shapes evenness implies flatness. To this end, we have introduced two methods. First, it may be possible to decompose vector configurations, as we have done for y-directed strips. Second, it may be possible to prove that an even set is arbitrarily extendable, as we have done for rectangular segments by using the concept of regularity.

Although decompositions, regularity and extensibility are useful tools, a full characterization of the shapes for which evenness implies flatness, remains to be established. For example, as to regularity, we do not know whether regularity is a necessary condition for the existence of single point extensions. Nor do we know how regularity is related to topological properties such as digital convexity or connectedness.

ACKNOWLEDGMENT

The author wishes to thank J. Van Campenhout, H. Peremans, and K. Audenaert. Their detailed comments helped to improve the readability of this correspondence.

REFERENCES

[1] S. H. Y. Hung, "On the straightness of digital arcs," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, pp. 203–215, 1985.
[2] C. E. Kim, "On cellular straight line segments," *Comput. Vision Graphics Image Processing*, vol. 18, pp. 369–381, 1982.
[3] ——, "Three-dimensional digital planes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 639–645, 1984.
[4] A. Rosenfeld, "Digital straight line segments," *IEEE Trans. Comput.*, vol. C-23, pp. 1264–1269, 1974.
[5] J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimensions I.* Berlin: Springer, 1970.
[6] P. Veelaert, "On the flatness of digital hyperplanes," *J. Math. Imaging and Vision*, vol. 3, pp. 205–221, 1993.
[7] ——, "Geometrical properties of three-dimensional digital planes," Tech. Rep. DG92-03, Univ. of Ghent, Electronics Lab., 1992.
[8] L. Dorst and A. smeulders, "Discrete representation of straight lines," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 450–462, 1984.
[9] M. E. Dyer, "Linear time algorithms for two- and three-variable linear programs," *SIAM J. Comput.*, vol. 13, pp. 31–45, 1984.
[10] C. E. Kim and I. Stojmenović, "On the recognition of digital planes in three-dimensional space," *Pattern Recognit. Lett.*, vol. 12, pp. 665–669, 1991.
[11] I. Stojmenović and R. Tošić, Digitization schemes and the recognition of digital straight lines, hyperplanes, and flats in arbitrary dimensions," in *Vision Geometry* (Comtemporary Mathematics Series Vol. 119), R. A. Melter, A. Rosenfeld, and P. Bhattacharya, Eds. New York: American Mathematical Society, 1991, pp. 197–212.

Simulated Annealing: A Proof of Convergence

Vincent Granville, Mirko Křivánek, and Jean-Paul Rasson

***Abstract*—We prove the convergence of the simulated annealing procedure when the decision to change the current configuration is blind of the cost of the new configuration. In case of filtering binary images, the proof easily generalizes to other procedures, including that of Metropolis. We show that a function Q associated with the algorithm must be chosen as large as possible to provide a fast rate of convergence. The worst case (Q constant) is associated with the "blind" algorithm. On the other hand, an appropriate Q taking sufficiently high values yields a better rate of convergence than that of Metropolis procedure.**

I. INTRODUCTION

There are different ways for simulating stochastic processes on a 2-dimensional grid [11]. The first attempts closely related to image processing are provided by [3] and [6]. However, the basic ideas had already appeared before in a more general context [13]. When a parameter called *temperature* is introduced an progressively decreased, then, beginning with an initial image (generally taken to be the observed noisy image) it is possible to obtain in an infinite number of steps a configuration of maximum a-posteriori probability which is called *MAP estimate* of the (unknown) true scene. In other words the cooling schedule allows restoration of images while the schedule with a fixed temperature is intended for image simulation. In most applications, the underlying model is a *Markov random field (MRF)* [5]. We do not give any details about MRF since these stochastic processes are now very popular and furthermore a description is not required for the understanding of the paper. The algorithms converging to the MAP estimate with a cooling schedule are generally called *simulated annealing*.

Many proofs of convergence of annealing schedules have already appeared in the literature [2], [4]–[6], [12], [14], [15]. The most popular algorithms are probably those based on the Gibbs Sampler [6] or on the Metropolis procedure [11]. Gidas [7] was one of the first to prove the convergence, but his proof contains many mistakes. Here, we develop a simple version of this algorithm and the major contribution is an original detailed proof of convergence. With respect to the terminology used in [15], the proposed algorithm is *homogeneous*: it is described by a sequence of homogeneous Markov chains; each Markov chain is generated at a fixed value of the temperature and the temperature is decreased in between subsequent Markov chains. However, the transition probabilities used here are rather different than usual: the decision to change the current configuration is based on the probability of that configuration, and not on the probability of the modified configuration as usual. Therefore, the fact that the algorithm will converge is much less intuitive and the proof requires some extra computations based on a lemma of Perron and Frobenius [16] as well as the analysis of the eigenvalues of a stochastic matrix. As a particular consequence, although convergence to a global maximum is guaranteed, the configurations which provide this maximum do not appear uniformly in limiting case.

Manuscript received March 11, 1992; revised August 8, 1993. Recommended for acceptance by Associate Editor A. Blake.

V. Granville and J.-P. Rasson are with the Département de Mathématique, F.U.N.D.P., Rempart de la Vierge 8, B-5000 Namur, Belgium.

M. Křivánek is with the Department of Computer Science, Charles University, CS-118 00 Praha 1, Czechoslovakia.

IEEE Log Number 9214427.

Although the major contribution of the paper is theoretical, we give a numerical application which will assess the theory developed here. In particular, it is shown how to proceed for improving the rate of convergence. We hope that the reader will gain an in-depth understanding of the logarithmic nature of annealing schedules.

II. Algorithm

From now on, we shall assume that x is a configuration of probability π_x of a particular random vector X. Typically, $\pi_x = P(X = x|Z = z)$ is the probability of the true scene conditionally to the (noisy) observed image. We have at our disposal an initial configuration $x(0)$ e.g., $x(0) = z$ and we would like after successive iterations to find a configuration which is as close as possible to $\arg\max_x \pi_x$. The maximum is attained by the MAP estimate of the true scene. The problem is that the number of configurations, say N, is very large e.g for a binary image, N is equal to 2^n where n is the number of pixels. Moreover the (Gibbs) probability π_x contains in general a constant factor which cannot be computed for practical reasons. Thus, it is impossible to use brute force and the simulated annealing algorithm should require the computation of only ratios of such probabilities. In fact our algorithm requires the computation of the ratios $\pi_{x(0)}/\pi_{x(t-1)}, t = 1, 2, \cdots$ instead of $\pi_{x(t)}/\pi_{x(t-1)}$ as usual, where t denotes the number of the current iteration and $x(t)$ is the current (updated) configuration. Since

$$\frac{\pi_{x(0)}}{\pi_{x(t)}} = \frac{\pi_{x(t-1)}}{\pi_{x(t)}} \frac{\pi_{x(0)}}{\pi_{x(t-1)}}, \tag{1}$$

the passage from iteration t to iteration $t+1$ requires the same basic computations in both cases. It can be argued that (1) will result in numerical inaccuracy if many successive applications are to be used. Fortunately, $\pi_{x(t-1)}/\pi_{x(t)} \to 1$ and in most applications, the numerator and the denominator considerably simplify. Moreover, it is always possible to take the logarithm of both sides of (1) in order to get more robust computations.

In this section, we suppose that each configuration x corresponds to a binary image containing n pixels labelled $1, \cdots, n$. At time (or step) $t, t = 1, 2, \cdots$ we build a new configuration $x(t)$ from the last one just by choosing at random one pixel i_t and possibly updating its assignment. Thus at each time t we have a dichotomic rule: we either make no change with the probability $1-\xi_t$ or we either change the assignment of one pixel chosen at random with the probability ξ_t. It is assumed that there is only one way to change the assignment of a pixel. This is the case if the algorithm is used to filter a binary image or (with a slight adaptation) to create a MAP segmentation on the observed image. Due to the dichotomic nature of the assignment rule, the algorithm is necessarily intended for a "binary" processing of the image. Therefore, it can be used for computing an optimal segmentation even if the image is not binary [10]. However, in case of a filtering version of the algorithm (like in this paper), the image need to be binary. The irreducibility condition requiring the finite number of steps and non-zero probability to pass from one configuration to any other is needed for the convergence. The algorithm is summarized as follows:

Algorithm

1) Begin with an initial configuration $x(0)$ and $t = 0$.
2) At step $t, t = 1, 2 \cdots$ build the configuration $x(t)$ as follows:
   2.1 Choose one pixel i_t at random
   2.2 Either keep the last configuration unchanged with the probability $1 - \xi_t$ Either change the assignment of pixel i_t with the probability ξ_t

Here, $\xi_t = (q\pi_{x(0)}/\pi_{x(t-1)})^{\phi(t)}$ where $q = \inf_x \pi_x / \sup_x \pi_x$ and $\phi(t)$ is a slowly increasing function. The expression $1/\phi(t)$ is called the *temperature* at time t. Notice that in case when $\pi_{x(t-1)}$ is large then the probability of the last configuration $x(t-1)$ is high and it is unlikely that this configuration will be changed in the current step. The converse is also true. In practice, raster scans of the image have been preferred to random selection of sites.

III. Proof of Convergence

Theorem 3.1 Let $\phi(t) = \lfloor \log_\mu t \rfloor$, where $\mu > 1$ is a sufficiently large integer. Then the algorithm converges almost surely to a global maximum of π_x.

Proof: We shall proceed in several steps. Let us consider the Markov chain constituted by nN states where the state (x, i) means that we are in the configuration x and that we intend to change the affectation (also called assignment) of pixel i. The set of all possible configurations will be denoted by Ω. Let $M(x, i; x', i'; t)$ be the transition probability from the state (x, i) at time $t-1$ to the state (x', i') at time $t (t \geq 1)$, and $M(t)$ be the $nN \times nN$ transition matrix at time t, containing the elements $M(x, i; x', i'; t)$. This Markov chain is aperiodic and irreducible but nonstationary. We also know that for (x, i) fixed there exists only one configuration $x'_{i,x} \neq x$ such that $M(x, i; x'_{i,x}, i'; t) \neq 0$. In fact, for all i, i' we have:

$$M(x, i; x', i'; t) = \begin{cases} \frac{1}{n}\{1 - (q\pi_{x(0)}/\pi_x)^{\phi(t)}\}, & \text{if } x' = x, \\ \frac{1}{n}(q\pi_{x(0)}/\pi_x)^{\phi(t)}, & \text{if } x' = x'_{i,x}, \\ 0, & \text{otherwise}. \end{cases}$$

Let $\tau_x \in \mathbb{R}_0^+$ $(x \in \Omega)$, $\sigma = \sum_{x\in\Omega} \tau_x \pi_x^{\phi(t)}$, and $p(t) = \frac{1}{n\sigma}\left(\tau_x \pi_x^{\phi(t)}, \cdots, \tau_x \pi_x^{\phi(t)}\right)_{x\in\Omega}$. It is understood that $p(t)$ is a stochastic vector with nN components, containing at most N distinct components (each component $\tau_x \pi_x^{\phi(t)}/n\sigma, x \in \Omega$ is repeated n times consecutively). Suppose that the parameters τ_x, $x \in \Omega$ satisfy the linear homogeneous system

$$\sum_{x\in C(x'), x\neq x'} \lambda_{x'x}\tau_x = n\tau_{x'}, \forall x' \in \Omega, \tag{2}$$

where $x \in C(x') \Leftrightarrow$ it is possible (only by changing the affectation of one pixel) to pass in one step from the configuration x to the configuration x' and $\lambda_{x'x}$ is the number of pixels for which we can pass in one step from the configuration x to the configuration x' by changing their affectation.

In the context of image filtering, $\lambda_{x'x} \in \{0, 1\}$ and therefore $M(t)$ is doubly stochastic and $\tau_x = 1$. In a previous version of the paper (see also [10]) the algorithm was more specifically designed for image segmentation, allowing $\lambda_{x'x}$ to take a value greater than 1. In any case it is easy to verify that $p(t)M(t) = p(t)$. The system (2) can be rewritten in the form $\Lambda\tau = 0$ where $\Lambda = (\lambda_{x'x})$, with $\lambda_{x'x} = -n$ if $x' = x$. But the sum of the components of any column of Λ is equal to zero and thus $|\Lambda| = 0$. Therefore the system (2) has an infinite number of solutions. If we put $\mathcal{A} := I + \frac{1}{n}\Lambda$, then $\mathcal{A}^T$ is a stochastic matrix which is irreducible and thereby $\rho(\mathcal{A}) \leq 1$, ($\rho(\mathcal{A})$ stands for the spectral radius of $\mathcal{A}$), and one of the eigenvalues is equal to 1. This last eigenvalue is associated to the eigenvector $\tau, \tau = (\tau_x)_{x\in\Omega}$. Thus, by the theorem of Perron and Frobenius we get $\tau > 0$, and the existence of a strictly positive solution to the system (2) is warranted.

Provided that $t \in [\mu^{k-1}, \mu^k[$ then $M(t)(p(t)$ resp.) is a constant matrix (vector resp.) and will be denoted by M_k (p_k resp.). We have established the fact that

$$p_k = p_k M_k, \forall k \in \mathcal{N}_0. \tag{3}$$

Let $\tilde{p}(t)$ be the probability vector at time t associated to the Markov chain described above. Let $\tilde{p}(0)$ be the initial probability vector. We obtain immediately that $\tilde{p}(t) = \tilde{p}(0)M(1)M(2)\cdots M(t)$. We would like to study the asymptotic behaviour of $\tilde{p}(t), t \to \infty$, but we have $\lim_{t-\infty} \tilde{p}(t) = \lim_{s-\infty} \tilde{p}(0) M_1^{\mu^1-\mu^0} M_2^{\mu^2-\mu^1} \cdots M_{s+1}^{\mu^{s+1}-\mu^s}$. Therefore, we shall concentrate on the proper member of the last identity.

If $M_{s+1}^{\mu^{s+1}-\mu^s}$ tends to be a matrix with identical rows, so it is for $M_1^{\mu^1-\mu^0} M_2^{\mu^2-\mu^1} \cdots M_{s+1}^{\mu^{s+1}-\mu^s}$ which is the product of a stochastic matrix $M_1^{\mu^1-\mu^0} M_2^{\mu^2-\mu^1} \cdots M_s^{\mu^s-\mu^{s-1}}$ with a matrix $M_{s+1}^{\mu^{s+1}-\mu^s}$ with identical rows in limiting case. As $\tilde{p}(0)$ is a probability vector, it follows that $\tilde{p}(t)$ converges when $t \to \infty$ to a row of the matrix $\lim_{s-\infty} M_{s+1}^{\mu^{s+1}-\mu^s}$.

Let $\omega = \mu^s, \mathcal{M}_\omega := M(\mu^s) = M_{s+1}$ and $\mathcal{M}_\omega^* = \lim_{t-\infty} \mathcal{M}_\omega^t$. We have $M_{s+1}^{\mu^{s+1}-\mu^s} = (\mathcal{M}_\omega)^{\omega(\mu-1)} = ((\mathcal{M}_\omega)^\omega)^{\mu-1}$ and thus to conclude the proof it is sufficient to prove that

$$\lim_{\omega-\infty} (\mathcal{M}_\omega)^\omega = \lim_{\omega-\infty} \mathcal{M}_\omega^*, \tag{4}$$

because $\mathcal{M}_\omega^*$ is a matrix with identical rows p_{s+1} (by virtue of (3)) and $p_{s+1} \to p$ as $s \to \infty$, where p is a vector such that only components corresponding to a global maximum of π_x are distinct from zero. We stress on the fact that the limiting density $p_{s+1}(s \to \infty)$ is not necessarily uniform on its support; this is a major difference with classical simulated annealing algorithms.

Now it remains to prove (4). In Lemma 3.2, we shall prove the convergence for a matrix $\mathcal{M}_\omega$ which is different and simpler than that described previously. However this new matrix is still worse (in term of rate of convergence, or in other words, in term of spectral radius) than the actual matrix. Therefore, if the convergence is attained with this "pseudo-matrix" (also noted $\mathcal{M}_\omega$), then the convergence is also garanteed for the true matrix. This pseudo-matrix is defined as the matrix which (after a suitable permutation of the states) corresponds to the irrealistic case when $n = 1$ and N is arbitrary. Therefore,

$$\mathcal{M}_\omega = \begin{pmatrix} 1-\omega^{-\alpha_1} & \omega^{-\alpha_1} & 0 & \cdots & 0 \\ 0 & 1-\omega^{-\alpha_2} & \omega^{-\alpha_2} & \cdots & 0 \\ 0 & 0 & 1-\omega^{-\alpha_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \omega^{-\alpha_{N-1}} \\ \omega^{-\alpha_N} & 0 & 0 & \cdots & 1-\omega^{-\alpha_N} \end{pmatrix} \tag{5}$$

and subsequently

$$\mathcal{M}_\omega^* = \frac{1}{\omega^{\alpha_1} + \cdots + \omega^{\alpha_N}} \begin{pmatrix} \omega^{\alpha_1} & \omega^{\alpha_2} & \cdots & \omega^{\alpha_N} \\ \vdots & \vdots & & \vdots \\ \omega^{\alpha_1} & \omega^{\alpha_2} & \cdots & \omega^{\alpha_N} \end{pmatrix},$$
$$\alpha_x = -\log_\mu(q\pi_{x(0)}/\pi_x) > 0. \tag{6}$$

□

(for ease of notation, the N configurations of Ω have been labelled $1, \cdots, N$; we use indistinctly the notation α_j or α_x if the configuration x has label j)

Lemma 3.2: Let $\mathcal{M}_\omega$ and $\mathcal{M}_\omega^*$ be matrices defined by (5) and (6). For μ sufficiently large it holds that $\lim_{\omega-\infty}\{\mathcal{M}_\omega\}^\omega = \lim_{\omega-\infty} \mathcal{M}_\omega^*$.

Proof: Let ρ_ω be the spectral radius of $\mathcal{M}_\omega - \mathcal{M}_\omega^*$. Then $\rho_\omega < 1$ (since $\lim_{t-\infty}(\mathcal{M}_\omega - \mathcal{M}_\omega^*)^t = 0$) and $\rho_\omega \sim \alpha - \gamma_1\omega^{-\beta_1} + \gamma_2\omega^{-\beta_2} + \cdots$ $(\omega \to \infty)$ where $0 < \beta_1 < \beta_2 < \cdots, \gamma_1, \gamma_2, \cdots \in \mathcal{R}_0$, $\gamma_1 > 0$ and $0 \le \alpha \le 1$. Suppose that $\alpha = 1$ (the other case is even easier). The characteristic equation can be rewritten asymptotically as

$$\lambda^N + (\delta_1 + \varepsilon_1\omega^{-\nu_1})\lambda^{N-1} + (\delta_2 + \varepsilon_2\omega^{-\nu_2})\lambda^{N-2} + \cdots + (\delta_N + \varepsilon_N\omega^{-\nu_N}) = 0. \tag{7}$$

where $\delta_1, \cdots, \delta_N; \varepsilon_1, \cdots, \varepsilon_N \in \mathcal{R}$, λ is any eigenvalue of $\mathcal{M}_\omega - \mathcal{M}_\omega^*$ and $\nu_1, \ldots, \nu_N > 0$. Let $\lambda^* = \rho\exp(\iota\varphi), \iota^2 = -1$, an eigenvalue of maximum modulus. Thus $\rho = \rho_\omega \sim 1 - \gamma_1\omega^{-\beta_1}$. If $Ne_N + \delta_1(N-1)e_{N-1} + \delta_2(N-2)e_{N-2} + \cdots + \delta_{N-1}e_1 \neq 0$ where $e_k := \exp(\iota k\varphi)$ and if $\sum_j \varepsilon_j \neq 0$ where $\sum_j$ is a sum taken over all indices j such that ν_j is minimum, then the asymptotic relation (7) holds only if $\beta_1 = \inf\{\nu_j; \varepsilon_j \neq 0\}$, else further asymptotic expansion is required under the entirely analogous proof. Taking μ large enough we take $\sup_j \alpha_j$ small enough and $\inf\{\nu_j; \varepsilon_j \neq 0\} < 1$. This is easily observed by developing the determinant $|(\mathcal{M}_\omega - \mathcal{M}_\omega^*) - \lambda I|$. Therefore $\rho_\omega \sim 1 - \gamma_1\omega^{-\beta_1}$ with $0 < \beta_1 < 1$ and subsequently $\lim_{\omega-\infty}(\rho_\omega)^\omega = \lim_{\omega-\infty}(\rho_\omega)^{\omega/N} = 0$. Without loss of generality, let ω be a multiple of N. Let R be any accumulation point of the sequence $(\mathcal{M}_\omega - \mathcal{M}_\omega^*)^{\omega/N}$ (N is fixed). The spectral radius of R is zero. Thus by Caley-Hamilton $R^N = 0$ and further $(\mathcal{M}_\omega - \mathcal{M}_\omega^*)^\omega \to 0$. But $\forall t \in \mathcal{N}_0$, $(\mathcal{M}_\omega - \mathcal{M}_\omega^*)^t = (\mathcal{M}_\omega)^t - \mathcal{M}_\omega^*$. In particular we have $(\mathcal{M}_\omega)^\omega \sim \mathcal{M}_\omega^*$.

IV. Further Details About q, μ and the Convergence

In practice q is not known. Geman and Geman [6] have to face a similar drawback in their Gibbs Sampler: instead of working with the ratio $q = \inf_x \pi_x / \sup_x \pi_x$ they use the quantity $\Delta = \sup_x \pi_x - \inf_x \pi_x$. In both cases, these quantities are just used for proving the convergence. In practice, the convergence may result even if q (resp. Δ) is chosen much greater (see section 5.1). Let us remark that the cooling schedule (that is, the *temperature*) is of order $n\Delta/\log t$ for the Gibbs Sampler. In our algorithm, it is of order $\log\mu/\log t$, where the constant μ has not been investigated in details. However, according to Lemma 3.2, there must be a constant κ $(0 < \kappa < 1)$, presumably depending on n, such that $\forall x, \alpha_x \le \kappa$. Therefore, we can take κ as follows:

$$\kappa = \sup_x \alpha_x = \sup_x \left\{ -\log_\mu \left(\frac{q\pi_{x(0)}}{\pi_x} \right) \right\} = \frac{\log(\sup_x \pi_x / q\pi_{x(0)})}{\log\mu}$$

and hence,

$$\mu = \left(\frac{\sup_x \pi_x}{q\pi_{x(0)}} \right)^{1/\kappa}, \qquad 0 < \kappa < 1. \tag{8}$$

Now it can be deduced from Lemma 3.2 that the sequence $(\pi_{x(t)})_t$ converges to $\sup_x \pi_x$ at least as fast as the sequence $(1 - \gamma_1 t^{-\beta_1})^t$ converges to 0, for some $\gamma_1 > 0$ and $0 < \beta_1 < 1$ depending on μ, n and N. If more than one configuration maximize π_x, then the sequence $(\pi_{x(t)})_t$ will converge to one of these maximizers, but the configurations with the highest τ_x have more chance to be reached. Let Ω_0 be the set of optimal configurations (those maximizing π_x). Then each $x_0 \in \Omega_0$ has a probability $\pi_{x_0}^* = \tau_{x_0} / \sum_{x \in \Omega_0} \tau_x$ to appear in limiting case (when $t \to \infty$). In case of image filtering, since $\tau_x = 1 (\forall x)$, the limiting distribution π^* is uniform as usual.

A possible generalisation of our algorithm consists in letting q depend on t. The theory also holds provided 1) $0 < q_t\pi_{x(0)}/\pi_{x(t-1)} < 1$, 2) $\exists \kappa_1, \kappa_2$, $0 < \kappa_1 \le \kappa_2 \dagger \forall t, \kappa_1 \le q_t \le \kappa_2$, 3) $\forall t, t' > 1, \lfloor \log t \rfloor = \lfloor \log t' \rfloor \Rightarrow q_t = q_{t'}$. For instance, q_t may practically be

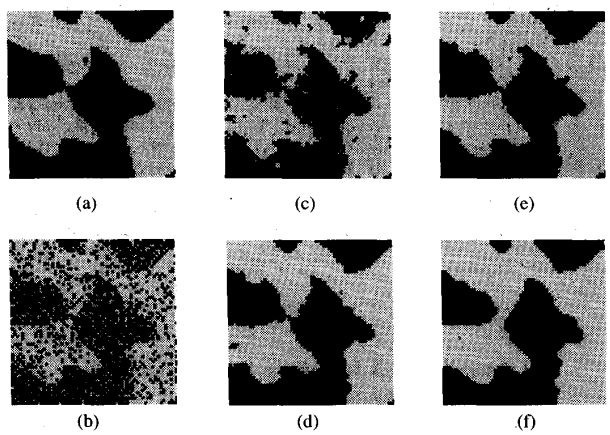

Fig. 1. True scene (a), observed image (b), and filtered image (c), (d), (e) using $10 \times n$ iterations of the annealing procedure, resp. with $Q = Q_1$, $Q = Q_2$, $Q = Q_4$.Comparison with the ICM procedure [9] is also provided (f).

TABLE I
NUMBER N_t OF WRONG PIXELS AND STATISTIC $U_t = \frac{1}{n}\log(\pi_{x(t-1)}/\pi_{x(0)})$ COMPUTED AFTER t ITERATIONS ($t = 0 \times n, 1 \times n, \cdots, 10 \times n$) OF THE ANNEALING PROCEDURE, RESP. WITH $Q = Q_1$, $Q = Q_2$,$Q = Q_4$

t/n	N_t				U_t			
	Q_1	Q_2	Q_4	ICM	Q_1	Q_2	Q_4	ICM
0	844	844	844	844	-	-	-	-
1	679	280	157	210	0.370	1.144	1.372	1.435
2	567	191	137	203	0.629	1.343	1.421	1.459
3	506	174	139	197	0.766	1.377	1.434	1.466
4	444	157	137	193	0.895	1.407	1.438	1.468
5	404	147	136	188	0.996	1.427	1.439	1.470
6	363	145	137	187	1.075	1.431	1.439	1.470
7	336	140	132	187	1.127	1.438	1.443	1.470
8	324	137	131	187	1.168	1.437	1.443	1.470
9	298	137	131	187	1.211	1.440	1.443	1.470
10	282	135	131	187	1.239	1.443	1.443	1.470
50	191	-	-	-	1.411	-	-	-

Comparison with the ICM procedure [9] is also provided.

a function of $\pi_{x(0)}, \cdots, \pi_{x(t-2)}$ but neither of $\pi_{x(t)}$ (in other case $M(t)$ is no more stochastic) nor $\pi_{x(t-1)}$ (in other case the relation $p(t)M(t) = p(t)$ is no more satisfied). A possible choice is

$$q_t = \frac{\min\{\pi_{x(0)}, \cdots, \pi_{x(t-2)}\}}{\max\{\pi_{x(0)}, \cdots, \pi_{x(t-2)}\}} \cdot \eta_t, \tag{9}$$

where η_t is a bounded sequence such that $0 < \lim_{t-\infty} \eta_t < 1$, with η_t sufficiently small to meet the requirement $0 < q_t \pi_{x(0)}/\pi_{x(t-1)} < 1$. This choice of q_t allows practical implementations which would otherwise be impossible since the true value of q is actually unknown.

V. NUMERICAL APPLICATION

In this section we apply the annealing procedure for filtering a binary noisy image. The true scene is modelled by the following simple h-nearest neighbors pairwise interaction MRF:

$$P(X = x) \propto \exp[-\beta n_{12}(x)], \beta > 0, \tag{10}$$

where $\beta = 1.5$ and $n_{12}(x)$ is the number of neighboring pairs with distinct colours in x (see Besag [1]) and $h = 8$. The noise process and the true image are assumed to be binary.

The mixture model of noise used here is described in [8], [9]. Under the assumption of an uniform noise, we have ([9], Theorem A):

$$P(X = x|Z = z) \propto \psi(\rho, x, z) \cdot P(X = x), \tag{11}$$

where z is the observed (noisy) image, $\rho (0 < \rho < 1)$ is a parameter (roughly speaking, $1 - \rho = \rho'$ is the proportion of noise in the observed image), $\psi(\rho, x, z) = (\rho + \rho'/c)^{\eta_{x,z}} \cdot (\rho'/c)^{n - \eta_{x,z}}$, $\eta_{x,z} = \#\{i, 1 \leq i \leq n \dagger x_i = z_i\}$, and $c = 2$ is the number of distinct colours in the image.

We are searching for the MAP estimate of the true scene. The MAP estimate is defined as the configuration x_* which maximizes the a-posteriori probability $\pi_x = P(X = x|Z = z)$. The initial configuration is $x(0) = z$ (this is nothing else than the conventional maximum likelihood classifier, see [9]). We have chosen a true scene x of $n = 64 \times 64$ pixels among the images at our disposal (taking $\phi(t) = 0$ in our simulated annealing algorithm would have produced a simulated true scene). The observed image z is a mixture of an uniform noise process and the true scene, with $\rho' = 0.4$. The filtered images (obtained after $10 \times n$ iterations, using raster scans of the image) are shown in Fig. 1. They were obtained with a parameter q depending on t and described in Section V-A. Numerical results are reported in Table I.

A. Practical Implementation of the Annealing Procedure

We generalize the theory developed so far. The proposed generalisation is valid only in the framework of image filtering; other generalisations should be investigated e.g. for image segmentation. Firstly, we release the assumption that the pixel i_t must be chosen at random at step t. To the contrary, it is now crucial (from a theoretical viewpoint) that $i_t = \varphi(i_{t-1})$ where φ is an appropriate one-to-one function. For instance $\varphi(i_t) = i_{t-1} + 1$ modulo (n) is the successor of i_{t-1} in a raster scan of the image. Hence, $M(x, i; x', i'; t) = 0$ if $(i, i') \neq (i_{t-1}, i_t)$ and assuming $(i, i') = (i_{t-1}, i_t)$, we have

$$\begin{cases} M(x, i; x, i'; t) & = & 1 - \{q_t \cdot \pi_{x(0)}/\pi_x\}^{\phi(t)} \\ M(x, i; x'_{i,x}, i'; t) & = & \{q_t \cdot \pi_{x(0)}/\pi_x\}^{\phi(t)} \\ M(x, i; x', i'; t) & = & 0 \text{ if } x \neq x' \neq x'_{i,x} \end{cases}$$

and

$$\xi_t = \left\{q_t \cdot \frac{\pi_{x(0)}}{\pi_{x(t-1)}}\right\}^{\phi(t)}, \qquad q_t = Q(\pi_{x(t-1)}, \pi_{x^*(t)}), \tag{12}$$

where $x^*(t)$ is the configuration that would be reached at step t if a change in the current configuration $x(t-1)$ was allowed and Q is a constant function taking the fixed value q. The generalisation consists of course in taking Q arbitrary. This approach yields a rich class of simulated annealing algorithms. Let us remark that thanks to the choice of i_t we can allow $q_t = Q$ to depend both on $x(t-1)$ and (implicitly) on $x(t)$: provided Q is a symmetric function, it can be seen that the condition $p(t)M(t) = p(t)$ still holds, with $\tau_x = 1$. The proof of convergence remains almost unchanged. In order to get a fast rate of convergence, $Q(\pi_{x(t-1)}, \pi_{x^*(t)})$ must be chosen as large as possible, with the restriction that $\forall i = 1, \cdots, n, \forall x \in \Omega$, $0 < Q(\pi_x, \pi_{x'_{i,x}}) \cdot \frac{\pi_{x(0)}}{\pi_x} < 1$. By maximizing Q, we also minimize $\sup_x \alpha_x = \sup_x\{-\log_\mu(Q\pi_{x(0)}/\pi_x)\}$ and hence β_1. Remember that the rate of convergence of the annealing algorithm is that of the sequence $(1 - \gamma_1 t^{-\beta_1})^t$, $\gamma_1 > 0$. Taking for instance $Q_1(\pi_x, \pi_{x'}) = q' \cdot \sqrt{\pi_x \pi_{x'}}/\pi_{x(0)}$. ($q'$ fixed), we get

$$\xi_t = \left(q' \cdot \sqrt{\frac{\pi_{x^*(t)}}{\pi_{x(t-1)}}}\right)^{\phi(t)}, \qquad q' \leq \inf_{i,x} \sqrt{\frac{\pi_x}{\pi_{x'_{i,x}}}} \tag{13}$$

and we can compute ξ_t directly (there is no need to use the numerically ill-conditioned formula (1)). Moreover the decision to modify the current configuration $x(t-1)$ is no more blind of the cost of the new configuration $x^*(t)$. Hence we get a reasonable practical rate of convergence. Furthermore, with the MRF model given by formula (10) we easily establish that $q' \leq \exp(-\beta h/2)$, where h is the MRF neighboring size (here $h = 8$ and $\beta = 1.5$). In the next examples, we have taken $q' = \exp(-\beta h/2)$. Finally, using the noise model (11), we get

$$\log \xi_t = \frac{\phi(t)}{2}\left\{2\log q' + (\eta_{x^*(t),z} - \eta_{x(t-1),z})\log\left(1+\frac{c\rho}{\rho'}\right) - \beta(n_{12}[x^*(t)] - n_{12}[x(t-1)])\right\}. \quad (14)$$

Other choices tested in this paper are $Q_2(\pi_x, \pi_{x'}) = \pi^{-1}_{x(0)} \min(\pi_x, \sqrt{q'\pi_x, \pi_{x'}})$ and $Q_3(\pi_x, \pi_{x'}) = q$; Q_2 yields the Metropolis algorithm. But Q_2 is not symmetric. However, with an appropriate adaptation of $p(t)$, we still keep the relation $p(t)M(t) = p(t)$. Not surprisingly, Q_2 yields a better rate of convergence than Q_1: this is explained by the fact that in any case, $Q_2 > Q_1$. For similar reasons, Q_3 yields the worst procedure and can not be implemented. Eventually, working with $Q_4(\pi_x, \pi_{x'}) = \pi^{-1}_{x(0)} \min(\pi_x, \sqrt{q'}\ (\pi_x + \pi_{x'})/2) \geq Q_2(\pi_x, \pi_{x'})$, we get a better rate of convergence than that of the Metropolis algorithm. The nearly optimal annealing schedule used here correspond to $\phi(t) = \log_\mu(t)$, where $\log \mu$ is equal to 16 (resp. 8, 4) when Q_1 (resp. Q_2, Q_4) is used.

ACKNOWLEDGMENT

We would like to thank anonymous referees for their helpful comments and suggestions, leading to a significant improvement of the correspondence. They encouraged the authors to investigate in more details the properties of annealing procedures, resulting in a deeper understanding of the convergence mechanism.

REFERENCES

[1] J. Besag, "On the statistical analysis of dirty pictures," *JRSS B*, vol. 48, pp. 259–302, 1986.
[2] M. E. Johnson, Ed., , *Simulated Annealing and Optimization: Modern Algorithms With VLSI, Optimal Design and Missile Defense Applications*. New York: American Science Press, 1989.
[3] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 5, no. 1, pp. 25–39, 1983.
[4] S. B. Gelfand and S. K. Mitter, "Simulated annealing," in *Proc. Advanced school CISM*, Udine, Italy, 1987, pp. 1–51.
[5] D. Geman, "Random fields and inverse problems in imaging," *Lectures Notes in Mathematics*, vol. 1427, pp. 117–196, 1988.
[6] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI- 6, no. 6, 721–741, June 1984.
[7] B. Gidas, "Non stationary Markov chains and convergence of simulated annealing algorithms," *J. Statist. Phys.*, vol. 39, pp. 73–131, 1985.
[8] V. Granville and J. P. Rasson, "A new modelisation of noise in image remote sensing," *Statist. and Probab. Lett.* vol. 14, pp. 61–65, 1992.
[9] ——, "A Bayesian filter for a mixture model of noise in image remote sensing," *Comput. Statist. Data Anal.* vol. 15, pp. 297–308, 1993.
[10] V. Granville, J. P. Rasson, and F. Orban-Ferauge, "Un modèle Bayesien de segmentation d'images," in *Proc. 4th Journées Scientifiques du Réseau de Télédétection de L'UREF*, Montréal, Oct. 1991 (to appear).
[11] X. Guyon, "Champs stationnaires sur $\mathcal{Z}^2$: Modéles, statistique et simulations," Université Paris I., 1985.
[12] B. Hajek, *Cooling schedules for optimal annealing, Math. Oper. Res.* vol. 13, pp. 311–329, 1988.
[13] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods*. London: Methuen, 1964.
[14] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 8, 799–813, Aug. 1989.
[15] V. Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications*. Dordrecht, D. Reidel, 1987.
[16] R. S. Varga, *Matrix Iterative Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1962.

Openings Can Introduce Zero Crossings in Boundary Curvature

P. F. M. Nacken

***Abstract*—In a recent paper, Chen and Yan presented a result on zero crossings of boundary curvature under morphological openings. In this correspondence, it is shown, by means of a counter example, that this result is not correct. An analysis of the counter example shows where Chen and Yan go wrong and determines a class of images for which their theorem holds.**

I. INTRODUCTION

This correspondence reports a problem encountered in a paper of Chen and Yan [2]. The paper by Chen and Yan contains a theorem on the behaviour of zero crossings of the curvature of the boundary of an object which is transformed through a family of morphological openings [4]. In this paper, a counter example to this theorem is given. This counter example is analysed to show where Chen and Yan go wrong and to determine a class of objects for which their theorem holds.

Multiscale image analysis frequently involves the tracing of features of an image X as X is degraded by some kind of blurring process [1], [3], [5]. It is desirable that the number of features or the amount of structure in the feature set becomes smaller as X is blurred.

An example of such a multiscale image analysis is studied by Chen and Yan. These authors define a multiscale analysis of binary images (in $\mathbb{R}^2$) by considering the zero-crossings of the curvature of the boundary contour. The blurring operator is a morphological opening with a disk of radius r. In the rest of this correspondence, it is assumed that objects are simply connected, i.e., contain no holes.

Let ∂X denote the contour of a set X. $B(r)$ is a (closed) disc of radius r. $Z(\partial X)$ is the number of zero crossings of the curvature of the contour and $CN(X)$ is the number of connected components of a set X. Irregular points of a contour are those points where the curvature has (positive or negative) infinite value. In [2] the following result is formulated:

Manuscript received April 20, 1992; revised August 31, 1993. This work was supported by the Foundation for Computer Science in the Netherlands (SION) with financial support from the Netherlands Organisation for Scientific Research (NWO). This work was part of a project in which the TNO Institute for Human Factors, CWI and the University of Amsterdam cooperate. Recommended for acceptance by Editor-in-Chief A. K. Jain.

The author is with the TNO Institute for Human Factors IZF, Kampweg 5, 3769 DE Soesterberg, The Netherlands.

IEEE Log Number 9215850.