

The Diffusion Equation Method of
Global Optimization is a Mean Field
Approximation to Simulated
Annealing

Griff L. Bilbro

Center for Communications and Signal Processing
Department Electrical and Computer Engineering
North Carolina State University

TR-91/15
September 1991

The Diffusion Equation Method of Global Optimization is a Mean Field Approximation to Simulated Annealing

Griff L. Bilbro

Abstract

We show that the *diffusion equation method* of global minimization of Kostrowicki and Pielak is equivalent to mean field annealing, a previously reported deterministic approximation to the usual simulated annealing algorithm. We derive the mean field approximation. We show how to apply it to real-valued problems and present the results for high dimensional image optimization problems involving thousands of variables.

1 Introduction

Recently in this journal, Kostrowicki and Piela[7] reported successful applications of their diffusion equation method to global minimization of low dimensional standard test problems in the Goldstein-Price, Hartman, Shekel, and Griewank families. In earlier work we[1, 5, 2] introduced mean field annealing and showed it to be a powerful and simple approach to a variety of difficult optimization problems. Mean field annealing is a simple deterministic approximation to simulated annealing.

In this paper we show that the algorithm of Kostrowicki and Piela is equivalent to mean field annealing. Their “Fourier-Poisson formula” is the mean field condition for approximating real-valued systems. Their “reversing procedure” in time is equivalent to the gradual reduction of temperature of simulated annealing. Kostrowicki and Piela have shown that the approach is useful for low-dimensional problems. We will show that mean field annealing is equally useful for high dimensional image optimization problems involving thousands of real variables. We also show that mean field annealing applies equally well to combinatorial optimization (*e.g.* the traveling salesman problem), which it converts to more tractable nonlinear optimization problems.

In Section 2 we will review the features of simulated annealing that are relevant to the development of mean field annealing. In Section 3 we will derive the mean field approximation

from information-theoretic considerations even though it is usually obtained with the tools of statistical mechanics[1, 5, 2]. We will show in Section 4 how the mean field approximation can be applied to both discrete-valued problems and real-valued problems. The formulation for real-valued problems will be shown to coincide with Kostrowicki and Piela's diffusion equation method. In Section 5 we will present results for a class of image processing problems involving many more variables than Kostrowicki and Piela considered.

2 Simulated Annealing

In simulated annealing, minimization is reformulated as a stochastic sampling problem. Instead of minimizing $U(s)$ with respect to $s \in S \subset R^N$, the associated Gibbs distribution

$$P[s] = \frac{1}{Z} \exp(-U[s]/T) \tag{1}$$

is maximized. Here the normalization in the domain³ S

$$Z = \int ds^N \exp(-U[s]/T) \tag{2}$$

depends on T only. The variable T is artificially introduced to create a family of distributions that emphasize maxima at low T . Since the exponential function is monotonically increasing, finding the mode of P at any T is logically identical to finding the minimum of U .

³When the domain S is discrete, the integral is replaced by a sum, but in this article we are concerned with the optimization of real variables.

The two formulations differ practically, however, because the sampling formulation admits the powerful Metropolis sampling procedure[9] which can efficiently produce instances of s that are statistically distributed according to Equation 1. At low enough T , every sample s is practically the global minimizer. This approach is feasible only if the sampling is begun at a high T and continued as the temperature is gradually reduced, a procedure called Simulated Annealing by Kirkpatrick, *et al.*, who introduced it[6].

3 Mean Field Approximation

In earlier work we[1, 5, 2] showed that simulated annealing could be accelerated with the mean field approximation. In this approach the important structure of P is approximated with a more convenient distribution P_0 for a sequence of falling values of T . In this section we provide an information-theoretic procedure for studying a given difficult P using an essentially arbitrary easy P_0 by minimizing the entropy of P_0 relative to P , or equivalently, the cross-entropy or Kullback-Leibler[8] distance between P_0 and P . This information-theoretic procedure leads to our previously successful approach based on the theoretical tools of statistical physics.

Assume we have another positive but otherwise arbitrary distribution $P_0[s, m]$. It is useful to choose P_0 to be easily analyzed and to have adjustable parameters represented by

some vector m . We rewrite P_0

$$P_0[s, m] = \frac{1}{Z_0} \exp(-U_0[s, m]/T), \quad (3)$$

where $Z_0 = \int ds \exp(-U_0[s, m]/T)$ which in general depends on m through U_0 .

The entropy of P_0 relative to P is

$$R = \int ds P_0[s, m] \ln \frac{P_0[s, m]}{P[s]}, \quad (4)$$

where we have suppressed the dependence of R on the vector of adjustable parameters m .

Using Equations 1 and 3, we rewrite Equation 4 as

$$R = \int ds \frac{\exp(-U_0/T)}{Z_0} (-U_0/T - \ln Z_0 + U/T + \ln Z). \quad (5)$$

We define the average with respect to P_0 of a function $\phi[s]$ as $\langle \phi \rangle = \int ds \phi \exp(-U_0/T)/Z_0$

and obtain

$$R = -\frac{1}{T} \langle U_0 - U \rangle - \ln Z_0 + \ln Z. \quad (6)$$

We define $F_0 = -T \ln Z_0$ and $F = -T \ln Z$ and obtain

$$R = \frac{1}{T} (F_0 - F + \langle U - U_0 \rangle), \quad (7)$$

It is known[8] that $R[m] \geq 0$ with equality holding if and only if $P_0 = P$. Here T is also positive so that

$$F \leq F_0 + \langle U - U_0 \rangle, \quad (8)$$

which is the basis of our mean field approximations to discrete, continuous, and even problems with both discrete and continuous variables.

The mean field approximation is obtained by minimizing Equation 7 with respect to m to find the tightest bound in Equation 8; mean field annealing involves tracking the minimum from high to low values of T [10]. In the case of discrete s_i as in graph coloring or binary image restoration[4] it is useful to choose

$$U_0 = - \sum_i m_i s_i, \tag{9}$$

but in the present context of problems with continuous s_i the simplest useful choice[5, 2] is

$$U_0 = \frac{1}{2} \sum_i (x_i - m_i)^2. \tag{10}$$

In either case the m_i are real.

4 Mean Field Annealing

Mean Field Annealing (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. SA works by gradually cooling an on-going 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[11, 1, 5, 2]. However because it is an approximation, MFA does not inherit any guarantee of convergence even when the analogous SA does converge.

4.1 Approximation of Systems of Discrete Variables

For combinatorial problems, the domain $S = \{0, 1\}^N$ is discrete and each solution might be represented as a vector s_1, s_2, \dots, s_N ones and zeros. A typical objective function might be

$$U[s] = - \sum_i h_i s_i - \sum_{ij} v_{ij} s_i s_j, \quad (11)$$

For this case, Equation 9 is appropriate since the average of any particular variable

$$\langle s_i \rangle = \frac{\sum_{s_i=0,1} s_i \exp(m_i s_i / T)}{\sum_{s_i=0,1} \exp(m_i s_i / T)} = \frac{1}{1 + \exp(-m_i / T)} \quad (12)$$

asymptotically approaches either one or zero for low T . This choice of U_0 is useful for combinatorial problems such as graph partitioning or binary image restoration[1, 4]. The vector m is chosen to minimize the right hand side of Equation 8 at each of a sequence of decreasing T 's The starting point of each new minimization is the terminal point of the previous minimization at slightly higher T .

4.2 Approximation of Systems of Real Variables

For the optimization of real-valued variables, the choice of form for the mean field must allow mean values of x between the extreme values as in Equation 10, $U_0 = \sum_i 1/2(s_i - m_i)^2$ or $U_0 = 1/2\|s - m\|^2$ with real means m_i . The average in Equation 8 inequality is weighted by the Gaussian density $Z_0^{-1} \exp(-U_0/T)$ now extends over all real configurations. Z_0 involves $\int_{-\infty}^{\infty} ds_1 \int_{-\infty}^{\infty} ds_2 \dots \int_{-\infty}^{\infty} ds_N$ but it factors so that the normalization is $Z_0 = \prod_i \sqrt{2\pi T}$. The logarithm of a product is a sum of logarithms so that $F_0 = -T \sum_i \ln \sqrt{2\pi T}$. The average U_0 is a sum of second moments $\langle U_0 \rangle = \sum_i T/2 = NT/2$. For this U_0 , both F_0 and $\langle U_0 \rangle$ are independent of m so the additive constant, $F_0 - \langle U_0 \rangle$, in Equation 8 is independent of the unknown m which becomes

$$F \leq W = \text{constant} + \langle U \rangle \quad (13)$$

and can be dropped. The result is that the mean field is defined by the minimum of

$$(\sqrt{2\pi T})^N \int ds U(x) \exp(-\|x - m\|^2/2T) \quad (14)$$

which is Kostrowicki and Piela's Equation 12, which they call the Fourier-Poisson formula.

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

ered. For large T fine structure of the original problem is averaged away and in this limit the 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.

5 Image Optimization

Optimization can be used to estimate a true image of N real pixel values from a similar image of noisy measurements, given certain generic information about the true scene and the noise. Let y be a noisy observation of a true scene s . If the observations are corrupted by independent, identically distributed, additive Gaussian noise, then $y_i = s_i + \epsilon_i$, where ϵ is a random variable distributed according to a density proportional to

$$\exp(-\|\epsilon\|^2/2\sigma^2). \quad (15)$$

If s is further known to be *e.g.* piecewise constant, then a *maximum a posteriori* s can be estimated by minimizing the negative of the logarithm of the Bayesian posterior density[3, 5]

$$U(s) = \frac{1}{2\sigma^2} \sum_i (y_i - s_i)^2 - \sum_{ij} V(s_i - s_j), \quad (16)$$

where the second sum is restricted to adjacent pixels i and j and $\epsilon_i = y_i - s_i$ has been used in the first term to express the noise in terms of the measurement and the unknown s . The function V depends on the problem but is positive in a small region around the origin and

is zero elsewhere. A useful choice for images is a Gaussian,

$$V(s_i - s_j) = \frac{b}{\sqrt{2\pi\tau}} \exp\left(-\frac{(s_i - s_j)^2}{2\tau}\right), \quad (17)$$

where b adjusts the strength of the prior expectation that adjacent pixels agree in value to within about τ units.

In the first term of Equation 16, all the averages in Equation 14 cancel except the one over s_i . The second term involves an integral over s_i and s_j of the product of three Gaussians which can be performed in succession using a table of integrals. The result is[3, 5]

$$\langle U \rangle = \frac{1}{2\sigma^2} \sum_i (y_i - m_i)^2 - \frac{b}{\sqrt{2\pi(T + \tau)}} \sum_{ij} \exp\left(-\frac{(m_i - m_j)^2}{2(T + \tau)}\right) \quad (18)$$

to within an additive term that is independent of m . The i^{th} component of the gradient of Equation 18 with respect to m is

$$\frac{\partial}{\partial m_i} \langle U \rangle = \frac{m_i - y_i}{\sigma^2} + \frac{2b}{\sqrt{2\pi(T + \tau)}^3} \sum_j (m_i - m_j) \exp\left(-\frac{(m_i - m_j)^2}{2(T + \tau)}\right), \quad (19)$$

where the sum on j is restricted to pixels adjacent to i .

At low $T + \tau$, Equation 18 typically exhibits so many local minima that conventional minimization techniques using Equation 19 terminate almost immediately from most starting points. Often the starting point and the associated terminal point cannot be distinguished by eye. However for large enough T , the second term becomes negligible so that $\langle U \rangle$ has a single minimum near $s = y$ and this minimum can be tracked if T is gradually reduced.

Using gradient descent to minimize $\langle U \rangle$ at each of a sequence of decreasing temperature, we obtain an low temperature estimate of the m that minimizes the original U .

Figure 1 is a plot of a 64×64 image of 4096 real variables which represent a step discontinuity of height 49.0. Figure 2 shows the same data degraded by additive Gaussian noise with zero mean and standard deviation 24.5, half the step height. Linear approaches to removing noise from Figure 2 invariably blur the step edge also. Nonlinear approaches based on Equation 16 exhibit local optima and require global optimization such as simulated annealing, which is is slow. Figure 3 is the result of minimizing Equation 16 by using mean field annealing of Equation 18. Clearly mean field annealing is able to recover the original image from the degraded image. Figure 3 was obtained using an initial $T = 24$ and a final $T = .24$ and a $v = 24$.

Piecewise-linear restorations are similarly obtained by replacing the first difference in the argument of the prior potential function V by the appropriate second difference[2].

6 Conclusions

The diffusion equation method of global optimization introduced by Kostrowicki and Piela has been related to the more familiar method of global optmization, simulated annealing. The mean field approximation was developed and used to convert stochastic simulated an-

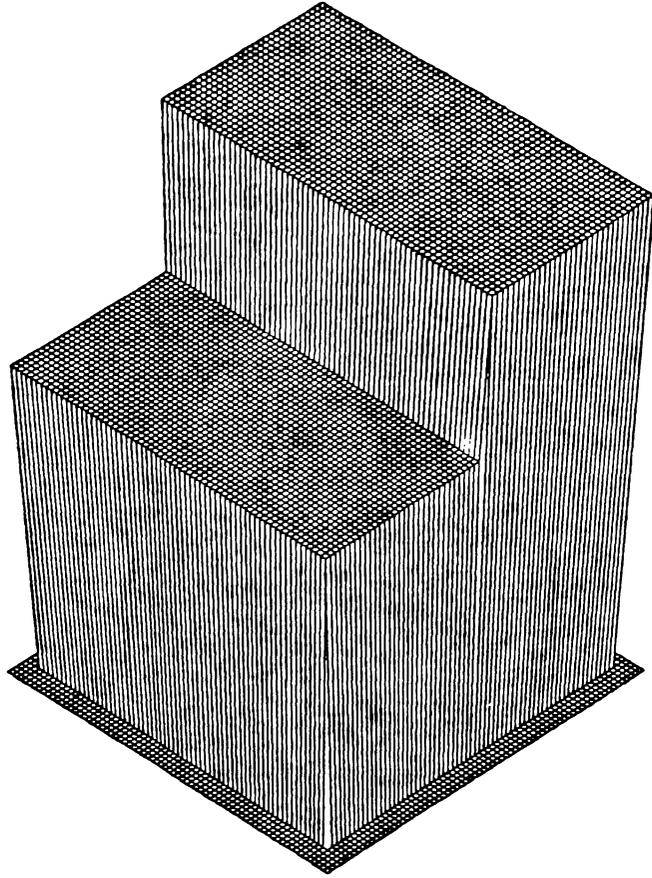


Figure 1: Original step image s

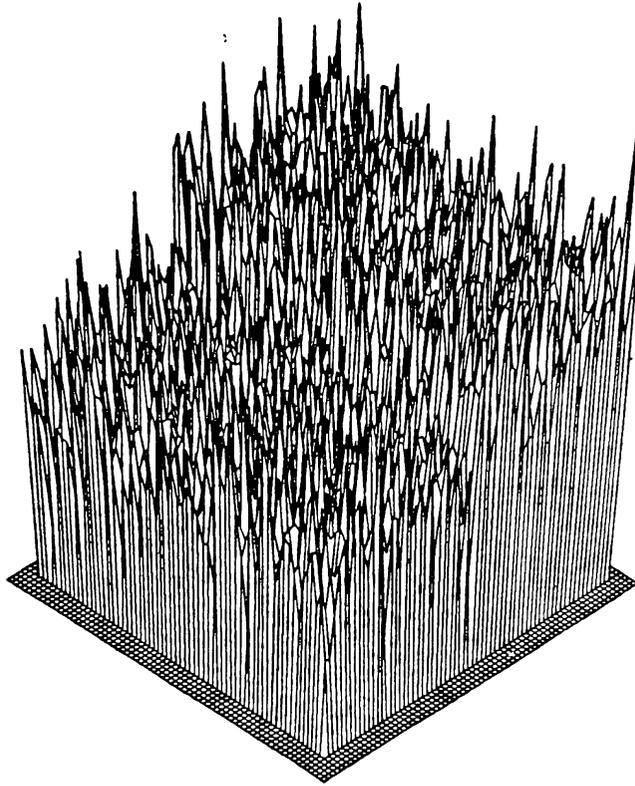


Figure 2: Degraded image y

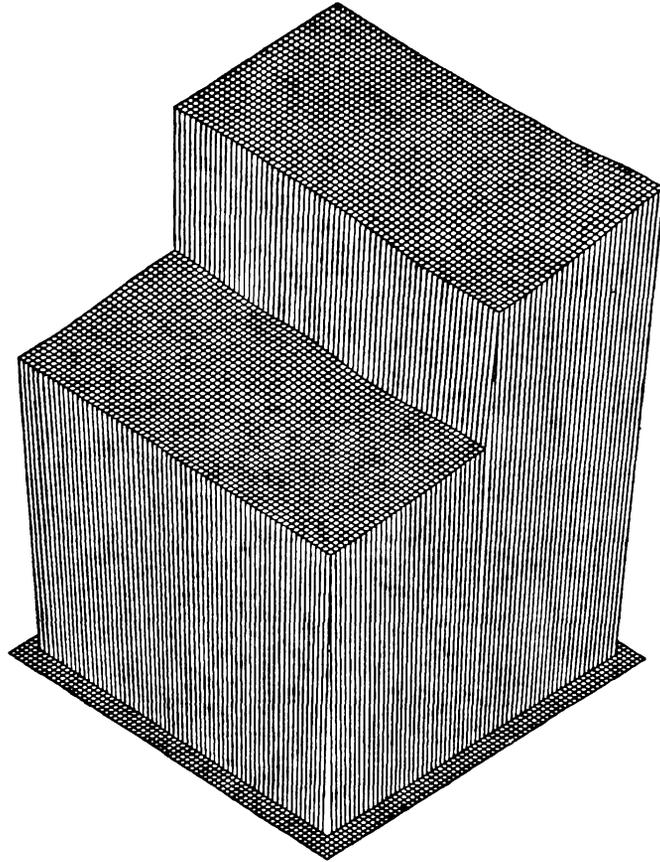


Figure 3: Mean field estimate of s , given y

nealing into a deterministic algorithm for optimizing real variables. Mean field annealing prescribes a sequence of local minimizations at decreasing “temperatures” of a Gaussian integral of the original objective function. Except for terminology (“time” for “temperature”), this is identical to Kostrowicki and Piela’s method. Experimental results for large image optimization problems were presented which support and extend Kostrowicki and Piela’s results.

7 Acknowledgements

We thank Stephen J. Garnier for writing the mean field annealing program and for producing the figures.

References

- [1] G. L. Bilbro, T. K. Miller, W. E. Snyder and D. E. Van den Bout, M. W. White, and R. C. Mann. Optimization by the Mean Field Approximation. In *Advances in Neural Network Information Processing Systems 1*, pages 91–98. Morgan-Kaufman, San Mateo, 1989. Reprinted from 1988 NIPS, Denver, CO.

- [2] Griff L. Bilbro and Wesley E. Snyder. Mean field annealing: An application to image noise removal. *Journal of Neural Network Computing*, 1990.
- [3] Griff L. Bilbro, Wesley E. Snyder, Steven J. Garnier, and James W. Gault. Mean field annealing: A formalism for constructing GNC-like algorithms. *IEEE Transactions on Neural Networks*, 3(1), 1992. scheduled to appear.
- [4] Griff L. Bilbro, Wesley E. Snyder, and Reinhold C. Mann. The mean field minimizes relative entropy. *Journal of the Optical Society of America A*, 8(2):290–294, February 1991.
- [5] H. P. Hiriannaiah, 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*, pages 1901–1912, December 1989.
- [6] S. Kirkpatrick, C. Gelatt, and M. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, May 1983.
- [7] J. Kostrowicki and L. Piel. Diffusion equation method of global minimization: Performance for standard test functions. *Journal of Optimization Theory and Applications*, 69(2):269–284, May 1991.
- [8] S. Kullback. *Information Theory and Statistics*. Wiley and Sons, 1959.

- [9] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Chem. Physics*, 21:1087–1092, 1953.
- [10] W. E. Snyder, G. L. Bilbro, and D. E. Van den Bout. New techniques in optimization. Technical Report NETR89-12, Center for Communications and Signal Processing, NCSU, Raleigh, 27695, 1989.
- [11] C. M. Soukoulis, K. Levin, and G. S. Grest. Irreversibility and metastability in spin-glasses. I. ising model. *Phys. Rev. B*, 28(3):1495–1509, August 1983.