Fast Stochastic Restoration of Gray Level Images

Griff L. Bilbro

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

TR-90/15
November 1990

Fast Stochastic Restoration of Gray Level Images

Griff Bilbro*
Department of Electrical and Computer Engineering
North Carolina State University, Raleigh, NC 27695-7914.

Abstract

A new stochastic technique is described for the Bayesian restoration of gray level images corrupted by white noise. The proposed technique is related to simulated annealing but generates candidates more efficiently for gray level images than either the Gibbs sampler or the Metropolis procedure. Convergence for the resulting homogeneous Markov chain is proved. Accelerated convergence is also proved in some cases. Experimentally the new technique is shown to restore floating point images in 1/50 of the time required for the usual simulated annealing. Experimental restorations of gray level images corrupted by white noise are presented.

1 Introduction

The theory of stochastic relaxation will be extended to allow generation of candidates from an arbitrary positive probability distribution or *generator*. The extended theory will be shown to contain the usual Metropolis procedure,

*Supported in part by U. S. Army Contract DAAL01-89-K-0906, by DARPA Contract MDA972-88-C-0047, and by the Center for Communications and Signal Processing at North Carolina State University.

the Gibbs sampler, and the "Cauchy machine" as special cases. Asymptotic correctness of the resulting algorithm will be demonstrated. When the generator closely approximates the true Gibbs distribution, convergence of the new algorithm will be shown to be exponentially fast. A simple improvement to the Metropolis generator will be shown to accelerate convergence by a factor of 50. As a practical example, the theory is applied to the Bayesian image restoration of piecewise smooth gray level images.

Stochastic relaxation with annealing[7] produces excellent restorations of noisy images but its utility is limited by long computation time. Consequently, pixel values are typically restricted to fewer than 5 discrete gray levels and often only 2 or 3 levels[16][10]. This is conspicuously less than ordinary gray level images with 256 gray levels. This report introduces a stochastic algorithm that is fast enough for the Bayesian restoration of gray level and floating point images.

Previous approaches to accelerating stochastic relaxation with annealing (SRA) include deterministic modeling of the relevant Gibbs distribution within the mean field approximation. The resulting mean field annealing (MFA) alogrithm has been previously found faster than SRA for some optimization problems[1] and has been applied to the Bayesian restoration of piecewise constant images[8], piecewise linear images[3], and binary images[4]. In image restoration the practical advantage of MFA over SRA appears clear[11], but sufficient conditions for convergence of MFA have remained elusive. Rigorous understanding of accelerated SRA seems more promising for the algorithm recently used to solve a parameter extraction problem for microwave transitors[2]. That algorithm was called Tree Annealing (TA) because it employed a binary tree to maintain an approximation to the relevant Gibbs distribution which it refined by annealing. Unfortunately TA cannot be extended to the high dimensional search space of image restoration even though it has been applied to other image problems[12]. The present paper combines these two earlier lines of development by using the fast stochastic techniques of TA with an MFA-like model of the Gibbs distribution. The new technique enjoys the speed of MFA and the robustness of SRA. Moreover convergence of the proposed technique can be understood theoretically.

The procedure proposed in this report differs from existing approaches. The descendents of TimberWolf raise efficiency by adjusting the range of the generator to prefer smaller perturbations at lower temperatures[9]. Vander-

bilt and Louie's algorithm raises efficiency with a generator which concentrates the search along a preferred direction[15]. The new theory will admit *any* positive generator, including the Lorentz distribution of the "Cauchy machine"[13]. The new algorithm is different from the Gibbs sampler which generates candidate values for a pixel according to the exact Gibbs distribution conditioned by the current values of other pixels[7] The cost of the Gibbs sampler may increase as fast as the number of gray levels times the square of the size of the cliques[16]. For images with many gray levels (many allowed states per pixel) or for complex prior distributions, it will be shown that correcting an approximate generator can be faster that using an exact generator.

2 Development of the general theory

In the familiar Metropolis simulation, a candidate state $F = y$ is generated by sampling a uniform distribution $\gamma(y|x)$ given the current state $F = x$. The candidate is accepted with probability

$$\min\left(1, \frac{\pi(y)}{\pi(x)}\right), \tag{1}$$

where π is the Gibbs density and γ is required to exhibit the symmetry[14]

$$\gamma(y|x) = \gamma(x|y) \tag{2}$$

to ensure asymptotic convergence to π. The Gibbs sampler[7] does not preserve Equation 2 and the following theory further relaxes the restrictions on γ.

Let the form of the Markov transition matrix from state x to state y be written as the product

$$P(y|x) = \gamma(y|x)a(y|x). \tag{3}$$

Here $\gamma(y|x)$ is the probability of generating state y as a candidate given x as the current state, and $a(y|x)$ is the probability of accepting y as the next state given x as the current state. The goal is to use P to sample the Gibbs distribution π and for this π must be a stationary state of P. Given γ, it will suffice to construct P so that the *detailed balance condition* is satistfied

$$P(y|x)\pi(x) = P(x|y)\pi(y). \tag{4}$$

Combining Equations 3 and 4 yields a condition on a

$$\frac{\pi(x)\gamma(y|x)a(y|x)}{\pi(y)\gamma(x|y)a(x|y)} = 1. \tag{5}$$

This does not uniquely specify a, which can be chosen to have a form similar to Equation 1

$$a(y|x) = \min(1, q(y|x)) \tag{6}$$

for some q which can be chosen to obey

$$q(y|x) = \frac{1}{q(x|y)} \tag{7}$$

like the ratio of functions in Equation 1. This allows the separation of Equation 5 into three cases $q(y|x) > 1$, $q(y|x) = 1$, and $q(y|x) < 1$ to obtain

$$q(y|x) = \frac{\gamma(x|y)\pi(y)}{\gamma(y|x)\pi(x)}. \tag{8}$$

2.1 Asymptotic convergence

The Markov chain resulting from the transition matrix informally constructed in the preceeding paragraph can be shown to have the following desirable behavior:

Theorem 1 : The Gibbs distribution π (Equation 28) is a stationary distribution of the transition matrix

$$P(y|x) = \begin{cases} \gamma(y|x)\min(1, q(y|x)) & \text{if } y \neq x \\ 1 - \sum_{z \neq x} P(z|x) & \text{if } y = x \end{cases}, \tag{9}$$

where q is given by Equation 8

$$q(y|x) = \frac{\gamma(x|y)\pi(y)}{\gamma(y|x)\pi(x)}.$$

and where γ is a positive definite but otherwise arbitrary distribution.

Proof: Consider the ratio of transitions between the states x and y for arbitrary probability density p

$$\frac{P(y|x)p(x)}{P(x|y)p(y)}. \tag{10}$$

By construction of q in Equation 8, this ratio is unity when $p = \pi$, the Gibbs' density. This implies that $p = \pi$ is a stationary distribution of P, since if

$$P(y|x)\pi(x) = P(x|y)\pi(y), \forall x, y, \tag{11}$$

then the sum over x is

$$\sum_x P(y|x)\pi(x) = \sum_x P(x|y)\pi(y). \tag{12}$$

But $\sum_x P(x|y) = 1$, so that the right hand member of this equation is simply $\pi(y)$ and thus π is a stationary state of P as desired. □

Note that only the ratio of π's appear in Equation 8, so that the normalization Z is not necessary for the procedure, just as in the usual Metropolis procedure. Moreover the Markov chain will actually converge to the Gibbs' distribution, as is shown in the following result.

Theorem 2: The finite homogeneous Markov chain generated by transistion matrix of Equation 9 convergerves asymptotically to $\pi(x)$.

Proof: First Equation 9 is shown to be a properly normalized transition probability. Since γ is a distribution $\sum_y \gamma(y|x) = 1$ and since $\min(1, q(y|x)) \leq 1$, then $\sum_{y \neq x} P(y|x) \leq 1$. Then by construction of the $y = x$ case, $\sum_y P(y|x) = 1$.

Now convergence is shown. It is known[14] that detailed (or local) balance Equation 4 guarantees the weaker condition of global balance. This in turn implies[6][14] that that π is the unique stationary distribution of the finite homogeneous Markov chain associated with P if P is *aperiodic* and *irreducible*. A Markov chain is aperiodic if[5]

$$\exists x : P(x|x) > 0. \tag{13}$$

By separately accounting for the omitted term in the sum in Equation 9, the $y = x$ member can be rewritten as

$$P(x|x) = 1 + \gamma(x|x) - \sum_y^{q(y|x) \geq 1} \gamma(y|x) - \sum_y^{q(y|x) < 1} \gamma(y|x) q(y|x). \tag{14}$$

Since γ is a distribution,

$$1 = \sum \gamma(y|x) = \sum_y^{q(y|x) \geq 1} \gamma(y|x) + \sum_y^{q(y|x) < 1} \gamma(y|x), \tag{15}$$

which is used to rewrite the first sum. A line of algebra results in

$$P(x|x) = \gamma(x|y) + \sum_{y}^{q(y|x)<1} \gamma(y|x)(1 - q(y|x)). \tag{16}$$

This implies that $P(x|x) > 0$ since γ is positive definite by assumption and $q(y|x) < 1$ under this summation. So that Equation 13 is satisfied by any x and the chain is therefore aperiodic.

A Markov chain is irreducible[14] (page 19) if and only if

$$\forall x, y \; \exists n : 1 \leq n < \infty \; \wedge \; P^n(y|x) > 0. \tag{17}$$

It was earlier shown that $P(x|x) > 0$. Furthermore $P(y|x) > 0$ for $y \neq x$ in Equation 9, because $\gamma(y|x) > 0$ by assumption and $q(y|x) > 0$ because it is the ratio of strictly positive quantities. Therefore $P(y|x)$ never vanishes and Equation 17 is satistfied with $n = 1$, so that the chain is also irreducible. It follows that in the limit of infinite transitions, the Markov chain induced by Equation 9 converges to this stationary Gibbs distribution. □

Theorem 1 shows that the proposed Metropolis-like procedure has the correct stationary distribution independently of the particular generator used in the procedure. Theorem 2 shows that the proposed procedure actually converges to this Gibbs distribution in infinite time. This is sufficient to show convergence of a "homogeneous simulated annealing algorithm" which involves an infinite homogeneous Markov chain at each of a sequence of temperatures decreasing to zero[14].

Theorem 3: Let

$$\pi(x) = \frac{1}{Z} \exp\left(-\frac{U(x)}{T}\right) \tag{18}$$

where Z is a normalization factor which depends only on the temperature T. Then the usual homogeneous annealing algorithm[14] based on Equation 9 converges to a global minimum of U with probability 1.

Proof: For any $T > 0$ the Markov chain associated with the proposed procedure converges to π at each temperature in infinite time by Theorem 2. In the limit of low temperature, π vanishes except on the set of globally minimal states of U[14]. It follows that if the temperature is reduced to zero, the distribution converges to a global minimum of U with probability 1 if the chain is infinitely long at each temperature.□

The proposed procedure contains the usual Metropolis procedure, the Gibbs sampler, and other schemes as special cases. In the first case, $\gamma(y|x)$ is uniform over the set of y in the *neighborhood of* x, the states that can be reached from x, and zero otherwise. The size of neighborhoods is fixed and y is a neighbor of x if and only if x is a neighbor of y, so that $\gamma(y|x) = \gamma(y|x)$ and the ratio cancels in Equation 8. This is also true when $\gamma(y|x)$ is taken to be a multivariate Gaussian[15] or a Lorentz distribution[13] since both are even in the difference $y - x$. In the case of the Gibbs sampler $\gamma(y|x)$ is taken to be the true Gibbs distribution for a single pixel conditioned by fixing all other pixel gray levels in their current state. The states $F = x$ and $F = y$ differ only in the gray level of a particular pixel and $\gamma(y|x) \propto \pi(y)$ with a proportionality constant equal to the joint probability of the fixed pixels. But $\gamma(x|y) \propto \pi(x)$ with the same proportionality constant since the same pixels are fixed. Therefore the entire numerator and denominator in Equation 8 cancel and $q = 1$ so that the Gibbs sample candidate is always accepted (but at the expense of actually evaluating every possible grey level for the pixel in question).

2.2 Finite time behavior

Consider the dynamics of a probability distribution stochastically evolving at fixed temperature T according to the transition matrix Equation 9, so that p is a function of time t. Because T is fixed, so is the Gibbs distribution π. The generator γ is also assumed fixed so that the transistion matrix P of Equation 9 is fixed also. The rate at which the probability of a particular state or image $F = x$ changes is the difference of the rate of transitions into x per unit time minus the rate out of x

$$\frac{dp(x,t)}{dt} = \sum_y P(x|y)p(y,t) - \sum_y P(y|x)p(x,t). \tag{19}$$

If the generator is identical to the target distribution $\gamma(y|x) = \pi(y), \forall x, y$, then $P(y|x) = \gamma(y|x)\min(1,1) = \pi(y)$, so that Equation 19 reduces to the special case

$$\frac{dp(x,t)}{dt} = \sum_y \pi(x)p(y,t) - \sum_y \pi(y)p(x,t) = \pi(x) - p(x,t), \tag{20}$$

a differential equation with a well known exponential solution, $p(x,t) = \pi(x) + (p(x,0) - \pi(x))e^{-t}$. In this case, $p(x,t)$ relaxes from an arbitrary initial distribution to the Gibbs distribution $\pi(x)$ exponentially fast with a characteristic time of unity or one update per pixel. This motivates the following result that if γ approximates π, the transition matrix Equation 9 may converge faster than the usual Metropolis procedure in which γ is uniform.

Theorem 4: If the generator γ of the finite transition matrix Equation 9 approximates the Gibbs distribution in the sense that the deviation

$$\delta(x|y) \equiv \frac{\gamma(x|y)}{\pi(x)} - 1 \tag{21}$$

obeys $|\delta(x|y)| \leq \overline{\delta}, \forall x, y$ for some positive constant $\overline{\delta} < \frac{1}{2}$, then the Markov chain induced by Equation 9 converges from arbitrary initial p to the Gibbs distribution faster than an exponential with characteristic time $\tau = 1/(1 - 2\overline{\delta})$.

Proof: Assuming positive distributions, substitution of Equation 9 into Equation 19 results in

$$\frac{dp(x,t)}{dt} = \sum_y (p(y,t)\pi(x) - p(x,t)\pi(y)) \min\left(\frac{\gamma(x|y)}{\pi(x)}, \frac{\gamma(y|x)}{\pi(y)}\right), \tag{22}$$

after a little algebra. If the deviation of the evolving distribution from the asymptotic Gibbs distribution is defined as

$$\epsilon(x,t) \equiv \frac{p(x,t)}{\pi(x)} - 1 \tag{23}$$

then Equation 22 can be rewritten in terms of δ and ϵ as

$$\frac{d\epsilon(x,t)}{dt} = -\epsilon(x,t) + \sum_y \pi(y)\,(\epsilon(y,t) - \epsilon(x,t))\min\,(\delta(y|x), \delta(x|y)). \tag{24}$$

At each time define $\overline{\epsilon}(t) = \max_x\,(|\epsilon(x,t)|)$. If $p \neq \pi$ the largest component of the fractional deviations $\epsilon(x,t)$ is either positive or negative. If it is positive then $\exists \overline{x}$ such that $\overline{\epsilon}(t) = \epsilon(\overline{x}, t)$ which is changing in time at the rate

$$\frac{d\overline{\epsilon}(t)}{dt} = -\overline{\epsilon}(t) + \sum_y (\pi(y)\epsilon(y,t) - \pi(y)\overline{\epsilon}(t))\min\,(\delta(y|x), \delta(x|y)). \tag{25}$$

But $\epsilon(y,t) \leq |\epsilon(y,t)| \leq \overline{\epsilon}(t)$ and $\min\left(\delta(x|y), \delta(y|x)\right) \leq \overline{\delta}$ so that

$$\frac{d\overline{\epsilon}(t)}{dt} \leq -\overline{\epsilon}(t) + 2\sum_y \pi(y)\overline{\epsilon}(t)\overline{\delta} \tag{26}$$

and $\sum_y \pi(y) = 1$ so that

$$\frac{d\overline{\epsilon}(t)}{dt} \leq -\left(1 - 2\overline{\delta}\right)\overline{\epsilon}(t) = -\frac{\overline{\epsilon}(t)}{\tau}. \tag{27}$$

On the other hand, if $\overline{\epsilon}(t)$ is attained for a negative $\epsilon(x,t)$, then for some $\overline{x}$, $\overline{\epsilon}(t) = -\epsilon(\overline{x},t)$ which leads again to Equation 27. Therefore for arbitrary probability distribution p and arbitrary time, the bound $\overline{\epsilon}(t)$ on the magnitude of the fractional deviation of p relative to π decreases faster than an exponential with characteristic time τ as claimed. □

Theorem 4 guarantees fast equilibration of the proposed stochastic relaxation procedure when a good enough generator γ can be found. This result applies only when it is possible to guarantee that γ has a $\overline{\delta} < 1/2$. However the following experimental work suggests a similar general dependence of convergence time on the quality of the generator.

3 Application to Bayesian restoration of images

Geman and Geman provided the general framework for stochastic Bayesian restoration of images modeled as Markov random fields[7]. Let F be the original gray level image and G be the observed image degraded by additive independent identically distributed (iid) Gaussian noise of zero mean and standard deviation σ. Model both F and G as Markov random fields so that if $G = g$ is the observed image of N pixels $g = (g_1, g_2, ..., g_N)^T$, then the probability of a particular sample realization of the original random field $F = f = (f_1, f_2, ..., f_N)^T$, is given by a posterior distribution which can be written as a Gibbs distribution

$$\pi(f) = \frac{1}{Z}\exp\left(-\frac{U(f,g)}{T}\right). \tag{28}$$

Here Z is a normalization factor which depends only on the temperature T, and the posterior energy function

$$U(f,g) = U_1(f,g) + U_2(f) \tag{29}$$

is the sum of a likelihood energy written in terms of the usual 2-norm

$$U_1(f,g) = \frac{1}{2\sigma^2} \parallel f - g \parallel^2 \tag{30}$$

and a prior energy written as a sum over cliques

$$U_2(f) = \sum_C V_C(f) \tag{31}$$

of potentials $\{V_C\}$.

The original image will be assumed to be piecewise smooth and modeled without line processes but with four kinds of cliques involving pairs of pixels: horizontal first neighbors, vertical first neighbors, and the two diagonal second neighbors. In this paper all the V_C have the same form $V_C(f) = v(\Delta_C)$ where $\Delta_C = f_{C1} - f_{C2}$ is the difference in gray values of the appropriate pixels and the scalar function $v(\Delta)$ is

$$v(\Delta) = \begin{cases} -(1 - |\Delta|/d)\ h & \text{if } |\Delta| < d \\ 0 & \text{otherwise} \end{cases}, \tag{32}$$

The constants d and h are the width and depth of every potential.

3.1 Algorithm

When the noise is additive iid Gaussian, the likelihood makes a computationally efficient generator because it is easy to sample a Gaussian. Rather than settle on the likelihood immediately, define the generator or probability of generating a sample realization of the original undistorted image $F = f$ in terms of a generation energy U_0

$$\gamma(f) = \frac{1}{Z_0} \exp\left(-U_0(f)\right) \tag{33}$$

where Z_0 is a normalization constant and

$$U_0(f) = \frac{1}{2s^2} \parallel f - g \parallel^2 \tag{34}$$

so that U_0 differs from U_1 only in the replacement of σ by s. Experimental results will show that $s = \sigma$ is the optimal assignment for the width of the generator. Note that γ has no factor of T in the exponent.

Algorithm A

Estimate the original image $f = (f_1, f_2, ..., f_N)^T$ given the observed image $g = (g_1, g_2, ..., g_N)^T$, an initial estimate of f, the standard deviation of the actual noise σ, the standard deviation of the generator s, a schedule $T_{initial}$, T_{final}, r, the clique potentials defined by the two parameters d and h.

1. $T \leftarrow T_{initial}$.
2. For every interior pixel location i
   (a) Generate a candidate from the likelihood $new \leftarrow Gaussian(g_i, s)$.
   (b) Calculate the change in generation energy $\Delta U_0 = (new - f_i)^2/(2s^2)$ from Equation 34 as the change in U_0 if the value f_i at i were replaced by the candidate value new.
   (c) Calculate the change in posterior energy ΔU as the change in U of Equation 29 if the value f_i at i were replaced by the candidate value new. This involves evaluating Equation 32 for the 8 cliques in Equation 31 to which i belongs.
   (d) Combine them as $\Delta E \leftarrow \Delta U/T - \Delta U_0$
   (e) If $Uniform(0, 1) < \exp(-\Delta E)$ then $f_i \leftarrow new$.
3. $T \leftarrow r * T$.
4. If $T > T_{final}$ go to 1.
5. Report the final estimate f. Stop.

Algorithm A uses a random number generator to generate Gaussian deviates of specified mean and standard deviation and uniform deviates in the unit interval. For simplicity in step 2, only *interior* pixels are updated. The error committed by this simplification prevents restoration at the edges of the image and degrades restoration near the edges. The effect is noticeable

in very small images such as Figures 4 and 5, but is not usually important for larger images. If edge effects are critical, the appropriate treatment of the smaller cliques near edges is known[7].

Algorithm A was implemented in C on a DECstation 3100, a RISC architecture desktop computer. It was applied to real-valued images even though some parts of the theory section are restricted to pixels which assume only discrete gray levels.

For the simple clique potential of Equation 32, satisfactory results are obtained anywhere in a region around $d = h = s = T_{initial} = \sigma$, $T_{final} = \sigma/10$, and $r = .9$ for high signal to noise or $r = .99$ for low signal to noise. It is sufficient to initialize the current state as the observed image, $f \leftarrow g$ with no preprocessing of any kind.

3.2 Dependence on the generator

Wolberg and Pavlidis observed that the Gibbs sampler has a computational complexity for each move that is linear in the number of gray levels that each pixel is allowed to search[16]. It is interesting to ask what the overall computation time for a complete restoration is as a function of the number of gray levels. Since the generator is Gaussian, it is approximately uniform in a region of width s around the observed image g and zero elsewhere. For $s >> \sigma$, the difference between g and f is insignificant because it is of order σ. In that regime, the generator is indistinguishable from a uniform perturbative generator around the current state f, so that the behavior of the usual Metropolis-based simulated annealing algorithm with uniform generator can be estimated. The dependence of total restoration time versus s is plotted in Figure 1 for the polygonal foreground of value 10 above a background of value 0 shown in Figure 2, degraded by additive iid Gaussian noise of zero mean and standard deviation 3, as shown in Figure 3. The upper curve (labeled "max") in Figure 1 shows the time to reliably obtain a restoration that differs from the original image by less than 1 gray level except for at most 1 percent of the pixels and corresponds to the restoration showns in Figure 4. The lower-curve (labeled "min") shows the time to obtain restorations with 98 percent of the pixels within 2 gray levels of the original and corresponds to Figure 5. For either curve in Figure 1, the optimal value of s is practically equal to the $\sigma = 3$ of the noise added to the original image. At $s = 128$, the generator is about as wide as in a Metropolis based simulated annealing

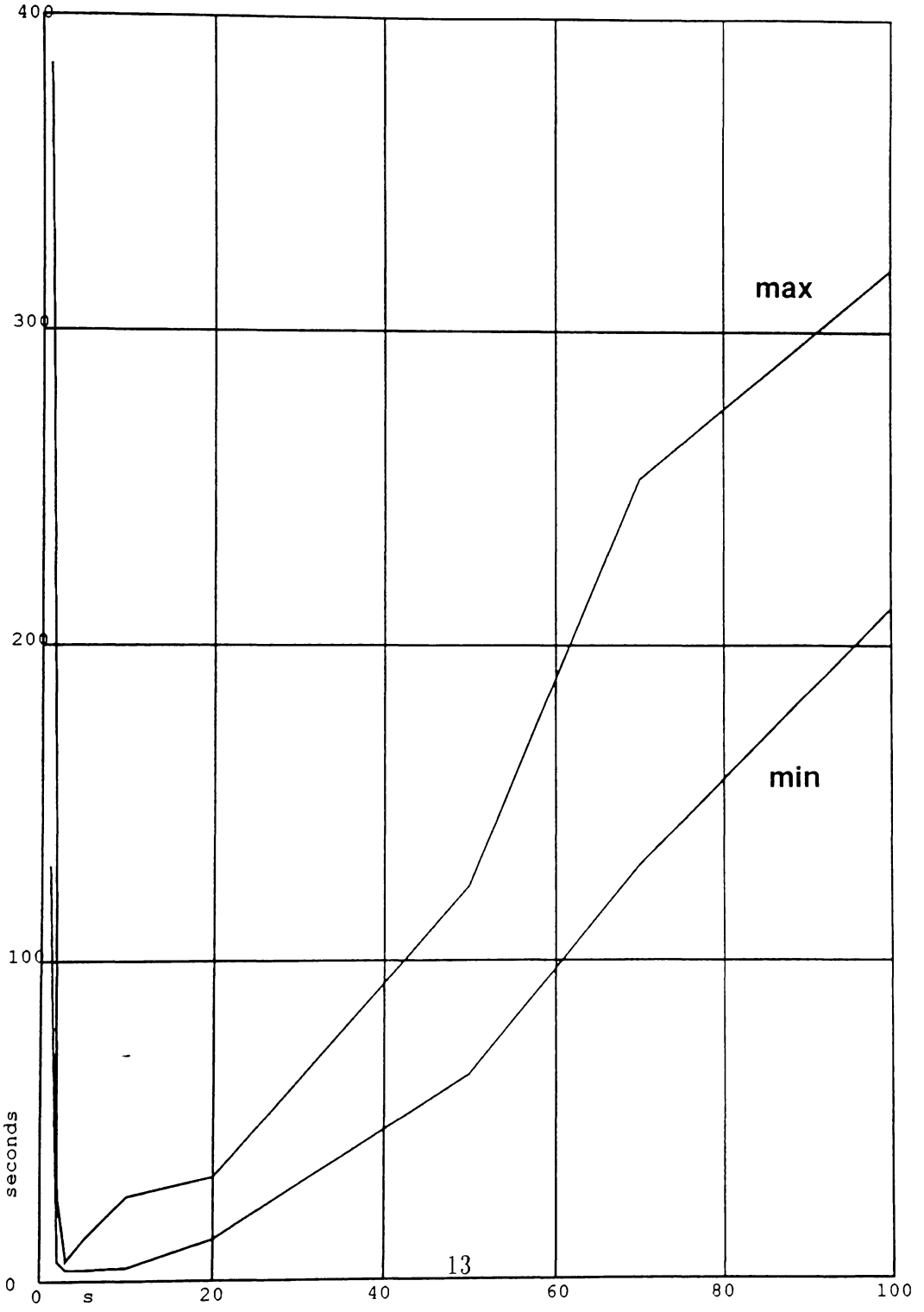

Figure 1: A plot of restoration time in seconds versus the standard deviation s of the generator. Fastest restoration occurs near $s = \sigma$. Restoration quality is somewhat subjective and the two curves attempt to bracket the time required for good restorations.

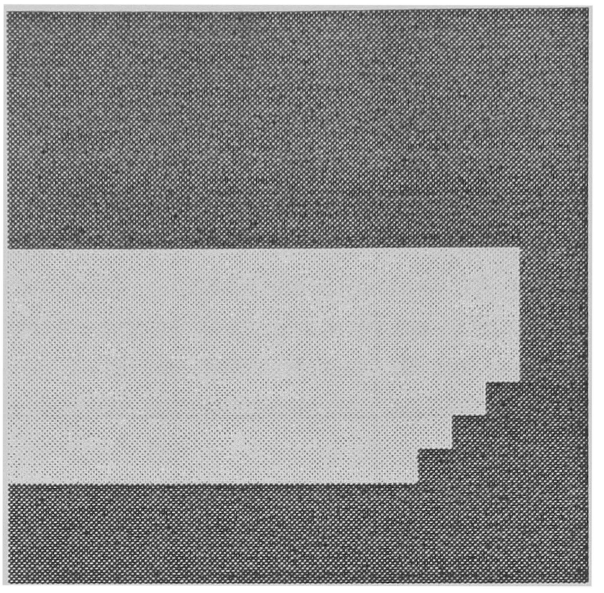

Figure 2: The original 17x17 image used in the experiments presented in Figure 1 of restoration time versus generator width s. Foreground at gray level 10, background at 0.
level 10, background at 0.

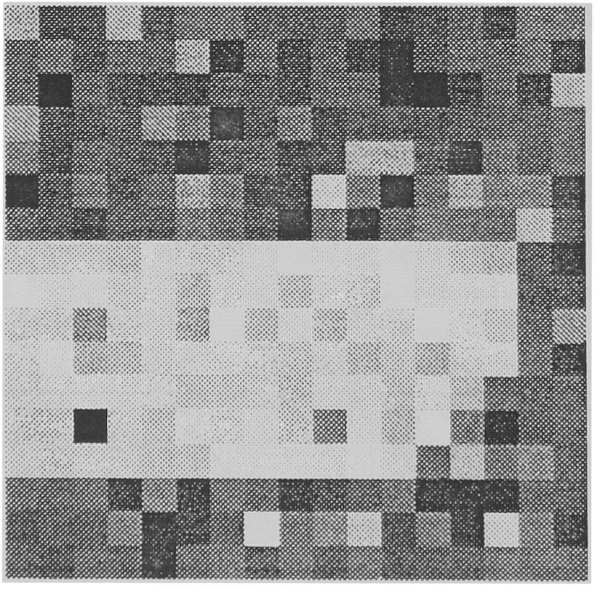

Figure 3: Figure 2 plus Gaussian noise of standard deviation $\sigma = 3$.

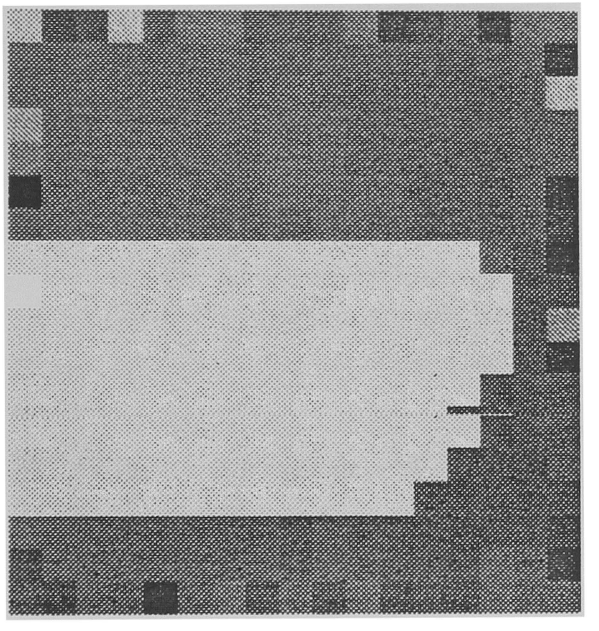

Figure 4: Restoration of Figure 3 requiring 3 seconds at $s = \sigma = 3$. Note that the edge pixels are not varied by the version of Algorithm A that is described in the text.

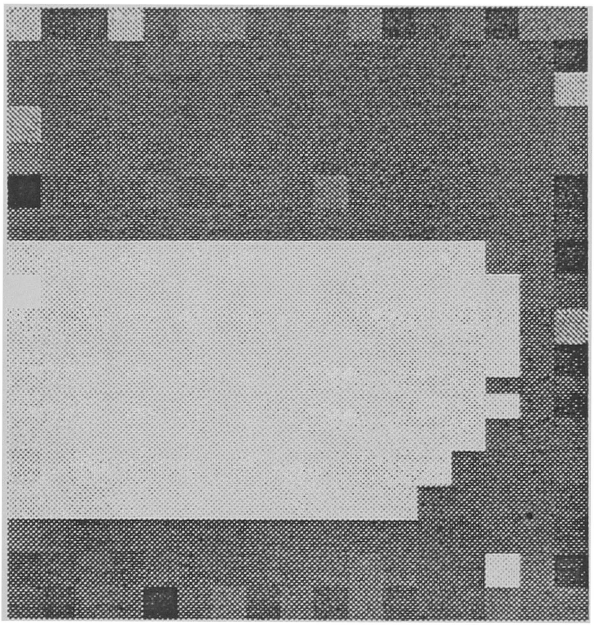

Figure 5: Restoration of Figure 3 requiring 1 second at $s = \sigma = 3$. Note that the edge pixels are not varied by the version of Algorithm A that is described in the text.

algorithm with uniform generator over the state space $(0, 1, 2, ..., 255)$. Figure 1 indicates that at $s = 100$, a uniform Metropolis algorithm runs about 50 times longer than at $s = \sigma$. It therefore appears that the total execution time is roughly linear in the effective size of the search space s, just as the time per update. Experience shows that shape of Figure 1 holds for larger images and in general. This particular experiment was conducted on so small an image because (17 by 17 pixels) only because of the inconveniently long run times except near the optimal value of s.

3.3 Restoration of images

In this section Algorithm A is applied to several types of realistic grey level images. Algorithm A was coded as described in Section 3.1 with the simple piecewise-smooth clique potential of Equation 32 and without line processes. The cooling rate was picked between $.9 < r < .99$ to trade off speed and quality. The other control parameters set as described in Section 3.1.

All the run tims can be halved by accelerating the cooling schedule with results *usually* comparable to those shown. Such faster schedules are not recommended however because 10 or 20 percent of the faster restorations are grossly erroneous; it is safer to run conservatively as recommended in Section 3.1, unless the better of two runs can be chosen. Slower restorations are marginally better.

3.3.1 A cartoon image

Figure 6 is a 64x64 cartoon image taken with a conventional video camera and frame grabber. The original image has a minimum value of 35, a maximum of 252, a mean of 42, and a standard deviation of 44. It was corrupted in Figure 7 with zero mean Gaussian noise of standard deviation 10 and restored in Figure 8 with Algorithm A in 23 seconds. Note the preservation of sharp edges and the smoothing of interior regions due to the simple piecewise smooth potential of Equation 32.

3.3.2 A Shepp-Logan head phantom

The Shepp-Logan head phantom of Figure 9 is a video scaled 64x64 image with a minimum value of 0, a maximum of 255, a mean of 157, and a standard

Figure 6: Original 64x64 cartoon video image.

Figure 7: Figure 6 plus Gaussian noise of standard deviation 10.

Figure 8: Restoration of Figure 7 requiring 23 seconds.

deviation of 46. The thin outer ring has intensity 255 and sets the dynamic range of the image. Inside the ring, the top circle has gray level 31, the range of the image. Inside the ring, the top circle has gray level 31, the two side ellipses have 37, and the tiny features near the center of the image have 47. Figure 9 was corrupted by iid Gaussian noise of standard deviation 5 (almost as much as the gray level difference of 6 between the circle and ellipse) as in Figure 10 and restored as shown in Figure 11 in 43 seconds.

3.3.3 A standard outdoor scene

The outdoor scene of Figure 12 is a 87x79 image with minimum value 32, maximum 232, mean 157, and standard deviation 46. It was corrupted in Figure 13 with noise of standard deviation 10 and restored as Figure 14 with Algorithm A and the piecewise smooth prior of Equation 32 in 47 seconds.

4 Conclusion

The theory of stochastic relaxation has been extended. The usual algorithm has been accelerated enough to make stochastic Bayesian restoration of real-

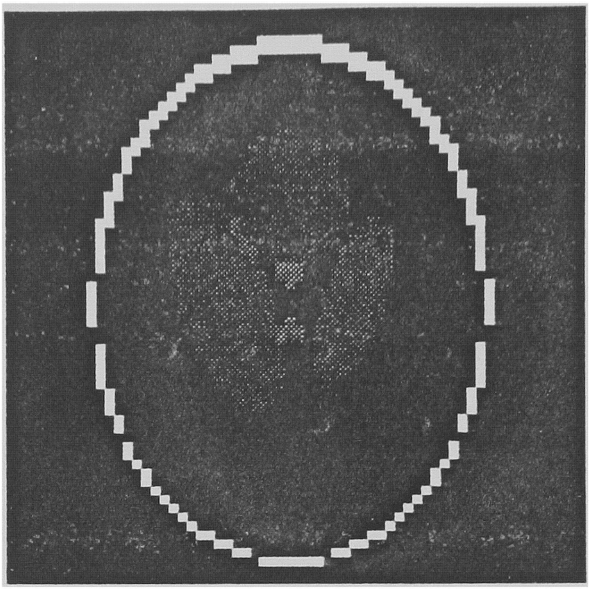

Figure 9: A video-scaled 64x64 synthetic Shepp-Logan head phantom image. Outer circle is at grey level 255; upper disk at level 31; ovals at level 37; smaller features at 47.

smaller features at 47.

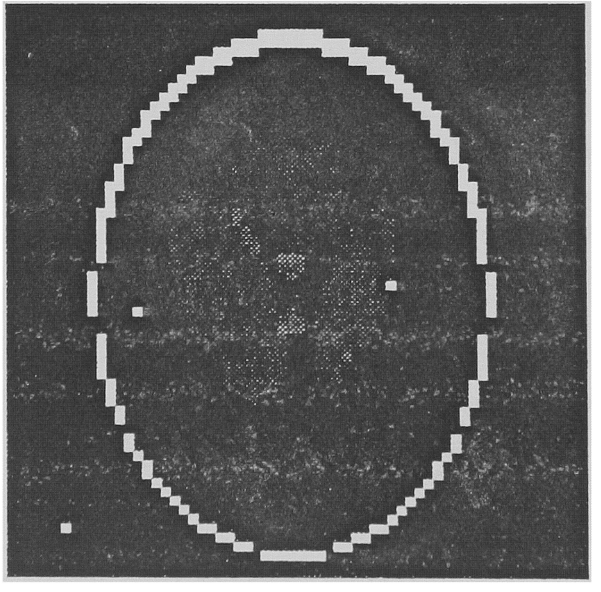

Figure 10: Shepp Logan plus Gaussian noise of standard deviation 5

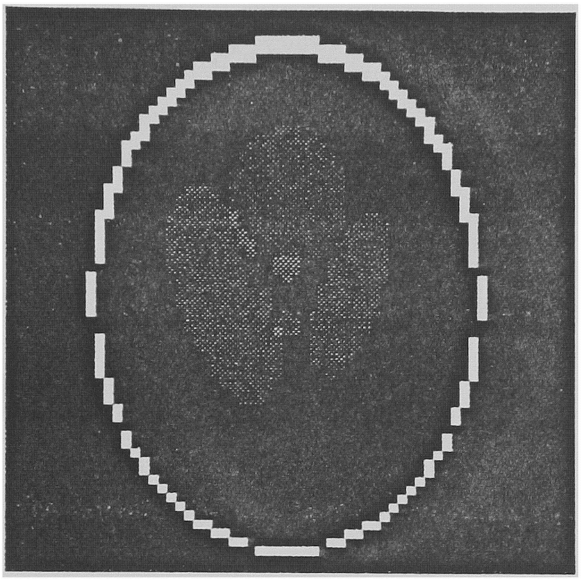

Figure 11: Restoration of Figure 10

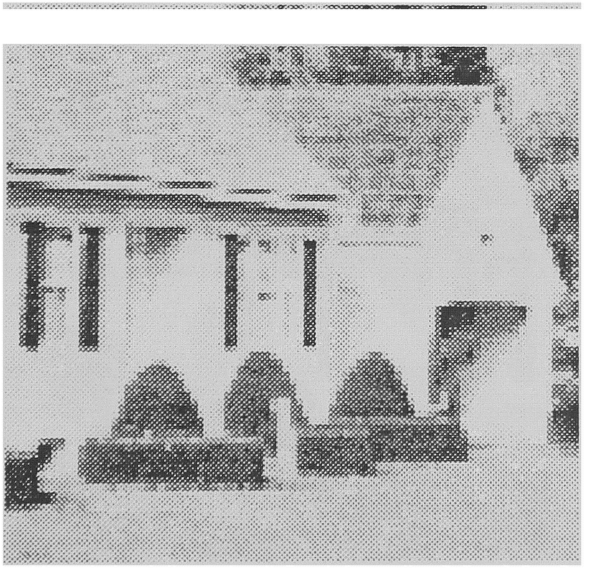

Figure 12: Standard image of a 87x79 section of the standard digitized outdoor scene.

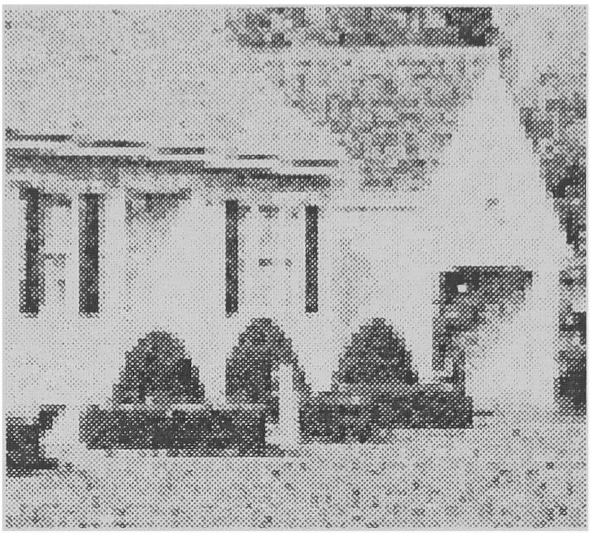

Figure 13: Observed data plus white Gaussian noise with standard deviation of 10

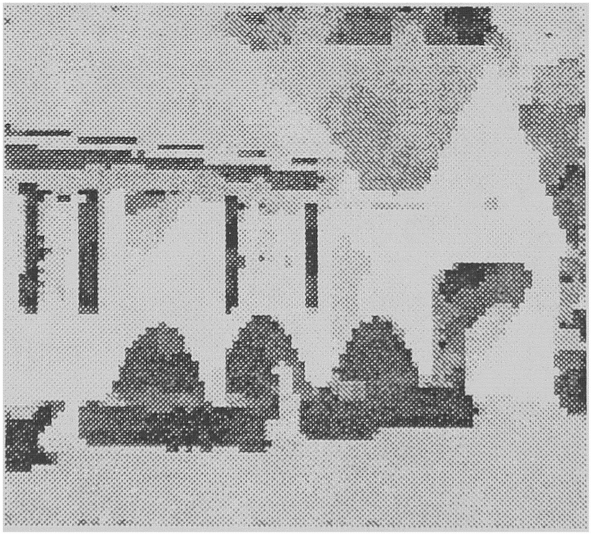

Figure 14: Restoration of Figure 13 requiring 47 seconds.

valued images practical on a serial processor.

Algorithm A produces piecewise smooth restorations quickly, instantiates the general theory, and demonstrates the advantage of the new theory, however three courses of future development are apparent in image processing. Algorithm A makes no use of explicit line fields with line-line interactions to favor smooth, closed boundaries[7]. Algorithm A uses the simplest clique potential even though more sophisticated clique potentials are available[7, 16, 8, 3]. Finally Algorithm A derives its speed just from the Bayesian likelihood, but the theory promises much faster convergence if a good enough generator can be found.

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-Kauffman, San Mateo, 1989. Reprinted from 1988 NIPS, Denver, CO.

[2] G. L. Bilbro, M. B. Steer, R. J. Trew, C.-R. Chang, and S. G. Skaggs. Extraction of the parameters of equivalent circuits of microwave transistors using tree annealing. *I.E.E.E. Trans. Microwave Theory and Techniques*, Nov. 1990. To appear.

[3] Griff L. Bilbro and Wesley E. Snyder. Mean field annealing: An application to image noise removal. *Journal of Neural Network Computing*, 1990. 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*. To appear.

[5] Romeo F. and A. Sangiovanni-Vencentelli. Probabalistic hill climbing algorithms: Properties and applications. In *Chapel Hill Conf. on VLSI*, Chapel Hill, NC, 1985. Computer Science Press.

[6] W. Feller. *An Introduction to Probability Theory and its Applications*, volume 1. Wiley, New York, 1950.

[7] D. Geman and S. Geman. Stochastic relaxation, Gibbs Distributions, and the Bayesian restoration of images. *IEEE Transactions on PAMI*, PAMI-6(6):721–741, November 1984.

[8] H. P. 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*, pages 1901–1912, December 1989.

[9] S. Hustin and A. Sangiovanni-Vicentelli. Tim, a new standard cell placement program based on the simulated annealing algorithm. In *International Workshop on Placement and Routing.* Microelectronics Center of North Carolina, 1988.

[10] J. Marroquin, S. Mitter, and T. Poggio. Probabilistic solution of ill-posed problems in computational vision. *Journal of American Statistical Association*, 82(397):76–89, March 1987.

[11] M. W. Roth. Survey of neural network technology for automatic target recognition. *I.E.E.E. Trans. Neural Networks*, 1(1):28–41, March 1990.

[12] W. E. Snyder, G. L. Bilbro, A. Logenthiran, and S. A. Rajala. Optimal thresholding a new approach. *Pattern Recognition Letters*, To appear.

[13] H. Szu and R. Hartley. Fast simulated annealing. *Physics Letters A*, 122(3,4):157–162, 1987.

[14] P.J.M Van Laarhoven and E. H. Aarts. *Simulated Annealing: Theory and Applications.* Reidel, Dordrecht, Holland, 1987.

[15] D. Vanderbilt and S. G. Louie. A monte carlo simulated annealing approach to optimization over continuous variables. *J. Comput. Phys.*, 36:259–271, 1984.

[16] G. Wolberg and T. Pavlidis. Restoration of binary images using stochastic relaxation with annealing. *Pattern Recognition Letters*, 3(6):375–388, December 1985.