its power and generality from that popular optimization procedure. MFA differs from SA by analytically approximating the relevant Gibbs distribution rather than stochastically simulating it.

We begin by reviewing SA and presenting its relationship to MFA. Then, we present three examples of the application of the MFA formalism; the third of which is the "weak membrane" objective function proposed by Blake and Zisserman[6].

Finally, we experimentally compare GNC applied to the weak membrane to the corresponding case of MFA applied to the "piecewise-constant" objective function [12]. The two algorithms are shown to produce similar results.

1.1 Background

We are concerned with an analysis and comparison of two recently-published techniques for improving the quality of images. By "improving", we mean the removal of noise from the image while preserving the sharpness of edges. The two techniques of interest are "Graduated Non-convexity", developed in a book[6] by Blake and Zisserman and recently[7] compared with statistical techniques (at least in one dimension), and "Mean Field Annealing", developed[1] and analyzed[4] by the authors.

Both techniques may be considered as image restoration techniques, although GNC was originally motivated by a desire to do edge detection. When considered as restoration methods, both techniques pose minimization problems in which some objective function, H, is to be minimized with respect to the unknown image, f. In both cases, we are given the measured image, g, and some *a priori* knowledge of the nature of the unknown image. Thus we must solve $min_f H (f | g)$

where $H = H_{gnc} = H_n + H_{pgnc}$ for GNC and $H = H_{mfa} = H_n + H_{pmfa}$ for MFA.

In both cases the "noise term" (to use the terminology of the Gemans[9]) is the same:

$H_n = \| f - g \|^2$ where $\| x \|^2 = \sum_i x_i^2$, and the prior terms (H_{pgnc} and H_{pmfa}) encapsulate our prior knowledge of the image statistics. Specific examples of prior terms are presented later in this paper.

Since both techniques utilize the same noise term (the constant σ which occurs in the MFA papers is irrelevant here), we will concentrate the analysis of this paper on the prior term(s).

In this paper, we will consider a particular instance of image restoration/edge detection, the case in which our *a priori* knowledge states that the image is uniform in brightness, except for step discontinuities. Blake and Zisserman [6] refer to this case as the "weak membrane", and the equivalent MFA instance is referred to [11] as "piecewise-uniform". Extension of the analysis of this paper to the "weak plate" (GNC) or "piecewise planar" (MFA) images is straightforward, but will not be developed here.

In the remainder of this section, we briefly describe both MFA and GNC, using, as much as possible, the same notation. In sections 2.0-4.0 we show how MFA leads to annealing algorithms for various types of objective functions. The reader is referred to the original papers for more extensive formal derivations. In Section 5.0, we show that GNC is derivable from MFA. In Section 6.0, we run both techniques on test images, corrupted by various amounts of noise, and compare the performance.

1.2 Graduated Non-convexity

In the "weak membrane" application of GNC, the minimization problem is

$$min_{f, l}H_{gnc} \text{ where} \tag{1}$$

$$H_{gnc} = H_n + S + P, \quad S = \lambda^2 \sum_i (\nabla (f_i))^2 (1 - l_i), \text{ and } P = \alpha \sum_i l_i \tag{2}$$

Here, the $l_i \in \{0, 1\}$ denotes a discontinuity in f at the ith pixel. That is, if $l_i = 1$, the pixel at point i is interpreted as an edge point. It has been shown [6] that minimizing H_{gnc} can be reduced to the following problem, which involves only continuous variables.

$$min_f(H_n + \sum_i v (\nabla (f_i))) \tag{3}$$

In (2) and (3), the ∇ represents any operator which returns a measure of the local "edginess" of the image. The 2-norm of the gradient is one such measure.

For the v function of (3), Blake and Zisserman obtain the "clipped parabola" of Figure 1.
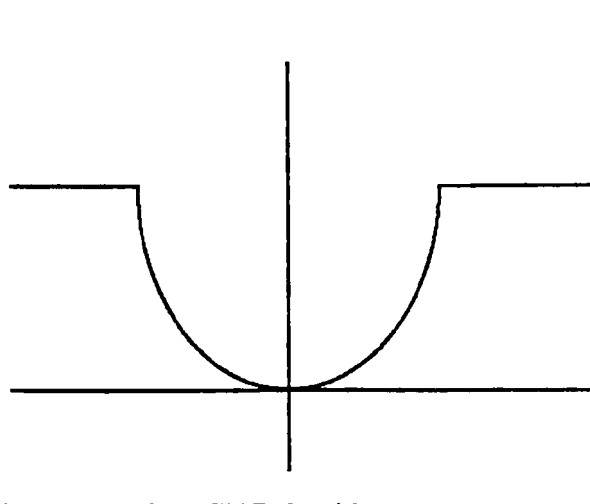

Figure 1 Prior energy of the GNC algorithm.

The minimization problem posed by Equation (3) is unsolvable by techniques such as gradient descent, since the function is in general not convex. That is, it may possess many minima. Instead, GNC approximates v with the piecewise-smooth function

$$v^*(t) = \begin{cases} \lambda^2 t^2 & \text{if } (|t| < q) \\ \alpha - c^* \, ((|t| - r)^2/2) & \text{if } q \le |t| < r \\ \alpha & \text{if } (|t| \ge r) \end{cases} \tag{4}$$

where $c = c^*/p,$

$$r^2 = \alpha \left(\frac{2}{c^*} + \frac{1}{\lambda^2}\right) \text{ and } q = \frac{\alpha}{\lambda^2 r} \tag{5}$$

and c^* is a scalar constant.

Equations(4) and (5) then define the algorithm. Reducing the parameter p from 1 to 0 steadily changes v^* until it becomes precisely equal to v. This produces a family of prior energies, illustrated in Figure 2.
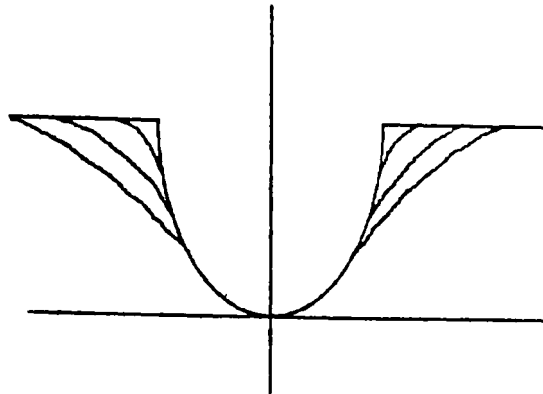
Figure 2 Smoothed approximations to the energy of Figure 1. Smaller values of p result in approximations which are closer to the desired prior.

The process of gradually reducing p begins by minimizing a function which is convex, and therefore has a unique minimum. Then, from that minimum, the local minimum is tracked continuously as p is reduced from 1 to 0.

1.3 Mean Field Annealing

The Mean Field Annealing approach described in [12] results in an effective objective function similar to that of GNC:

$$H_{mfa} = H_n - \frac{b}{\sqrt{2\pi T}} \sum_i exp\left(-\frac{(\nabla(f_i))^2}{T}\right)$$

(6)

Here, b is a scalar which defines the relative importance of the prior and noise terms, whose optimal value is dependent on the noise in the image[2], and T is a parameter that is reduced in the course of the algorithm, much like the p of GNC.

Plotting the prior term from MFA in Figure 3, we note a striking similarity between Figure 2 and Figure 3. In both cases, we have an energy function which increasingly penalizes the presence of gradients in the image. In the GNC case, the prior retains its original shape, and the "annealing" process (that is, the lowering of p) results in succes-
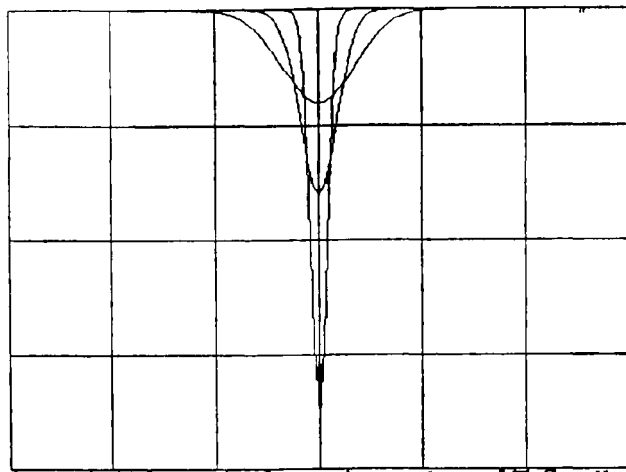
Figure 3 Prior energy for MFA, for various values of T. Smaller T results in sharper peaks.

sively closer fits to the predetermined shape of the prior. In the MFA case, the shape of the prior itself changes, retaining a constant area, but becoming narrower as T is reduced.

In the next sections, we discuss the mathematics of MFA and show both its relationship to both simulated annealing and to GNC. MFA can be applied directly to the H of equation (1) and results in another family of curves like Figure 1 (see Figure 6).

2.0 Simulated Annealing and MFA

Although Simulated Annealing is slow, it is simple, general, easy to apply to new problems, and remarkably successful even when theoretical conditions for convergence are not met[13]. SA works by gradually cooling an ongoing stochastic simulation of a Gibbs distribution. Mean field theory provides a deterministic approximation to a Gibbs distribution which also can be cooled in the same way to produce a Mean Field Annealing (MFA) algorithm. Many SA algorithms can be converted to analogous MFA algorithms that run in 1/50 the time required by the SA version [1,4,8,11,12,14]. However, because it is an approximation, MFA does not inherit any guarantee of convergence even when the analogous SA does converge.

Historically, treatments of the mean field have been restricted to "Ising-like" systems described by an objective function (also sometimes denoted as an "energy function" or "Hamiltonian") involving binary vector $s = \{s_i\}_{i=1}^{i=N}$ with energy

$$H(s) = -\sum_1 h_i s_i - \sum_{ij} v_{ij} s_i s_j. \tag{7}$$

But recently[3,4], we have extended MFA to a wider class of problems. This extension depends on Peierls' inequality $F \leq W$ which bounds the exact "free energy" at temperature T of a system of s described by Equation(7)

$$F = -T \ln \sum_{\{s\}} exp\left(-\frac{H}{T}\right)$$

by the "Weiss free energy"

$$W = F_0 + \langle H - H_o \rangle.$$

Here H_o is an arbitrary function of s. The expectation is taken with respect to the density $Z_o^{-1} exp\left(\frac{-H_o}{T}\right)$ over all possible configurations of the s's normalized by the factor

$$Z_o = \sum_s exp\left(-\frac{H_o}{T}\right), \text{ which is also used to define } F_0 = -T \ln Z_0.$$

In the context of statistical mechanics the function F characterizes the equilibrium of the system described by Equation(7) at temperature T. The utility of the bound $F \leq W$ is that it remains valid even if H_0 depends on adjustable parameters in addition to s. It can be shown[4] that equality obtains if and only if $H_0 = H$. Therefore, we may choose an H_0 which is dependent on some set of adjustable parameters $\{x_i\}$, and choose those parameters to minimize W.

In this sense, adjusting H_0 to minimize W adjusts H_0 to most closely resemble H. Peierls' inequality was originally derived in the context of statistical mechanics, and restricted to problems with discrete-valued variables, but we have recently shown how to choose H_0 to treat continuous variables as well [3,12] and have shown, by constructing an information-theoretic derivation[4], that no analogy to physics is necessary to justify the inequality.

MFA converts an optimization problem into the limiting member of a family of optimization problems. Instead of directly varying H, MFA varies a certain weighted average

of H. The width of averaging kernel depends on a scalar called the temperature T. For small T the kernel approaches a Dirac delta function, and in the limit the original problem is recovered. For large T, the fine structure of the original problem is averaged away, and in this limit the approximating objective becomes convex even if the original objective was not. MFA works when the large-T optimum approaches the best (or at least a good) low-T optimum as T is reduced. Geiger and Girosi first reported a relation between MFA and GNC optimization of the weak membrane. They showed that MFA can be applied to the weak membrane to obtain an algorithm that is qualitatively identical to GNC for the same problem[10], although they did not make use of Peierl's inequality. In this paper we show that using the MFA derivation approach produces GNC-like algorithms for many optimization problems.

3.0   A binary problem

The H of Equation (7) is useful for graph bisection if v_{ij} is the adjacency matrix, h_i reflects some externally specified preference for assigning node i to a particular partition, and the value of s_i determines the assignment of node i [1,8]. A slightly more complicated form of H has been used to restore binary images[4]. Here we consider the simplest one-dimensional case where, $\forall i, h_i = h$, and all v vanish except $v_{ij} = v$, $(\forall j) j = i+1$ so that

$$H = -h \sum_{i=1} s_i - \sum_{i=1}^{i=N} s_i s_{i+1}$$ with the s connected in a ring $s_{N+1}=s_1$. For concreteness

we take $s_i \in \{-1, 1\}$. [5]

If $0<h<v$, then a descent algorithm cannot reliably find the minimum of H unless its moveset includes flips of at least 2 v/h adjacent variables. To see this, consider global minimum configuration $s_i = 1$ for all i, with energy $H(0,N) = (-h-v) N$. Now the configuration $s_i = -1$ for all i, with higher energy $H(N,0) = (h-v)N$ is a local minimum since flipping k consecutive variables to 1 raises the energy to $H(N-k,k) = H(N,0) -2kh + 4v > H(N,0)$ for $k< 2v/h$. Flipping non-consecutive variables is higher still. Thus, there is a "barrier" between the "all down state" and the global minimum "all up state". Evidently, a success-

---

5 Any other kind of random binary variable in a Hamiltonian can be algebraically transformed to the range $\{1,+1\}$, although special consideration is required if a factor of s^2 occurs anywhere.

ful descent algorithm requires ingenuity to construct its moveset. MFA provides an alternative that requires no such ingenuity.

A mean field approximation of the associated Gibbs density is obtained by choosing $H_o = -\sum_i x_i s_i$ with adjustable mean field parameters $x = \{x_i\}_{i=1}^{i=N}$. The sum over configurations becomes $\sum_s \phi(s) \equiv \sum_{s_1 = \pm 1} \sum_{s_2 = \pm 1} \cdots \sum_{s_N = \pm 1} \phi(s)$ for any function $\phi(s)$

so that $Z_0 = \prod_i 2 \cosh x_i / T$,

from which $F_0 = -T \sum_i \ln(2 \cosh x_i / T)$. The average value, $<s_i>$ under H_0 is related to x_i by the function m

$$m(x_i) \equiv \tanh(x_i / T) \tag{8}$$

which tends toward the algebraic sign of x_i for small enough T so that the original binary character of the problem is recovered from MFA at $T = 0$. Now $\langle H_0 \rangle = -\sum_i x_i m(x_i)$

and

$$\langle H \rangle = -h \sum m(x_i) - v \sum m(x_i) m(x_{i+1}) \tag{9}$$

Combining these, we obtain a bound on the free energy $F \leq W$ where

$$W = \sum_i -T \ln 2 \cosh \frac{x_i}{T} - h m(x_i) - v m(x_i) m(x_{i+1}) + x_i m(x_i) \tag{10}$$

which can be minimized with respect to the mean field x. The resulting self-consistency condition on x

$$x_i = h + v m(x_{i-1}) + v m(x_{i+1}) \tag{11}$$

can be converted to an equivalent condition for the expectations $m_i = m(x_i)$,

$$m_i = \tanh\left(\frac{h + v m_{i-1} + v m(x_{i+1})}{T}\right). \tag{12}$$

The minimum of W tells us about the equilibrium at a given T, but to find the minimum of H, we must in general anneal. For our shift invariant choice of h and v, the low W states have, for all i, shift invariant $m_i = m$ as well. Equation (12) becomes

$$m = \tanh \left((h + 2vm) / T \right) \qquad (13)$$

which is graphically solved in Figure 4 for $h=v=1$ at several values of T. Figure 4 shows that Equation (13) has one solution for high T but three solutions below a certain "critical" T. It can be shown that the middle solution is a maximum of W and the two extreme solutions are minima. Here the rightmost solution determines the global minimum as T falls. One way to describe this is that MFA varies the "non-convexity" of the optimization problem as a function of T.
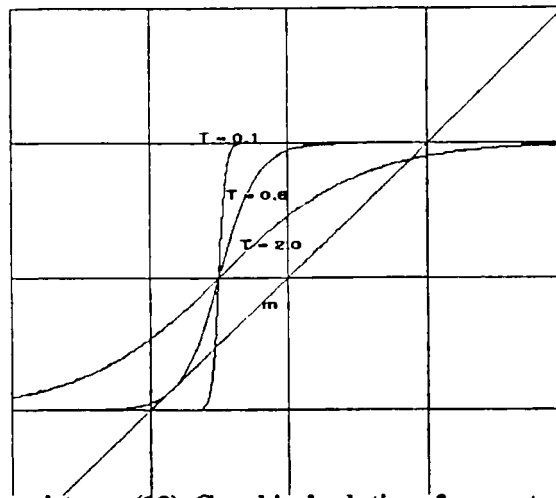

Figure 4 Self consistency (13). Graphical solutions for m = tanh((h + 2vx) / T). The straight line is y=x. The curves are y = tanh((h + 2vx) /T). Intersections between the straight line and curves indicate extrema of W. At high T, there is only one solution. At low T, the problem becomes non-convex. MFA tracks the first intersection to the global minimum.

4.0 A Real-Valued Problem

The objective or Hamiltonian $H(u) = \dfrac{1}{2} \sum_i (u_i - d_i)^2 - \lambda \sum_i V(u_i - u_{i+1})$ (14)

is minimized by a piecewise-constant approximation to d, i.e., when real variables u_i can be chosen to resemble data d_i and to frequently coincide with neighbors $u_{i\pm 1}$. If $V(x)$ is chosen to be a Gaussian of sufficiently small width centered at $x=0$. The following mean

field analysis can be performed for any such Gaussian, and in the limiting case of zero width, when V becomes a Dirac delta function, the result of this procedure coincides with the direct analysis of $V(x) = \delta(x)$; therefore, we will restrict ourselves to that simpler case here. For this problem, the simplest useful choice of H_0 is

$$H_0 = \frac{1}{2} \sum_i (u_i - x_i)^2 \tag{15}$$

with real means x_i. The average in Peierls' inequality is weighted by the Gaussian density $Z_0^{-1} exp(-H_0/T)$ and now extends over all real configurations. Z_0 involves $\int_\infty^\infty du_1 \int_\infty^\infty du_2 \ldots \int_\infty^\infty du_N$, but it factors so that the normalization is $Z_0 = \prod_i \sqrt{2\pi T}$. The logarithm of a product is a sum, so that $F_0 = -T \sum_i ln \sqrt{2\pi T}$. The average H_0 is a sum of second moments $\langle H_0 \rangle = \sum_i T/2 = (NT)/2$. For this H_0, both F_0 and $<H_0>$ are independent of x. All the x dependence appears in $<H>$ which is the sum of a data term $\frac{1}{2} \sum_i (u_i - d_i)^2 \geq (\frac{NT}{2})$ and an interaction that involves the average of a Dirac delta $\langle \delta(u_i - d_i)^2 \rangle$

$$\langle \delta(u_i - u_{i+1}) \rangle$$
$$= \frac{1}{2\pi T} \int \left(exp \left(-\frac{(u_i - x_i)^2}{2T} - \frac{(u_{i+1} - x_{i+1})^2}{2T} \right) \delta(u_i - u_{i+1}) \, du_i du_{i+1} \right) \tag{16}$$

which can be integrated once with the Dirac delta

$$\langle \delta(u_i - u_{i+1}) \rangle = \frac{1}{2\pi T} \int \left(exp \left(-\frac{(u_i - x_i)^2}{2T} \right) exp \left(\frac{-(u_{i+1} - x_{i+1})}{2T} \right) du_i \right) \tag{17}$$

and then evaluated with a table of integrals to obtain

$$\langle \delta(u_i - u_{i+1}) \rangle = \frac{1}{\sqrt{4\pi T}} exp(-(x_i - x_{i+1})^2 / (4T)). \tag{18}$$

Combining these results,

$$W = \text{constant} + \sum_i \left(\frac{(x_i - d_i)^2}{2} - \frac{\lambda}{\sqrt{4\pi T}} exp\left(-\frac{(x_i - x_{i+1})^2}{4T} \right) \right) \qquad (19)$$

where the constant is independent of x and can be dropped. This W has the same form as the original H except that the Dirac delta has been replaced by a Gaussian which is plotted in Figure 5 for several T. Note that the Gaussian tends to a Dirac delta as T vanishes. Figure 5 should be compared with Figure 6 where the non-convexity of the clipped parabola is also reduced by increasing T.

In both cases, MFA prescribes a solution (minimum) first to be found at high T, where the problem will be convex. That solution is the initial state for another minimization at a slightly lower T. Thus, MFA replaces a non-convex problem with a family of problems, of increasing non- convexity. Two-dimensional versions of this formulation have been used to remove noise from corrupted observations of real-valued images known to be piece-wise-constant[12]. MFA has been used to identify the most effective form for the delta function interaction in two dimensions[11]. Piecewise-linear restorations are similarly obtained by replacing the first difference in the argument of the delta function by the appropriate second difference[2].

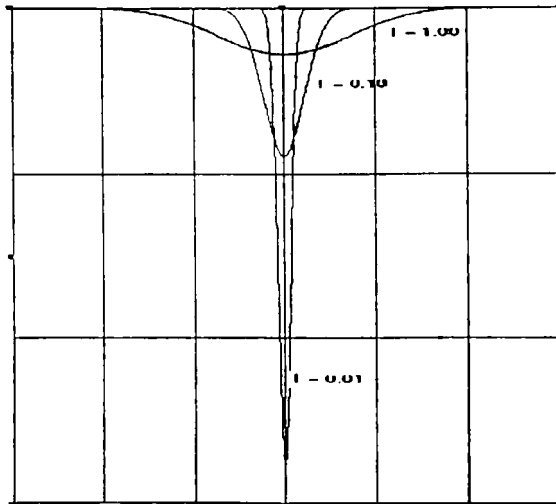## 5.0   A Mixed Problem - The Weak Membrane

Figure 5   The interaction potential for Equation (19) plotted for several T.

In this section, we consider a problem - Blake and Zisserman's weak membrane[6] - which involves both continuous variables (image brightness values), and binary variables, "line processes", in which the presence of a 1 indicates an edge. We will present the deri-

vation of this section in somewhat more detail than those presented in the previous section, so that the exact approach will be more easily followed. In one dimension, the weak membrane is[6]

$$H(u, l) = \sum_{i=1}^{i=N} [(u_i - d_i)^2 + \lambda^2 (u_i - u_{i+1})^2 (1 - l_i) + \alpha l_i]$$ (20)

for line field $l_i \in \{0, 1\}$, real intensity u_i, and for simplicity, $u_{N+1} \equiv u_1$. It is possible to treat both u and l as random fields and apply mean field theory to both, but Blake and Zisserman are principally concerned with non-convexities due to discrete l, so we restrict H_o to depend only on l, and its estimating parameter, x,

$$H_o(x) = -\sum_i x_i l_i .$$ (21)

To set up the Mean Field problem, we must determine and evaluate each term in Peierls' inequality. As usual, we will take averages using the Gibbs density: $\dfrac{exp(-H_0/T)}{Z_o}$ and now the normalization is over all real configurations:

$$Z_0 = \sum_{l_1 = 0, 1} \sum_{l_2 = 0, 1} \cdots \sum_{l_N = 0, 1} exp(-\frac{H_0}{T})$$ (22)

$$= \sum_{l_1 = 0, 1} \sum_{l_2 = 0, 1} \cdots \sum_{l_N = 0, 1} \prod_{j=1}^{N} exp(-\frac{x_j l_j}{T})$$ (23)

$$\prod_{i=1}^{N} \sum_{l_i = 0, 1} exp(\frac{x_i l_i}{T}) = \prod_{i=1}^{N} (1 + exp(x_i/T))$$ (24)

From this we have $F_o = -T\sum_i ln(1 + exp(x_i/T))$.

It will prove convenient later to evaluate the expectation of l_i :

$$\langle l_i \rangle = \frac{1}{Z_o} \sum_{l_1 = 0, 1} \sum_{l_2 = 0, 1} \cdots \sum_{l_{i-1} = 0, 1} \sum_{l_{i+1} = 0, 1} \cdots \sum_{l_N = 0, 1} [exp(-H_o/T)]$$

---

6 Although Blake and Zisserman refer to the one-dimensional version of this model as the *weak string*, for simplicity of nomenclature, we will use the term *membrane* for both, since our derivation is readily extendible to two dimensional functions.

We abbreviate the series of summations for notational convenience and expand the argument of the exponential

$$\langle l_i \rangle = \frac{1}{Z_0} \sum_{\{l|\, l_i=1\}} exp\,(x_i/T) \prod_{j \neq i} exp\,(\frac{x_j l_j}{T}) \tag{25}$$

Simplifying:

$$\langle l_i \rangle = \frac{exp\,(x_i/T) \prod_{j \neq i} (1 + exp\,(x_j/T))}{\prod_k (1 + exp\,(x_k/T))} = \frac{e^{x_i/T}}{1 + e^{x_i/T}}$$

which we define to be $\langle l_i \rangle \equiv \sigma\,(x_i)$.

Then, $\langle H_0 \rangle = -\sum_i x_i \sigma\,(x_i)$, and

$$\langle H \rangle = \sum_{i=1}^{N} (u_i - d_i)^2 + \lambda^2 (u_i - u_{i+1})^2 (1 - \sigma\,(x_i)) + \alpha\sigma\,(x_i) . \tag{26}$$

Substituting these terms into Equation (9), we have

$$W(u,x) = \sum_i -Tln\,(1 + exp\,(\frac{x_i}{T})) + (u_i - d_i)^2 + \lambda^2 (u_i - u_{i+1})^2 (1 - \sigma\,(x_i)) + \alpha\sigma\,(x_i) + x_i \sigma\,(x_i) . \tag{27}$$

We may minimize W at a given T by setting

$$\frac{\partial W(u,x)}{\partial x_i} = -\frac{T exp\,(\frac{x_i}{T}) \frac{1}{T}}{1 + exp\frac{x_i}{T}} + \lambda^2 (u_i - u_{i+1})^2 (\frac{-\partial\sigma}{\partial x_i}) + \frac{\alpha\partial\sigma}{\partial x_i} + \sigma\,(x_i) + x_i(\frac{\partial\sigma}{\partial x_i}) = 0 . \tag{28}$$

From (28), the first term cancels $\sigma(x_i)$, yielding

$$(\alpha + \hat{x}_i - \lambda^2 (u_i - u_{i+1})^2)\,(\frac{\partial\sigma}{\partial x_i}) = 0, \tag{29}$$

where we are dropping the explicit dependence of σ on x.

Since $\dfrac{\partial\sigma_i}{\partial x_i}$ never vanishes for nonzero T, we have

$$\hat{x}_i = \lambda^2 (u_i - u_{i+1})^2 - \alpha \tag{30}$$

for this set of x. Collecting the terms of W which multiply σ, we find

$$W(u, x) = \sum_i \left[-Tln\,(1 + exp\,(\frac{x_i}{T})) + (u_i - d_i)^2 + \lambda^2 (u_i - u_{i+1})^2 + (-\lambda^2 (u_i - u_{i+1})^2 + \alpha + x_i)\,\sigma_i \right] \tag{31}$$

but from (29), the coefficient of σ_i is 0 at $x_i = \hat{x}_i$, so that

$$\hat{W}(u) = \sum_i \left[(u_i - d_i)^2 - Tln\,(1 + exp\,(\frac{\hat{x}_i}{T})) + \lambda^2 (u_i - u_{i+1})^2 \right]$$

$$= \sum_i \left[(u_i - d_i)^2 - Tln\left(1 + exp\left(\frac{\lambda^2 (u_i - u_{i+1})^2 - \alpha}{T}\right)\right) + \lambda^2 (u_i - u_{i+1})^2 \right] \tag{32}$$

The reader may have some concern over the use of a derivative to find $min_x W$, since we have, at least superficially, no assurance that W is convex in x. However, from Equation (30), we see that W is minimized by a unique x, for finite T.

The MFA methodology prescribes that in order to find $min_u H(u, l)$, we anneal

$min_{u,x} W(u, x) = min_u \hat{W}(u)$ on T. The last two terms of Equation (32) are independent of the data and reflect our prior knowledge of the nature of the image. These terms may therefore be considered as prior potentials [9] or effective potentials [10]

$$W_{PP} = \left(-Tln\left(1 + exp\frac{(\lambda^2 (u_i - u_{i+1})^2 - \alpha)}{T}\right) + \lambda^2 (u_i - u_{i+1})^2 \right). \tag{33}$$

W_{PP} is plotted vs. $(u_i - u_{i+1})$ for several values of T in Figure 6. Note that for $T < 0.3$, this plot is qualitatively identical to the GNC smoothing of the weak membrane. Thus, at low T, the MFA-derived objective approximately reproduces Blake and Zisserman's piecewise quadratic heuristic function. This similarity was first noted by Geiger and Girosi [10], although they arrived at the result in a less general way.

6.0 Experimental findings

The conceptual similarity of the piecewise constant MFA and the weak membrane GNC prompted us to compare the two algorithms experimentally. In this section when we refer to MFA, we mean the piecewise constant formulation given by Equation 6, and simi-
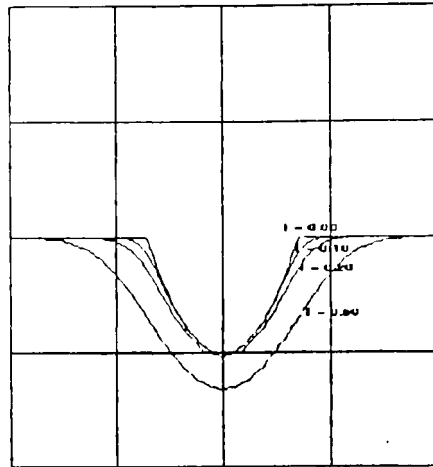
Figure 6 A plot of the last two terms of (35) against the difference $u_i - u_{i+1}$ for several values of T.

larly, when we refer to GNC, we mean the weak membrane formulation given by Equation 2.

The two approaches each were used to restore the same image with various signal-to-noise ratios. On each application of MFA and GNC to a noisy image, the respective parameters were varied to achieve the best possible image restoration. Several hundred runs with distinct parameter values were completed for each equation. We found that for each noisy image there existed some parameter set for each algorithm such that the restored images were of comparable quality.

The resulting image quality achieved is depicted in the following figure with the original image (Figure 7), the image corrupted with SNR= 2 (Figure 8), the GNC restored image (Figure 9), and the MFA restored image (Figure 10).
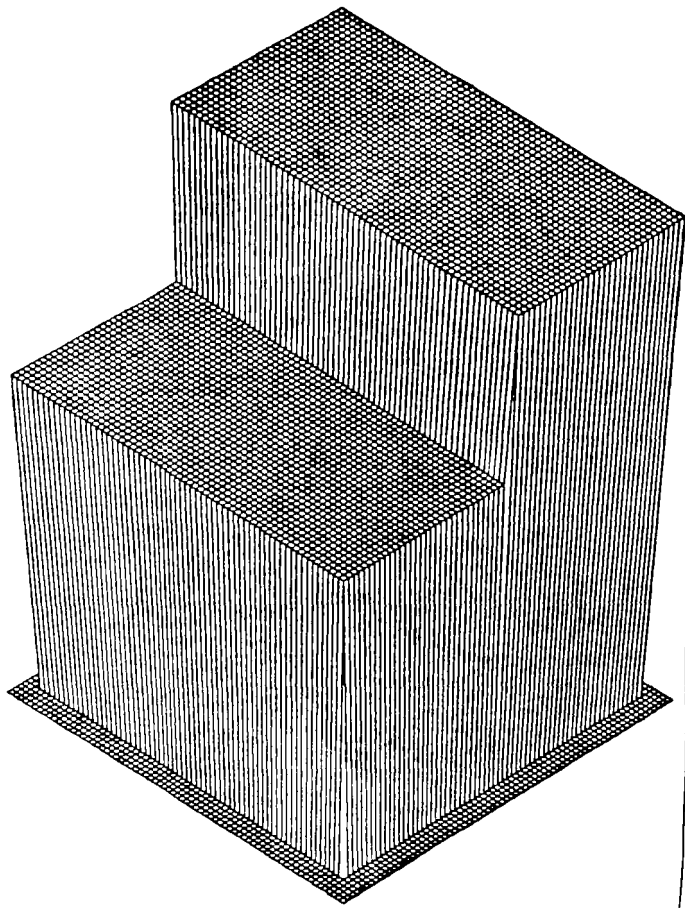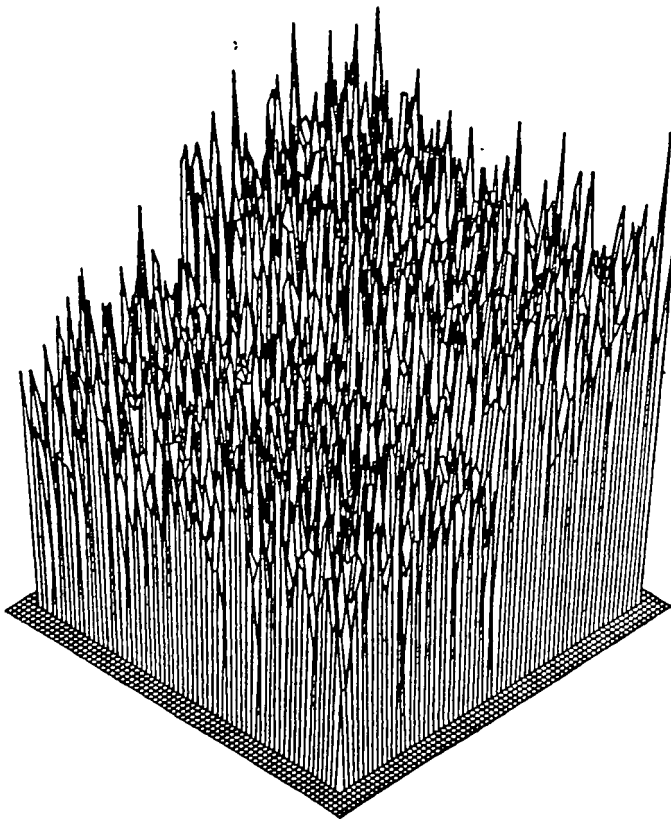
Figure 7 Original step image

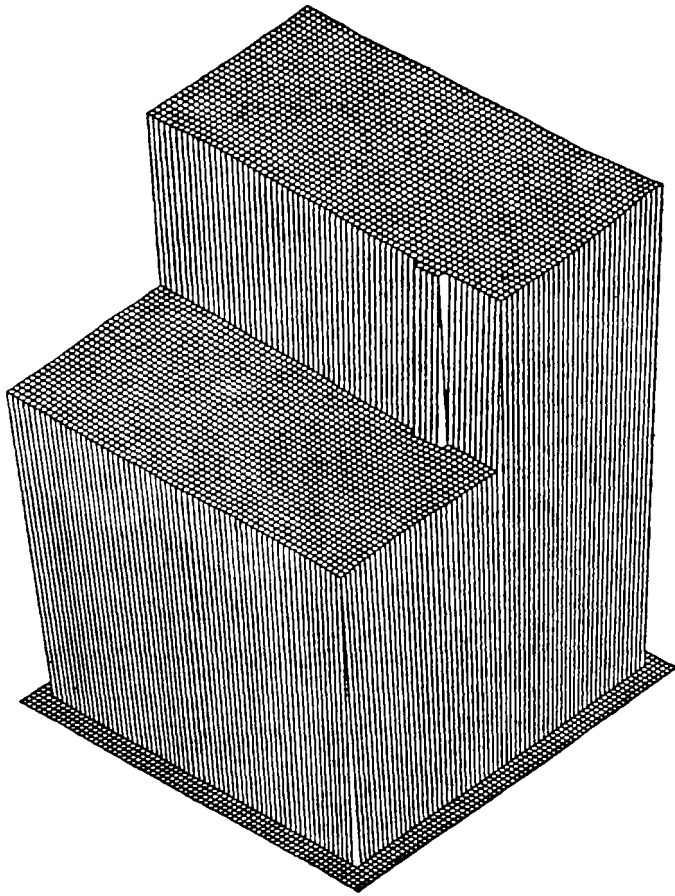Figure 8 Image after addition of random noise. SNR=2.

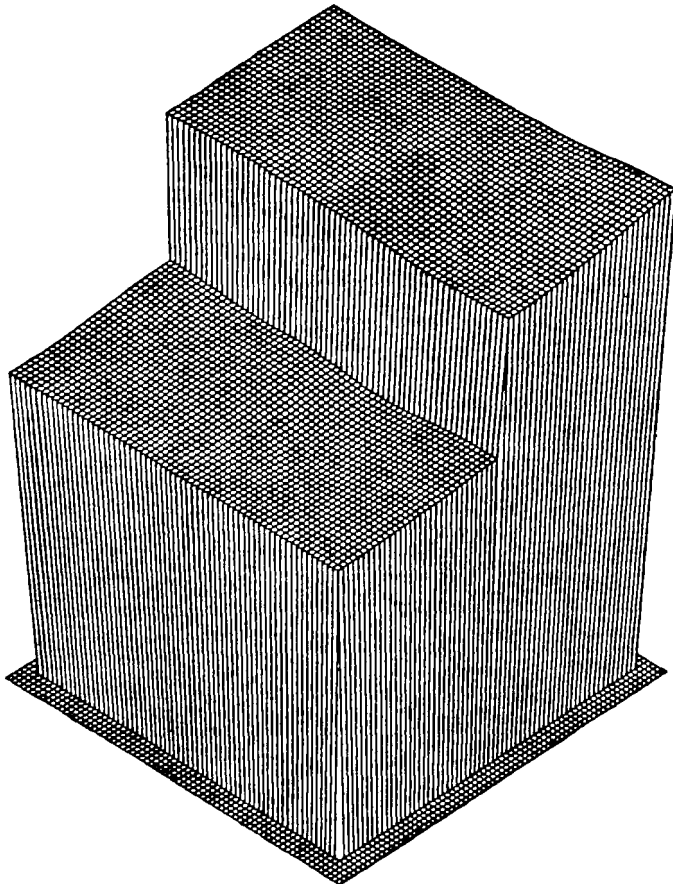Figure 9 Restoration using GNC.

Figure 10 Restoration using MFA.

Coding of the GNC algorithm found in [6], which performs descent using successive over-relaxation, and using an implementation of MFA, which also uses SOR, we found the execution times of MFA to be roughly ten times faster than GNC for high noise cases (SNR < 3). For cleaner images, SNR >=4, the GNC execution times were faster.

We found it more difficult to find the GNC parameter set which produced the best restored image than it was to find the corresponding MFA parameter set. However, this may simply reflect our greater familiarity with the behavior of the MFA algorithm.

7.0 Conclusion

In this paper we have illustrated the application of the mathematics of mean field theory to produce a deterministic approximation to simulated annealing. This approach, which we call Mean Field Annealing (MFA), when applied to a suitable objective function, results in an annealing algorithm. Typically MFA algorithms have an execution time speed up of 50:1 over the corresponding simulated annealing algorithm.

In addition, we have shown that the application of MFA results in algorithms that, for the same problems, are qualitatively identical to GNC algorithms. This MFA, GNC relationship provides new insight that helps unify our understanding of optimization problems. We see that simulated annealing, for problems with discrete variables which it treats stochastically, is related to MFA, which treats both continuous and discrete variables deterministically. In turn, we see that since MFA is related to GNC, which owes its derivation to physical analogies such as line processes and weak membranes, and therefore GNC also has a relationship to simulated annealing.

Further, we demonstrated the use of MFA to derive optimization algorithms with examples of both combinatorial and continuous optimization, as well as the weak membrane example, with mixed, binary and continuous variables. In the latter case we compared experimentally the corresponding MFA and GNC algorithms applied to the restoration of a simple, noisy image. The principal conclusion of this paper is that the mathematics of MFA, using annealing on an expectation resulting from Peierls' inequality, provide a powerful and general tool for deriving optimization algorithms.

8.0 References

1 G. Bilbro, R. Mann, T. Miller, W. Snyder, D. Van den Bout, and M. White, "Optimization by Mean Field Annealing", in *Advances in Neural Information Processing Systems* Morgan-Kauffman, San Mateo, 1989

2 G. Bilbro and W. Snyder, "Range Image Restoration using Mean Field Annealing" in *Advances in Neural Network Information Processing Systems*, Morgan-Kauffman, San Mateo 1989.

3 G. Bilbro and W. Snyder "Mean Field Annealing, an Application to Image Noise Removal" *Journal of Neural Network Computing*, forthcoming.

4 G. Bilbro and W. Snyder "Mean Field from Relative Entropy" *Journal of the Optical Society of America, A.* forthcoming.

5 G. Bilbro and W. Snyder, "Fusion of Range and Reflectance Image Data Using Markov Random Fields" IEEE International Symposium on Intelligent Control, Arlington, VA, August, 1988.

6 A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, 1987.

7 A. Blake, "Comparison of the Efficiency of Deterministic and Stochastic Algorithms for Visual Reconstruction", *IEEE PAMI*, vol. 1, no 1. January, 1989.

8 D.E. Van den Bout, and T.K. Miller, "Graph Partitioning using Annealed Neural Networks" *IEEE Trans. Neural Networks* 1(2):192-203, 1990

9 S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and Bayesian Restoration of Images", *IEEE PAMI*, vol. 6, no 6, 1984.

10 D. Geiger and F. Girosi, "Parallel and Deterministic Algorithms for MRFS: Surface Reconstruction and Integration", AI Memo NO 1114, MIT, 1989.

11 H. Hiriyannaiah, "Signal Reconstruction using Mean Field Annealing", Ph.D. thesis, North Carolina State University, 1990

12 H. Hiriyannaiah, G.L. Bilbro, W.E. Snyder, and R.C. Mann, "Restoration of Piecewise Constant Images via Mean Field Annealing", *Journal of the Optical Society of America A*, pp. 1901-1912, December 1989.

13 P.J.M. van Laarhoven and E.H.L. Aarts, *Simulated Annealing: Theory and Applications*, D. Reidel, Norwell, Mass 1988.

14 C.M. Soukoulis, K. Levin, and G.S. Grest, "Irreversibility and Metastability in Spinglasses. I. Ising Model" *Phys. Rev. B.* 28(3):1495-1509.