Simulated Annealing, Acceleration Techniques, and Image Restoration

Marc C. Robini, *Member, IEEE,* Thierry Rastello, and Isabelle E. Magnin, *Member, IEEE*

***Abstract*— Typically, the linear image restoration problem is an ill-conditioned, underdetermined inverse problem. Here, stabilization is achieved via the introduction of a first-order smoothness constraint which allows the preservation of edges and leads to the minimization of a nonconvex functional. In order to carry through this optimization task, we use stochastic relaxation with annealing. We prefer the Metropolis dynamics to the popular, but computationally much more expensive, Gibbs sampler. Still, Metropolis-type annealing algorithms are also widely reported to exhibit a low convergence rate. Their finite-time behavior is outlined and we investigate some inexpensive acceleration techniques that do not alter their theoretical convergence properties; namely, restriction of the state space to a locally bounded image space and increasing concave transform of the cost functional. Successful experiments about space-variant restoration of simulated synthetic aperture imaging data illustrate the performance of the resulting class of algorithms and show significant benefits in terms of convergence speed.**

***Index Terms*— Discontinuity recovery, ill-posed inverse problems, image restoration, metropolis dynamics, simulated annealing.**

I. Introduction

THE IMAGE restoration problem is to recover the original image from a blurred and noisy observation. It is widely documented [1]–[10] and usually addressed under assumptions of stationarity. In this paper, we consider a general linear image observation model (i.e., no distinction is made between the space-variant and space-invariant cases) together with a stationary noise process. We also assume that the PSF is known precisely although blur identification is quite often an ambitious challenge. Most of the time, the associated discrete image restoration problem is both ill-conditioned (this follows from the ill-posedness of the underlying continuous inverse problem) and underdetermined. Thus, the information provided by the data alone is generally not sufficient to determine a satisfying solution and it is then necessary to introduce some *a priori* assumption about the original image by imposing further constraints derived from an appropriate model. This approach stems from regularization theory [11], [12] and find some interpretation in a Bayesian framework [1], [3]. A popular and quite realistic image model is based on the common observation that most real scenes are locally smooth [5]; it leads to the extensively studied, so-called edge-preserving regularization [13] and results in either a convex or nonconvex optimization problem [1], [5], [9], [14]–[19]. In this last case which is considered here, one can employ deterministic (hence suboptimal) algorithms such as iterated conditional modes (ICM) [2], graduated nonconvexity (GNC) [19], mean-field annealing (MFA) [20], or ARTUR [13] and its recent generalization [21]. An alternative is to use stochastic relaxation with annealing, which can be shown to be asymptotically optimal when properly tuned [1], [22], [23]. Well-known types of stochastic algorithms are the Metropolis dynamics [24] (see [25] for a general setting) and the Gibbs sampler [1]. Yet, despite the continuous development of high performance computer equipments, annealing algorithms are widely reported to exhibit a very low convergence rate.

Our concern is to show that the convergence speed of Metropolis-type annealing algorithms can be significantly increased while staying in a rigorous theoretical framework. One should note here that this class of algorithms is quite rarely used in the image restoration field. Indeed, as a result of the amazing work of Geman and Geman [1], there is a strongly marked preference for the Gibbs sampler. However, for large discrete support PSF's and a typical 256-level dynamic range, the exact implementation of the Gibbs sampler becomes practically unfeasible. The computational load is reduced to a great extent by using the Metropolis dynamics, which may also be preferred because of its tremendous simplicity, but other practical problems can then be encountered.

We shall first consider the critical issue of finite-time convergence. It was proved in [26] that conventional logarithmic cooling schedules generally perform poorly as soon as one deals with a finite amount of available computing time and that exponential schedules are to be preferred. This mathematical justification for exponential schedules is of remarkable importance since it concerns most real applications of simulated annealing. We try to state this result in an accessible way and we clearly show that the best achievable convergence rate exponent is reached asymptotically for suitably adjusted exponential cooling schedules. Subsequently, two theoretically justified acceleration techniques are successively considered. The first one allows to overcome the specific difficulties associated with the Metropolis single-site updating dynamics by restricting the space of allowable images to certain subsets [27]–[29] which are particularly well adapted to the representation of locally smooth images. The second

Manuscript received December 2, 1997; revised January 15, 1999. This work was supported by Électricité de France under Contract #2L7297/01/EP832 and has been performed in accordance with the scientific trends of the National Research Group GDR ISIS of the Centre National de la Recherche Scientifique. The associate editor coordinating the review of this manusript and approving it for publication was Prof. Patrick L. Combettes.

The authors are with CREATIS, CNRS research unit (UMR 5515), INSA 502, 69621 Villeurbanne Cedex, France (e-mail: marc.robini@creatis.insa-lyon.fr).

Publisher Item Identifier S 1057-7149(99)07558-2.

technique constitutes the essential contribution of the paper: as demonstrated in [30], [31], an inexpensive speed-up can be obtained by applying any increasing concave transform to the cost functional to be minimized. Still, to our knowledge, the effectiveness of this method has not yet been tested experimentally. We also establish a new result that allows one to compare, at least theoretically, the relative performance of any two possible transforms.

The paper is organized as follows. The image restoration problem is defined and discussed in the next section. Metropolis-type annealing algorithms are presented in Section III together with the main convergence results, a particular emphasis being made on their finite-time behavior. Section IV is devoted to acceleration techniques and the peripheral issue of the selection of exponential cooling schedules is raised in Section V. Experimental results about restoration of simulated synthetic aperture imaging data appear in Section VI, followed by concluding remarks.

II. Edge-Preserving Image Restoration

Consider the linear image observation model expressed as

$$d_e(r,u) = \iint_D x^0(r',u')h(r,u;r',u')\,dr'\,du', \qquad (1)$$

$$d(r,u) = d_e(r,u) + \eta(r,u), \qquad (r,u) \in D' \qquad (2)$$

where $D \subset \mathbb{R}^2$ and $D' \subset \mathbb{R}^2$ are given domains, x^0 represents the original image to be recovered, d is the available experimental data and η is a stationary noise process; the integral equation (1) relating x^0 to the exact data d_e is a particular case of a Fredholm equation of the first kind and the kernel h is the two-dimensional (2-D) impulse response, or PSF, of the imaging system. Typically, the domain D is sampled on a regular grid S and the discrete image restoration problem is formulated as the estimation of the true image $x^0 = \{x_s^0; s \in S\}$ from data d according to the degradation model

$$d_s = (Hx^0)_s + \eta_s, \qquad s \in S' \qquad (3)$$

where $S' \subset S$ (see, e.g., [5]), H is a discrete blur operator representing the PSF and $(Hx)_s \doteq \Sigma_t\, h_{s,t}x_t$. The transformation from x^0 to d induces a severe loss of information. Since we are primarily concerned with the cases in which the support of the discrete kernels $h_{s,\cdot}$, $s \in S'$, extends over a large area, the linear system (3) is underdetermined. Moreover, it is well known that first kind Fredholm integral equations, and hence the initial continuous problem, are ill-posed in the sense of Hadamard [12]. As a direct consequence, the recovery of x^0 is an ill-conditioned inverse problem; immediate solutions such as the least squares one with minimum norm, or the generalized inverse if H^tH is singular, are not satisfactory because the noise component amplification is most often unacceptable.

In many image restoration methods, the above difficulties are overcome by including some *a priori* information about the true image x^0. The framework proposed here belongs to this category. Let $\Lambda \doteq \{0,\cdots,L-1\}$ be the set of gray levels, or intensity range, we define $\Lambda^{|S|}$ to be the (digital) total image space, where $|A|$ refers to the cardinal number of a finite set A. The solution minimizes a cost functional, or "energy function" U: $\Lambda^{|S|} \to \mathbb{R}$ defined by

$$U(x) = \lambda\Phi(x) + \sum_{s\in S'}((Hx)_s - d_s)^2 \qquad (4)$$

where $\lambda \in \mathbb{R}_+^*$ (i.e., $\mathbb{R}_+\backslash\{0\}$) is the usual "smoothing parameter" and the "regularization term" Φ should be beneficial for both the removal of blur and the recovery of discontinuities in the image. From a Bayesian point of view, $\hat{x}^0 = \arg\min_{x\in\Lambda^{|S|}} U(x)$ is the maximum *a posteriori* (MAP) estimate arising when η is a white Gaussian noise and the density for the prior distribution of x^0 is proportional to $\exp(-\lambda'\Phi(x))$. We restrict our attention to the case of an approximately piecewise constant original image and we adopt the first-order model

$$\Phi(x) = \sum_{\langle s,t\rangle} \phi(\mathcal{D}^{(1)}_{\langle s,t\rangle}(x)/\Delta). \qquad (5)$$

Here, $\Delta \in \mathbb{R}_+^*$ is a scale parameter and $\mathcal{D}^{(1)}_{\langle s,t\rangle}$ is a linear approximation to a first-order derivative in an horizontal, vertical, or diagonal direction (the summation is over pairs of pixels defined on the eight-nearest neighbor system). Note that, in a more general setting, one can consider linear combinations with higher-order derivatives [5], [9] which promote, for instance, the formation of piecewise planar and quadric areas.

The function ϕ: $\mathbb{R} \to \mathbb{R}$ in (5) is known as the "potential function" in the Bayesian framework; it is commonly taken to be even, increasing in $\mathbb{R}_+$ and such that $\phi(0) = 0$. Many ϕ functions that allow the preservation of edges have been proposed in the literature. Some authors advocate the use of convex functions [14]–[17] in order to ensure the well-posedness of the inverse problem, while some others prefer using nonconvex potentials [5], [9], [18], [19], [32] which yield to sharper edges but may introduce instability into the inversion process. Thus, it is not surprising that the conditions for the ϕ function to preserve discontinuities are somewhat contradictory. Stevenson *et al.* [17] argue that ϕ should be convex and that it should verify $\phi(v) < v^2$ for large values of v. In [5], Geman and Reynolds generalize the result of Blake and Zisserman [19] about implicit line processes and impose the condition $\lim_{v\to\infty}\phi(v) < +\infty$. Their work is further extended by Charbonnier *et al.* [13] who specify conditions that are quite similar to the ones proposed by Li [33]. In order to favor accurate reconstructions in the vicinity of discontinuities, we use the function

$$\phi(v) = \frac{|v|}{1+|v|} \qquad (6)$$

introduced in [5], which is strictly concave in $(0,+\infty)$. It follows from this last property that the functional U is most often nonconvex. Indeed, strictly speaking, the set $\Lambda^{|S|}$ is nonconvex because Λ is discrete and therefore we cannot define any convex optimality criterion on such a set. The formidable optimization problem involved in minimizing U is tackled with a dynamic Monte Carlo method, namely the Metropolis sampler with annealing, as described in the next three sections.

III. Metropolis-Type Annealing Algorithms

This section is primarily intended to provide a clear-cut justification for exponential cooling schedules. In addition, the presented material allows us to introduce some useful notations and to set the basis that will be used throughout the rest of the paper.

Consider a real-valued function U to be minimized on a finite set E called the configuration space (or state space) and let $q = \{q(x,y); x,y \in E\}$ be a symmetric and irreducible Markov kernel on E, which we will call the *communication kernel*. Clearly, if $q(x,x) = 0$ for all $x \in E$, then $\mathcal{N}(E) = \{\mathcal{N}_x(E);\ x \in E\}$, $\mathcal{N}_x(E) = \{y \in E \mid q(x,y) > 0\}$, is a neighborhood system on E. In practice, the communication kernel is usually defined by first selecting a neighborhood system on the configuration space and then imposing

$$q(x,y) = \begin{cases} |\mathcal{N}_x(E)|^{-1}, & \text{if } y \in \mathcal{N}_x(E), \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Define for any positive real parameter β the Markov kernel $p_{U,\beta}$ on E such that

$$p_{U,\beta}(x,y) = \begin{cases} q(x,y)\exp(-\beta(U(y)-U(x))^+), \\ \qquad \text{if } y \neq x, \\ 1 - \displaystyle\sum_{z \in E\setminus\{x\}} p_{U,\beta}(x,z), \\ \qquad \text{otherwise.} \end{cases} \tag{8}$$

where $a^+ \doteq \max\{a,0\}$. The unique equilibrium probability measure for the transition $p_{U,\beta}$ is the Gibbs distribution with energy U at temperature β^{-1}: $\pi_{U,\beta}(x) = \exp(-\beta U(x))/Z_{U,\beta}$, where $Z_{U,\beta} = \Sigma_{z\in E}\exp(-\beta U(z))$. As $\beta \to \infty$, one can check that $\pi_{U,\beta}$ tends to the uniform distribution Π_U on the set $E_{\min} \doteq \{x \in E \mid U(x) = \min_{z\in E} U(z)\}$ of global minima of U. Now consider a divergent increasing sequence $(\beta_n)_{n\in\mathbb{N}^*}$ of inverse temperatures called the cooling schedule. A Metropolis-type annealing algorithm associated with $(E,U,q,\nu_0,(\beta_n)_{n\in\mathbb{N}^*})$ is a discrete-time, nonstationary Markov chain $(X_n)_{n\in\mathbb{N}}$ with initial law of X_0 given by ν_0 and transitions $\mathrm{P}(X_n = y \mid X_{n-1} = x) = p_{U,\beta_n}(x,y)$, $x, y \in E$. The key idea is that, for sufficiently slowly increasing cooling schedules, the law of X_n should be close to π_{U,β_n} and we can expect that

$$\lim_{n\to\infty} \inf_{x\in E} \mathrm{P}(X_n \in E_{\min} \mid X_0 = x) = 1. \tag{9}$$

An early result of Geman and Geman [1] showed that (9) holds if $\Sigma_{n\in\mathbb{N}^*}\exp(-\beta_n R) = \infty$ with $R > 0$ large enough (though these authors considered the Gibbs sampler, the proof readily applies to the Metropolis dynamics), thus leading to logarithmic cooling schedules of the form $\beta_n = \beta_0 \ln(n+1)$, where $0 < \beta_0 \leq R^{-1}$. The best value for the constant R has been computed by Hajek [22] and a theorem by Chiang and Chow [23] completes this result by specifying when the annealing sequence $(X_n)_{n\in\mathbb{N}}$ has an asymptotic limit distribution that gives strictly positive mass to every configuration $x \in E_{\min}$. Also, a necessary and sufficient condition for strong ergodicity (i.e., $\lim_{n\to\infty} \mathrm{P}(X_n = y \mid X_m = x) = \Pi_U(y)\ \forall\ m \in \mathbb{N}$, $x \in E$) can be found in [34].

More crucial for practical use of simulated annealing is the convergence rate, as one always deals with finite-time horizon cooling schedules. Catoni [26] shows that the convergence measure

$$M_\alpha(n) \doteq \sup_{x\in E} \mathrm{P}(U(X_n) \geq \min_{z\in E} U(z) + \alpha \mid X_0 = x) \tag{10}$$

$\alpha \in \mathbb{R}_+^*$, cannot decrease faster than some limited power of n^{-1}. More precisely, for any triplet (E,U,q), there exists $k_1 \in \mathbb{R}_+^*$ such that

$$\inf_{\beta_1\leq\beta_2\leq\cdots\leq\beta_n} M_\alpha(n) \geq k_1 n^{-D_\alpha^{-1}(E,U,q)} \tag{11}$$

where the definition of the difficulty D_α of the energy landscape (E,U,q) at level $\alpha > 0$ is based upon the notion of cycle decomposition of the state space introduced in [35] (see Appendix A). The constant D_α is sharp, since it is possible to construct finite-time cooling schedules $(\beta_n^N)_{1\leq n\leq N}$ that are almost optimal in the sense that $M_\alpha(N) \leq k_2 N^{-D_\alpha^{-1}(E,U,q)}$ holds for some positive constant k_2 and α small enough [26]. This is technically achieved with piecewise logarithmic sequences that are closely linked to the energy landscape and hence extremely difficult to identify. A noticeable consequence is that conventional logarithmic schedules will not give an expression (10) decreasing as the optimal power of n^{-1} in the general case. Alternatively, though Hajek's result [22] proves that an exponential cooling schedule cannot generally grant (9), one can select

$$\beta_n^N = \beta_0 \exp(n\xi), \tag{12}$$

$$\xi = N^{-1}(\ln N - \ln(\alpha\beta_0)) \tag{13}$$

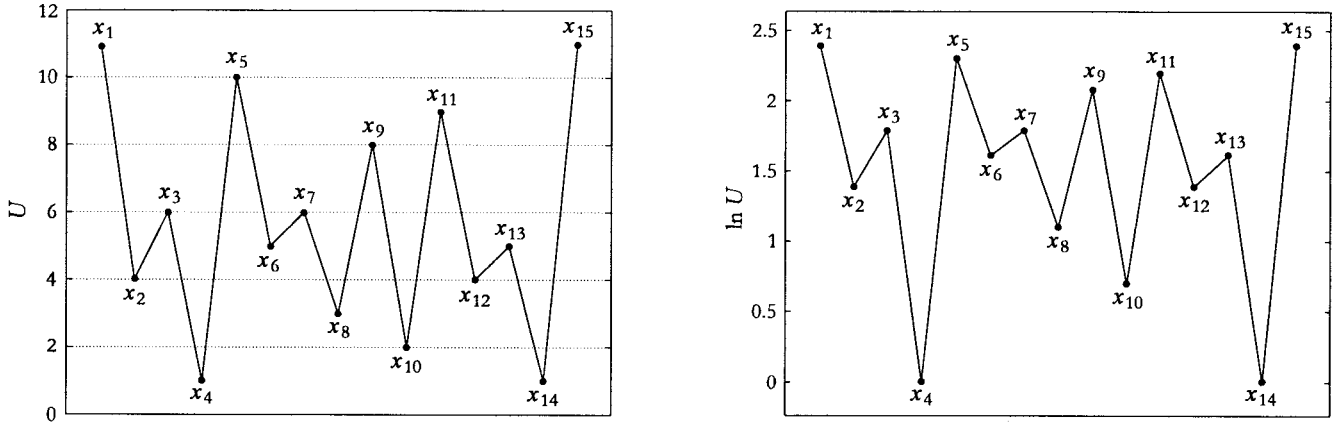
where $\beta_0 \in \mathbb{R}_+^*$ is independent of the horizon N, such that $M_\alpha(N)$ is of order $N^{-D_\alpha^{-1}(E,U,q)}$ as N increases and for α small enough. This particularly interesting result follows directly from the material presented in [26] (see Appendix B) which provides, to our knowledge, the first rigorous mathematical justification for the widely used exponential cooling schedules. Moreover, it turns out that these schedules are to be preferred to logarithmic ones as soon as one deals with a finite amount of computing time.

IV. Acceleration Techniques

In straightforward applications of the Metropolis dynamics to image processing $(E = \Lambda^{|S|})$, the communication kernel (7), which will be denoted by q_Λ, is defined by

$$\mathcal{N}_x(E) = \left\{ y \in \Lambda^{|S|} \mid\ |\{s \mid x_s \neq y_s\}| = 1 \right\}. \tag{14}$$

The corresponding annealing algorithms $(\Lambda^{|S|},\ U,\ q_\Lambda,\ \nu_0,\ (\beta_n^N)_{1\leq n\leq N})$ generally perform poorly and one typically introduces computational short cuts at the expense of a loss of the theoretical convergence properties. However, this can be elegantly avoided by means of the two rigorously justified acceleration techniques described below. The first one, initially motivated by the experimental success of the "union" algorithm from [5], consists in carrying out the minimization process on a restricted image space [27]–[29], say $\Omega_\delta^{|S|}$, through the use of a specific communication kernel q_δ (the corresponding material will be briefly summarized since we do



(a) (b)

Fig. 1. Increasing concave transform of the cost functionnal: (a) original energy landscape (E, U, q); (b) distorted energy landscape $(E, \ln U, q)$.

not make any contribution to this existing method). The second technique follows from an idea of Azencott [30], [31]: an inexpensive speed-up may be obtained by monotonic concave distortion of the energy function, i.e., using the composite of U and a real-valued function φ, increasing and strictly concave, rather than U itself. Taking both of these modifications into account leads to a class of annealing algorithms $(\Omega_\delta^{|S|}, \varphi \circ U, q_\delta, \nu_0, (\beta_0 \exp(n\xi))_{1 \leq n \leq N})$, which perform significantly better than $(\Lambda^{|S|}, U, q_\Lambda, \nu_0, (\beta_n^N)_{1 \leq n \leq N})$ for both logarithmic and exponential cooling schedules, as will be illustrated by experiments in Section VI.

A. Restricted Image Spaces

For the choice of the communication kernel q_Λ (7), (14), a new candidate is uniformly generated over the entire gray level range. Hence, especially for low temperatures, $\mathrm{P}(X_n = X_{n-1})$ is close to one and many iterations are required before a site has a sensible chance to undergo a large intensity change. As a result of this high rejection rate, image discontinuities are difficult to alter and "outlier" intensity values resulting from the early stages of the annealing process subsist quite frequently. In order to overcome these problems, the total image space $\Lambda^{|S|}$ can be restricted to a locally bounded image space $\Omega_\delta^{|S|}$ at level $\delta \in \mathbb{N}^*$ which consists of all the configurations $x \in \Lambda^{|S|}$ such that, $\forall\ s \in S$

$$\min_{t \in \mathcal{N}_s(S)} x_t - \delta \leq x_s \leq \max_{t \in \mathcal{N}_s(S)} x_t + \delta \tag{15}$$

where $\mathcal{N}(S)$ is a predefined neighborhood system on S [29]. Clearly, the use of a four- or eight-nearest neighbor system together with a reasonably small level δ (we shall take $L = 256$ and $\delta = 5$ in our experiments) permits the formation of piecewise smooth images and is thus in accordance with the prior information introduced in Section II. Another quite similar possibility is to make use of the class of subspaces defined in [27], [28], such that $\min_{t \in \mathcal{N}_s(S)} |x_s - x_t| \leq \delta$ $\forall\ s \in S$. Still, we prefer to resort to (15) because of the greater richness of the associated restricted image spaces.

Let us assume that δ is fixed and $X_{n-1} = x \in \Omega_\delta^{|S|}$. The Metropolis dynamics working on $\Omega_\delta^{|S|}$ is implemented in the following manner. The new candidate y at time n is generated from x by affecting to a randomly selected site s a random gray level value in the set $\omega_s^\delta(x) = \{l \in \Lambda \mid (x_{\setminus\{s\}}, l) \in \Omega_\delta^{|S|}\}$, where $x_{\setminus\{s\}}$ stands for x excluding site s. The computation of $\omega_s^\delta(x)$ is quite straightforward as soon as one is aware of the fact that a site s is not only bounded by its neighbors, but also provides bounds for them. However, it leads to a rather complex-looking mathematical description which is reproduced in Appendix C for convenience. The main point to retain is that the communication kernel $q_\delta = \{q_\delta(x, y);\ x, y \in \Omega_\delta^{|S|}\}$ (46) can be shown to be symmetric and irreducible [29, Lemma 2]. Therefore, the convergence results exposed in Section III hold for a locally bounded image space $\Omega_\delta^{|S|}$ at level $\delta \in \mathbb{N}^*$ together with q_δ.

B. Increasing Concave Transform of the Cost Functional

Recall from Section III that it is possible to select finite-time cooling schedules such that the convergence measure $M_\alpha(N)$ (10) is of order $N^{-D_\alpha^{-1}(E,U,q)}$, where the inverse difficulty $D_\alpha^{-1}(E, U, q)$ is the optimal annealing speed exponent. Hence, predictably enough, the more difficult the energy landscape, the lower the convergence rate. Let φ be a function that is strictly increasing in an interval covering at least the range of U. Clearly, from the definitions given in Appendix A, we have $\mathcal{C}(E, \varphi \circ U, q) = \mathcal{C}(E, U, q)$ and the minimization of U on the state space E can therefore be handled equally well by using either the composite of U and φ or U itself. At this point, the interesting thing is that the difficulty of $(E, \varphi \circ U, q)$ at level $\alpha \in \mathbb{R}_+^*$, given by

$$D_\alpha(E, \varphi \circ U, q) = \max\Big\{ d_{\varphi \circ U}(\Gamma) \mid \Gamma \in \mathcal{C}(E, U, q) \backslash \{E\}, \\ U(\Gamma) \geq U(E) + \alpha \Big\} \tag{16}$$

appears to be less than or equal to $D_\alpha(E, U, q)$ when φ is strictly concave, thus leading to a potential speed-up for annealing. Roughly speaking, the underlying idea is that an increasing concave transform of the cost functional exaggerates the depth of lower energy minima. As an example, let us consider the simple energy landscape depicted in Fig. 1(a):

the configuration space is defined by $E=\{x_i; i=1,\cdots,15\}$, the possible values for the energy U are the integers ranging from 1 to 11, and the communication kernel is such that $q(x,y)>0$ if and only if $x=x_i$ and $y=x_{i\pm1}$. Using a logarithmic transform, one obtains the energy landscape $(E,\ln U,q)$ displayed in Fig. 1(b). The global minima x_4 and x_{14} now appear much deeper compared to x_{10} and it seems to be easier to exit from the cycle $\{x_6,\cdots,x_{10}\}$ so that the convergence toward $E_{\min}$ is facilitated. Another way to show this is to examine the effect of the transform on D_α by letting, for instance, $\alpha=\frac{1}{2}$. From the definitions (37)–(41), we have

$$D_{1/2}(E,U,q)=d_U(\{x_6,\cdots,x_{10}\})=\frac{U(x_{11})-U(x_{10})}{U(x_{10})-U(E)}=7$$

whereas $D_{1/2}(E,\ln U,q)=d_{\ln U}(\{x_6,\cdots,x_{10}\})\approx 2.17$, and it follows that $M_{1/2}(N)$ is divided by about $N^{0.318}$ when $(E,\ln U,q)$ is considered.

Putting aside the trivial case for which $D_\alpha(E,U,q)=0$, all of the above is made precise in the following theorem, adapted from [30], whose proof is given in Appendix D.

Theorem 1: For any energy landscape (E,U,q) and any level $\alpha\in\mathbb{R}_+^*$ such that the set $\{\Gamma\in\mathcal{C}(E,U,q)\backslash\{E\}\mid U(\Gamma)\geq U(E)+\alpha, H_U(\Gamma)>0\}$ is not empty, for any continuously differentiable function φ: $(\mathcal{A},\mathcal{B})\rightarrow\mathbb{R}$ with $\mathcal{A}<U(E)$ and $\mathcal{B}>\max_{z\in E}U(z)$, increasing and strictly concave in $(\mathcal{A},\mathcal{B})$, the level sets $\mathcal{C}(E,U,q)$ and $\mathcal{C}(E,\varphi\circ U,q)$ are equal and $D_\alpha(E,\varphi\circ U,q)<D_\alpha(E,U,q)$.

Immediate candidates for the distortion of the energy function are

$$\varphi_1^{\tau,a}(u)=(u-a)^{1/\tau},\tag{17}$$

$$\varphi_2^a(u)=\ln(u-a)\tag{18}$$

where $\tau\in(1,+\infty)$ and $a\in(-\infty,U(E))$. The problem of choosing suitable values for τ and a together with the fact that many other families of functions are conceivable bring us naturally to the question of whether a theoretical mean of comparison can be found. As regards this points, it is mentioned in [30] that larger values of $-\varphi''/\varphi'$ should lead to greater decreases in the difficulty of the energy landscape, though the motivations for this suggestion do not appear to be very clear. Still, we can prove the following theorem (see Appendix E).

Theorem 2: Let (E,U,q), α, $\mathcal{A}$, and $\mathcal{B}$ be as specified in Theorem 1. Let φ_i and φ_j be some real valued functions, twice continuously differentiable and strictly increasing in $(\mathcal{A},\mathcal{B})$. If

$$\left(-\frac{\varphi_i''}{\varphi_i'}\right)(u)<\left(-\frac{\varphi_j''}{\varphi_j'}\right)(u)\quad\forall\, u\in\left(U(E),\max_{z\in E}U(z)\right)\tag{19}$$

then $D_\alpha(E,\varphi_j\circ U,q)<D_\alpha(E,\varphi_i\circ U,q)$.

Hence, φ_j is a "better" candidate than φ_i if condition (19) holds, which we denote by $\varphi_j\succ\varphi_i$. Obviously, we have $\varphi_2^a\succ\varphi_1^{\tau',a}\succ\varphi_1^{\tau,a}\ \forall\, a<U(E), \tau'>\tau>1$, and, by extension to the family of functions $(\varphi_2^a)_{a<U(E)}$, it appears that one should rather concentrate on the family defined by

$$\varphi_3^{\tau,a,b}=\ln((b-a)^\tau-(b-u)^\tau),\tag{20}$$

with $\tau\geq 1$, $a<U(E)$ and $b>\max_{z\in E}U(z)$, which satisfies $\varphi_3^{\tau',a,b}\succ\varphi_3^{\tau,a,b}\ \forall\ \tau'>\tau\geq 1$. Furthermore, straightforward calculations show that the parameter a (respectively, b) should be as close as possible to its upper (respectively, lower) bound for best results. However, care should be taken with the unconditional use of the mean of comparison provided by Theorem 2 because the behavior of the constants involved in Theorem 3 (Appendix B) is unclear under distortion of the energy function. Also, note that it may not be fruitful to spend too much time in trying to find the best possible distortion function according to (19). In order to support this last remark, consider for instance the family $(\varphi^{\tau,a})_{\tau\in\mathbb{R}_+^*,a\in\mathbb{R}}$ defined by $\varphi^{\tau,a}(u)=-\exp(-r(u-a))$. Though these functions have the remarkable property that $-(\varphi^{\tau,a})''/(\varphi^{\tau,a})'=\tau$, they are practically unfeasible even for small values of τ.

V. Practical Selection of the Cooling Schedule

In practical situations, the critical constants of the energy landscape are generally unknown. Therefore, it is not possible to rely on the theoretical convergence results laid out in Appendix B to find some guidelines for the choice of the parameters of the cooling schedule. Usually, the length N of the annealing chain is fixed in advance, depending on the available computing resources. The cooling schedule (12), (13) can be written as

$$\beta_n^N=\beta_{\min}\left(\frac{\beta_{\max}}{\beta_{\min}}\right)^{n/N}\tag{21}$$

and one is then left with the problem of finding suitable values for the initial and final temperature values $T_{\max}=\beta_{\min}^{-1}$ and $T_{\min}=\beta_{\max}^{-1}$. This key issue has been addressed by many authors in the "early age" of simulated annealing [36]. However, all the proposed techniques are based either on empirical rules or on theoretically unjustified assumptions and none of them stands apart from the others. This leads us to believe that conceptually simple selection criterions should be preferred to sophisticated ones as far as the implementation of the algorithm is concerned. In the following paragraphs, we discuss our approaches for setting the initial and final temperature values in the experiments about image restoration.

A. Initial Temperature Value

An intuitive but widely used means of selection of the initial temperature value is to determine $\beta_{\min}$ in such a way that most transitions are accepted; i.e., $\Sigma_{z\in E\backslash\{x\}}p_{U,\beta_{\min}}(x,z)$ is close to one for almost all $x\in E$. This qualitative criterion can be assessed via the acceptance ratio associated with a Markov chain and defined up to time m by

$$\chi((X_n)_{n\in\mathbb{N}};m)=m^{-1}\sum_{n=1}^{m}\mathbf{1}_{\{X_n\neq X_{n-1}\}}.\tag{22}$$

In the case of a Metropolis algorithm (i.e., $\beta=T^{-1}=$ constant), one verifies that (22) converges to

$$\chi^\infty(T)\doteq\sum_{x\in E}\pi_{U,T^{-1}}(x)\sum_{z\in E\backslash\{x\}}p_{U,T^{-1}}(x,z)\tag{23}$$

in probability. Hence, rigorously speaking, the selection of $\beta_{\min}$ reduces to the estimation of $T_{\max}$ such that $\chi^{\infty}(T_{\max})$ is equal to a given value $\chi_{\max} \in (0,1)$ preferably close to one. It may be argued that the precise knowledge of $T_{\max}$ is not crucial as rough approximations (see, e.g., [37]) appear to be generally satisfactory. Nevertheless, precise estimates are necessary to allow comparison between the distortion functions introduced in the previous section. We propose to simulate a discrete-time stochastic process $(X_n)_{n\in N}$ on E with uniform initial distribution and transitions

$$P(X_n = y \mid X_{n-1} = x) = p_{U,T_k^{-1}}(x,y), \qquad k = \left\lfloor \frac{n-1}{N_0} \right\rfloor, \qquad N_0 \in \mathbb{N}^*$$

where $\lfloor \cdot \rfloor$ is the floor operator; the sequence $(T_k)_{k\in\mathbb{N}}$ being adaptively generated such that the current acceptance ratio $\chi_k \doteq \chi((X_n)_{n\geq kN_0}; N_0)$ hopefully converges to $\chi_{\max}$ in probability. Let us assume temporarily that χ^{∞} is approximately linear in $[T_a, T_b]$, with $\chi^{\infty}(T_a) < \chi_{\max} < \chi^{\infty}(T_b)$, and that we have two values T_0, $T_1 \in [T_a, T_b]$. Then, clearly, if $N_0 \to \infty$ (which implies that $\chi_k \to \chi^{\infty}(T_k)$), the convergence can be obtained by putting

$$T_{k+1} = T_k + (\chi_{\max} - \chi_k)\frac{T_k - T_{k-1}}{\chi_k - \chi_{k-1}} \tag{24}$$

which corresponds to the secant model for finding a root of the function $\chi^{\infty}(T) - \chi_{\max}$. In practical situations, the behavior of χ^{∞} with respect to the temperature is dependent on the energy landscape (E, U, q). However, provided that $q(x,x)$ remains constant for all $x \in E$, one can show that there exists $\tilde{T} < +\infty$ such that χ^{∞} is strictly monotonic increasing in $(\tilde{T}, +\infty)$. Still, N_0 is finite and the generation mechanism (24) should be slightly modified in order to stabilize the search. This can be done be forcing T_{k+1} to lie in an interval defined on the basis of the current temperature value T_k; for instance, it is quite reasonable to impose $T_{k+1} \in [T_k/2, 2T_k]$ in addition to (24), so that

$$T_{k+1} = \min\left\{2T_k, \max\left\{\frac{T_k}{2}, T_k + (\chi_{\max} - \chi_k) \cdot \frac{T_k - T_{k-1}}{\chi_k - \chi_{k-1}}\right\}\right\}. \tag{25}$$

The simulation is initialized by choosing a large value for T_0 and setting $T_1 = T_0/2$. The stopping criterion is given by either $k = k_{\max}$ or $|T_k - T_{k-1}| \leq \epsilon T_{k-1}$ during a fixed number of iterations. Though the convergence of χ_k does not seem possible to establish, we never experienced any case of divergence using the above scheme with N_0 sufficiently large.

B. Final Temperature Value

Suppose that the true image x^0 is known and assume further that x^0 is an isolated local minimum of (E, U, q), i.e.,

$$x^0 \in E_{\text{loc}} \doteq \{x \in E \mid U(z) - U(x) > 0 \ \ \forall\, z \in E, \ q(x,z) > 0\}. \tag{26}$$

Then, $p_{U,\beta}(x^0, z)$ is known for all $z \in E$, $\beta \in \mathbb{R}_+$, and, in our view, one very attractive method of selecting $\beta_{\max}$ would be to choose among those values of β for which

$$\sum_{z\in E\setminus\{x^0\}} p_{U,\beta}(x^0, z) \leq \chi_{\min} \tag{27}$$

where $\chi_{\min} \in (0,1)$ is close to zero. However, x^0 is unknown in practice and it is impossible to guarantee that (26) holds for all noise realizations η, even though $\mathrm{P}(x^0 \in E_{\text{loc}})$ may be arbitrarily close to one if x^0 belongs to some prototypical image class and λ is large enough [5]. Instead, we have chosen to select $\beta_{\max}$ such that $\Sigma_{z\in E\setminus\{x^{\text{loc}}\}} p_{U,\beta_{\max}}(x^{\text{loc}}, z) = \chi_{\min}$ for a configuration $x^{\text{loc}} \in E_{\text{loc}}$ obtained from coordinate-wise deterministic minimization. Because $f(\beta; x) \doteq \Sigma_{z\in E\setminus\{x\}} p_{U,\beta}(x,z)$ is decreasing and strictly convex in $\mathbb{R}_+^*$ with respect to β for all $x \in E_{\text{loc}}$, $\beta_{\max}$ can be estimated using any common root finding algorithm (we use the secant method). The workload involved in such an estimation process is quite reasonable. Assuming for convenience that the support of $h_s, ., .$ is M^2 pixels wide for all $s \in S'$, then $f(\beta; x^{\text{loc}})$ can be computed in $O(M^2|\{z \in E\setminus\{x^{\text{loc}}\} | q(x^{\text{loc}}, z) > 0\}|)$ operations, which reduces for instance to $O(M^2(L-1)|S|)$ operations per iteration if $E = \Lambda^{|S|}$ and $q = q_{\Lambda}$ (7), (14).

VI. Experiments

We present experiments on simulated synthetic aperture imaging (SAI) data which compare the relative performances of annealing algorithms for the different options evoked in this paper and demonstrate the merit of the class of algorithms $(\Omega_{\delta}^{|S|},\ \varphi \circ U,\ q_{\delta},\ \nu_0,\ (\beta_0 \exp(n\xi))_{1\leq n\leq N})$. In this context, a 2-D scene, or reflectivity distribution, is imaged by moving a small antenna along a linear track usually called azimuth. This radiating element emits a single wide-bandwidth signal and receives the echoes coming back from the different scatterers. The procedure is repeated for equally spaced azimuthal positions, such that a large synthetic antenna is formed and spatially sampled along a lateral direction. The juxtaposition of the echo signals makes it possible to form an image whose temporal (axial) direction is identified by range. In synthetic aperture radar (SAR) imaging, a high range resolution is achieved by correlating each temporal line with the emitted wideband signal (i.e., matched filter processing). In addition, the azimuthal resolution can be improved by integrating all the echoes along the synthetic antenna [38]. However, since the PSF is not stationary along the axial direction, severe difficulties are encountered for wide-range SAI systems. Consequently, SAI is most often performed on limited axial areas for space-invariance approximation [39] and it would be very useful to image wider areas under the best possible resolution. Though efforts are made in this direction in SAR imaging [40], [41], efficient wide-range SAI data processing techniques are scarce; this need arises in nondestructive testing (NDT) and medical ultrasound imaging applications as well [42], [43].

We consider the side-looking case represented in Fig. 2 such that the following material adapts easily to most types of SAI systems. The stationary target distribution x^0 to be imaged lies on the bottom $(z = 0)$ and the synthetic aperture is

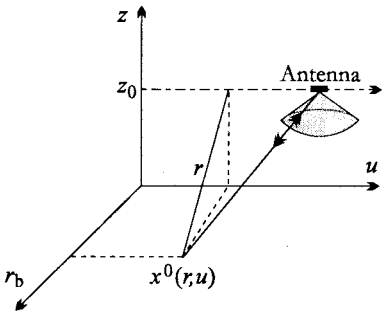


Fig. 2. Geometry of a side-looking synthetic aperture imaging system.

formed by the motion of a real antenna, or transducer, along a straight linear track located at height z_0. This lateral track, or azimuth, is specified by u and the slant range is identified by r. The "start-stop" approximation is adopted here, i.e., the temporal Doppler effect due to the sensor movement between transmission and reception of a signal is ignored. Note that $z_0 = 0$ in the case of NDT or medical ultrasound imaging applications, unlike SAR imaging systems where a cosine correction factor must be considered in order to convert from slant range r to bottom range r_b. The PSF of the system is the ensemble of echo signals received from a single reflecting point $\delta(r - r',\ u - u')$. It can be expressed as

$$h(r,u;r',u') = \theta(r,r',u-u')g(r-\zeta(r',u-u')) \cdot \exp(-2j\omega_c c^{-1}(\zeta(r',u-u')-r')), \quad (28)$$
$$\zeta(r',v) \doteq ((r')^2+v^2)^{1/2}$$

with the following interpretation: $\zeta(r',v)-r'$ is the hyperbolic range migration, g is the envelope of the (possibly processed) received signal with carrier angular frequency ω_c, c denotes the wave propagation speed and the beam pattern θ of the physical antenna is defined by

$$\theta(r,r',v) = \mathcal{F}^{-1}\left[\operatorname{sinc}^2\left(\frac{2\pi\mathcal{L}f_r v}{c\zeta(r',v)}\right)\right] \quad (29)$$

where $\mathcal{F}^{-1}$ denotes the one-dimensional (1-D) inverse Fourier transform and $\mathcal{L}$ is the length, or diameter, of the antenna. Clearly, the PSF (28) is space-invariant with respect to the azimuthal direction and we can write $h(r,u;r',u') = h(r,u-u';r',0) \doteq h(r,u-u';r')$. The exact data are then given by

$$d_e(r,u) = \iint x^0(r',u')h(r,u-u';r')\,dr'\,du' = \int x^0(r',u) \odot_u h(r,u;r')\,dr' \quad (30)$$

where "$\odot_u$" denotes the 1-D convolution operator in the u-direction. Within a narrow-range segment around $r = r_0$, which dimension can be taken to be equal to the depth of focus of the synthetic aperture [44], the PSF can be treated as space-invariant such that (30) reduces to a classical 2-D convolution: $d_e(r,u) \approx x^0(r,u) \odot_r \odot_u h(r,u;r_0)$. However, this approximation is not considered here and (30) together with (28) and (29) will serve as a reference for the two examples that follow.

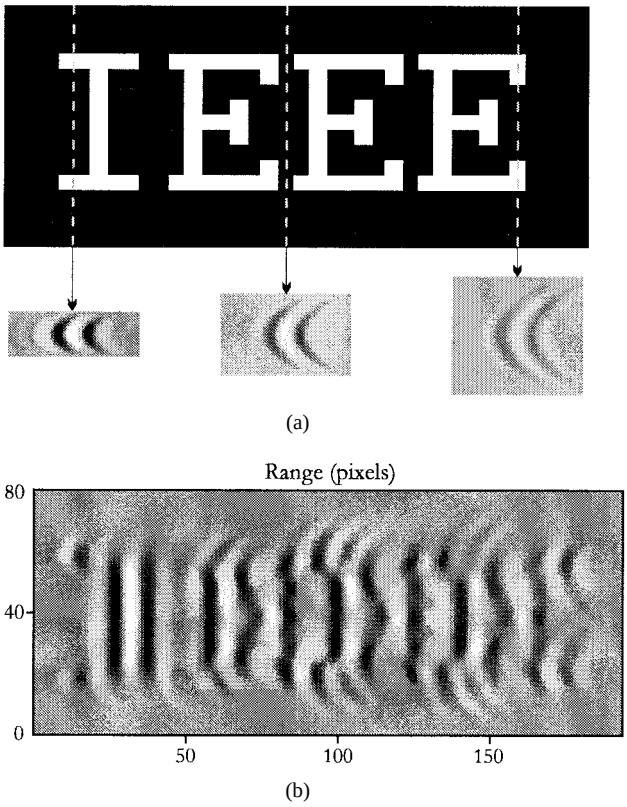


Fig. 3. "IEEE" image: (a) original image and PSF of the imaging system at different range values; (b) data (25 dB BSNR).

We shall first consider the nonphysical simulation depicted in Fig. 3. The original image x^0 is a text sample with maximum gray level value equal to 128. Following (3), the data were generated by degrading x^0 with a discrete blur operator H derived from the above image formation model and a zero-mean, white Gaussian noise with variance σ^2 specified by a 25 dB blurred signal-to-noise ratio (BSNR). Recall that

$$\mathrm{BSNR} \doteq 10 \cdot \log\left(\frac{1}{|S'|}\sum_{s\in S'}\frac{((Hx^0)_s - \overline{d_e})^2}{\sigma^2}\right) \quad (31)$$

where $\overline{d_e} \doteq |S'|^{-1}\,\Sigma_{s\in S'}\,Hx^0$ is the mean of the exact data. The total image space $E = \Lambda^{|S|}$ is defined by a regular grid S of size 122×234 ($|S'| = 80 \times 192$) and $L = 256$. The second example (see Fig. 4), for which $|S| = 146 \times 192$ and $|S'| = 100 \times 150$, was obtained in a similar way with a 0 dB BSNR. It is inspired by the nondestructive evaluation of a sample containing holes parallel to the control surface, as shown in Fig. 5. The data displayed in Fig. 4(b) then correspond to a B-scan image resulting from the combination of A-scan signals recorded for equally spaced locations of a broad-band, unfocused transducer acting both as a transmitter and as a receiver. The -3 dB main lobe width α_{-3} of the transducer's beam pattern is of 11°; this angle determines the lateral length of the synthetic aperture defined by $l_{\mathrm{sa}}(r') = 2r'\tan(\alpha_{-3}/2)$ for a point scatterer located at the axial distance r' from the transducer. Here, the range interval of the image area is [20 mm, 100 mm] such that l_{sa} varies from 3.8 to 18.9 mm.

In every experiment, the length N of the annealing chain is equal to$6000 \cdot |S|$. For exponential cooling schedules, the

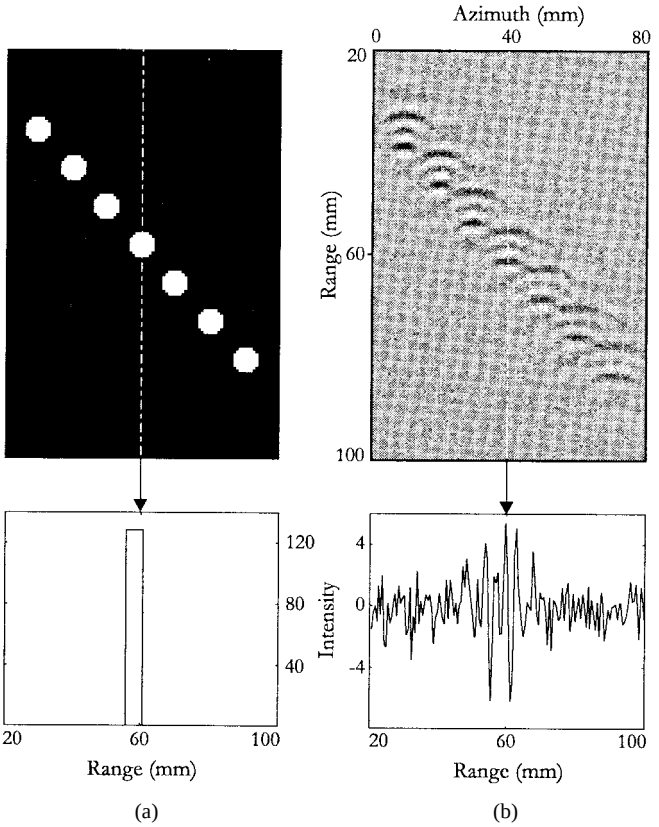


Fig. 4. Simulated B-scan image: (a) original image; (b) data (0 dB BSNR).

initial and final temperature values are selected according to the methods set out in Section V with $\chi_{\max} = 0.8$ and $\chi_{\min} = 5 \cdot 10^{-4}$. The starting point X_0 of the annealing algorithm is given by the configuration obtained by the end of the search of the initial temperature; this search being initialized with the realization of a uniform white noise process or the corresponding locally bounded approximation [29] if the minimization is to be carried out on $\Omega_\delta^{|S|}$. In this last case, $\mathcal{N}(S)$ is the four-nearest neighbor system and we choose $\delta = 5$ which appears to be adequate for $L = 256$ [29]. For the two examples considered here, the edges are of 128 gray levels and we have found that setting $\Delta = 256$ gives good results. Indeed, due to the concavity of the ϕ function (6), many other choices are possible, most of which are suitable for a wide range of discontinuity amplitudes. More important is the selection of the hyperparameter λ which balances the data fidelity term against the roughness penalty. When the energy function is not distorted, one may suggest to make use of the method devised in [5] by taking the mean sum of squared blurring coefficients $\Sigma_t h_{s,t}^2$ over internal pixels $s \in S'$. Note that since the prior (5) includes diagonal pairs of adjacent pixels, we must then take into account the correction given in [45]. However, the values thus obtained most often yield to over-biased solutions. This result can be partially explained for restricted image spaces, but further investigations that are beyond the scope of this paper remain necessary. Hence, we decided to select λ by trial and error in order to produce restorations which appear best visually, resulting in $\lambda = 1$ for the "IEEE" image and $\lambda = 4$ for the simulated B-scan. Still, in both cases, satisfactory solutions are obtained for any value with the same order of magnitude.

A. "IEEE" Image

Here, we consider the annealing algorithms

$$\mathcal{A}_{\ln}^{\Lambda} \doteq (\Lambda^{|S|}, U, q_\Lambda, \nu_0, (\beta_0' \ln(n+1))_{1 \le n \le N}), \tag{32}$$

$$\mathcal{A}_{\exp}^{\Lambda} \doteq (\Lambda^{|S|}, U, q_\Lambda, \nu_0, (\beta_0 \exp(n\xi))_{1 \le n \le N}) \tag{33}$$

$$\mathcal{A}_{\exp}^{\Omega} \doteq (|\Omega_5^{|S|}, U, q_5, \nu_0, (\beta_0 \exp(n\xi))_{1 \le n \le N}), \tag{34}$$

$$\mathcal{A}_{\exp}^{\Omega,\varphi} \doteq (\Omega_5^{|S|}, \varphi \circ U, q_5, \nu_0, (\beta_0(\varphi) \exp(n\xi(\varphi)))_{1 \le n \le N}) \tag{35}$$

in order to assess the successive benefits coming with exponential cooling schedules, restricted image spaces and increasing concave transforms of the cost functional.

Let us first start with a few visual examples. The restoration result shown in Fig. 6(a) was obtained with the algorithm $\mathcal{A}_{\exp}^{\Omega}$, $\beta_0 = 0.1182$, $\xi = 6.3135 \cdot 10^{-8}$. It clearly demonstrates the potential of the method, as the recovery of the original image [Fig. 3(a)] from the degraded data [Fig. 3(b)] is an ambitious challenge. Not surprisingly, the quality of the restoration appears to be less satisfactory when coordinate-wise deterministic minimization is used [see Fig. 6(b), the starting point is a constant image]. The result displayed in Fig. 6(c) does not show any improvement; it was achieved with $\mathcal{A}_{\ln}^{\Lambda}$, where β_0' is such that $\beta_0' \ln(N+1) = \beta_0 \exp(N\xi)$, using the starting point X_0 of $\mathcal{A}_{\exp}^{\Omega}$. This leads to the rather common observation that conventional Metropolis-type annealing algorithms are easily stucked in poor local minima. In order to provide motivations for taking the shift-variance of the PSF into account, restoration was also performed using the mid-range PSF throughout the image. The minimization was carried out on $\Omega_5^{|S|}$ with an exponential cooling schedule and the corresponding result given in Fig. 6(d) is to be compared with Fig. 6(a).

The convergence rates shown in Fig. 7 illustrate the dynamic behavior of the annealing algorithms (i) $\mathcal{A}_{\ln}^{\Lambda}$, (ii) $\mathcal{A}_{\exp}^{\Lambda}$, and (iii) $\mathcal{A}_{\exp}^{\Omega}$. For suitable comparison, the starting point X_0 of (ii) is the same as for (i) and (iii). It turns out that exponential cooling schedules are to be preferred and the restriction of the image space gives significant benefits in terms of convergence speed. The additional speed-up resulting from increasing concave transform of the cost functional is depicted in Figs. 8 and 9. Fig. 8 displays the convergence rates achieved with $\mathcal{A}_{\exp}^{\Omega,\varphi}$ in the following cases: (iii) $\varphi \circ U = U$ (see also Fig. 7), (iv) $\varphi = \varphi_1^{2,a}$, (v) $\varphi = \varphi_1^{4,a}$ (17), (vi) $\varphi = \varphi_2^a$ (18), where $a = |S'|\sigma^2 \approx ||Hx^0 - d||_2^2$. Note that the parameters of the cooling schedule vary widely under distortion of the energy function; for instance, $\beta_0(\varphi_2^a) = 6.0875 \cdot 10^3$ and $\xi(\varphi_2^a) = 3.7897 \cdot 10^{-8}$. As predicted by the theory, further acceleration is achieved with $\varphi = \varphi_3^{\tau,a,b}$ (20) for increasing values of τ (see Fig. 9): (vi) $\tau = 1$ (see also Fig. 8), (vii) $\tau = 15$, (viii) $\tau = 30$, (ix) $\tau = 40$. The free parameter b was taken to be equal to $\frac{3}{2} U(\tilde{X})$ where $\tilde{X}$ is a sample of the uniform distribution on $\Lambda^{|S|}$. This turns out to be a useful rule of thumb although it can be objected that $b < \max_{z \in E} U(z)$

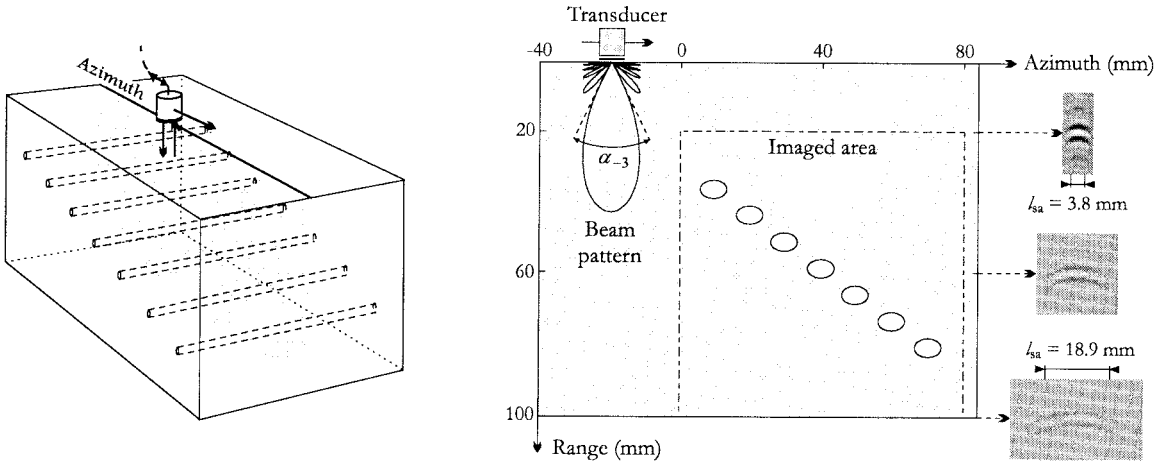


Fig. 5. Simulated experimental setup for obaining the B-scan displayed in Fig. 4(b) (nondestructive testing of a sample containing holes parallel to the control surface) and representations of the resulting PSF at different range values.

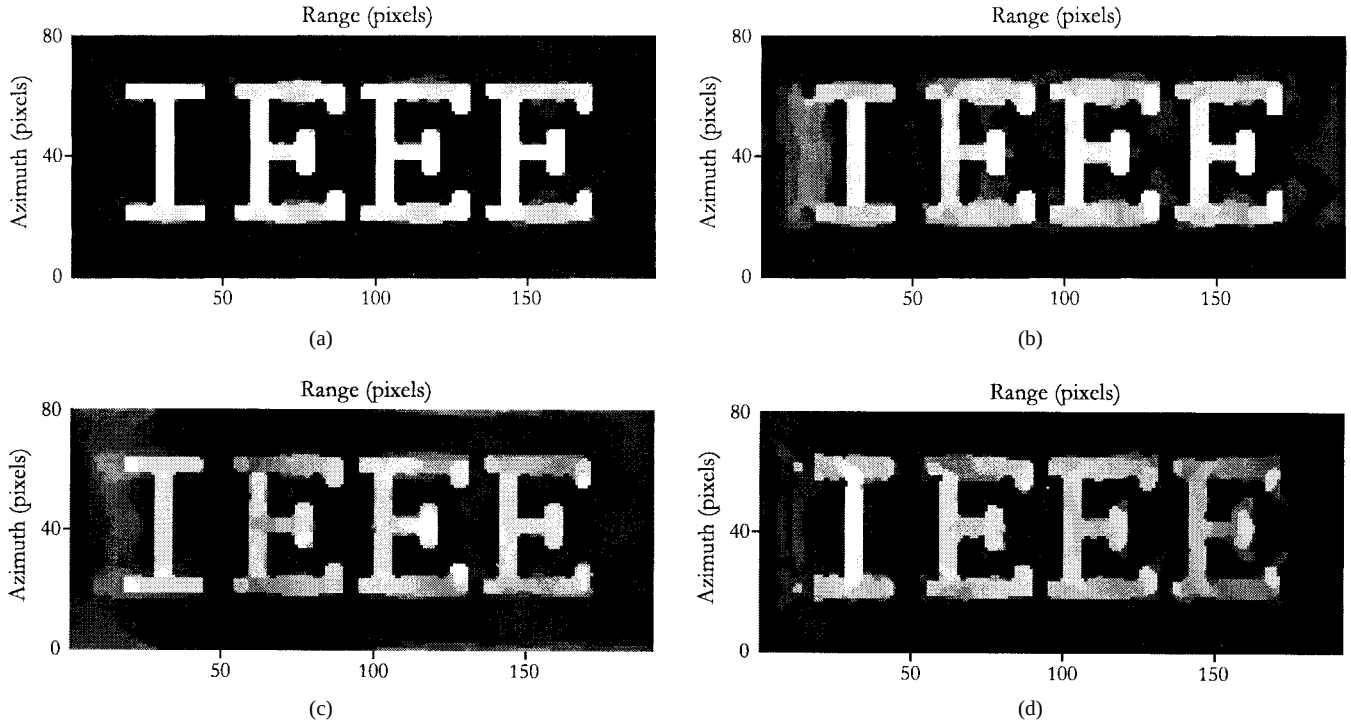


Fig. 6. Restoration results for the "IEEE" image (see Fig. 3). (a)–(c) Space-variant restoration: (a) exponential cooling schedule and locally bounded image space [algorithm $\mathcal{A}^{\Omega}_{\exp}$ (34)]; (b) Local minimum of the energy landscape $(\Omega_5^{|S|}, U, q_5)$; (c) Logarithmic cooling schedule and total image space [algorithm $\mathcal{A}^{\Lambda}_{\ln}$ (32)]. (d) Space-invariant restoration using the mid-range PSF: exponential cooling schedule and locally bounded image space.

is highly probable. Actually, b can take any value greater than the maximum energy level $\Upsilon \doteq \max_{1\leq n\leq N} U(X_n^N)$ of the annealing chain. Since we always deal with finite temperature values, it is very unlikely that $U(\tilde{X}) < \Upsilon$ except, possibly, at the beginning of the chain. It follows that $b = \frac{3}{2}U(\tilde{X})$ is secure, even during the search of the initial temperature value.

B. Simulated B-Scan Image

Similar experiments were conducted on the example shown in Fig. 4. The restoration result obtained with $\mathcal{A}^{\Omega}_{\exp}$ appears in Fig. 10(a); the associated energy value $U(X_N^N)$ is equal to 18 642.5. The low performances of coordinate-wise deterministic minimization [see Fig. 10(b), $U(X_N^N) = 18\,896.9$] and $\mathcal{A}^{\Lambda}_{\ln}$ [see Fig. 10(c), $U(X_N^N) = 24\,199.6$] illustrate its high sensitivity to the choice of the starting point. Note that the images in Figs. 10(a) and (b) exhibit comparable energy values though they are visually far apart. The space-invariant restoration result making use of the mid-range PSF is displayed in Fig. 10(d). As it could be expected, significant distorsions with respect to the true image are observed for both low and high range values. Finally, in order to appreciate the overall speed-up resulting from the restriction of the image space together with an increasing concave transform of the energy function, Fig. 11 shows the convergence rates achieved with

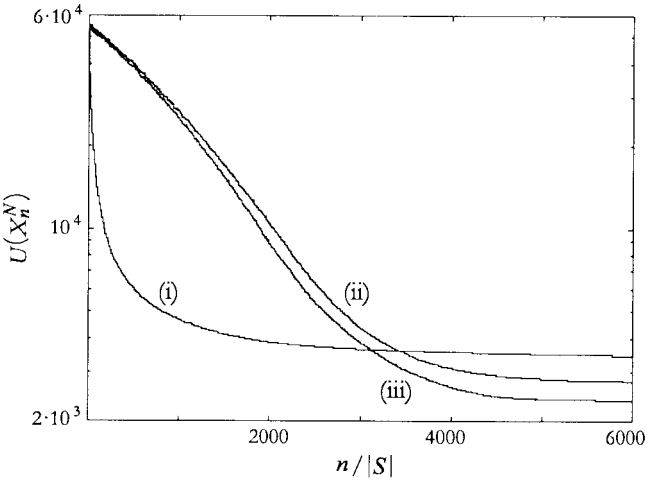


Fig. 7. Logarithmic cooling schedule versus exponential cooling schedule and acceleration by restriction of the image space: convergence rates achieved with the algorithms (i) $\mathcal{A}_{\ln}^{\Lambda}$ (32), (ii) $\mathcal{A}_{\exp}^{\Lambda}$ (33), and (iii) $\mathcal{A}_{\exp}^{\Omega}$ (34).

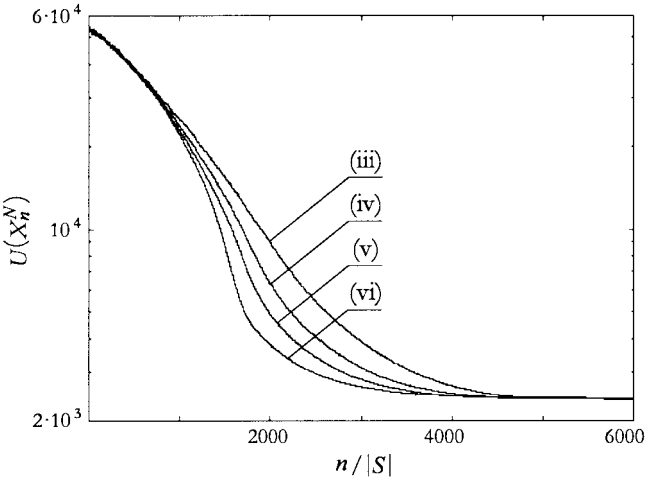


Fig. 8. Increasing concave transform of the cost functional. Convergence rates achieved with the algorithm $\mathcal{A}_{\exp}^{\Omega,\varphi}$ (35) for typical φ functions: (iii) $\varphi \circ U = U$ (see also Fig. 7); (iv) $\varphi \circ U = (U-a)^{1/2}$; (v) $\varphi \circ U = (U-a)^{1/4}$; (vi) $\varphi \circ U = \ln(U-a)$; where $a = 1466.3$.

(i) $\mathcal{A}_{\exp}^{\Lambda}$ (33) and (ii) $\mathcal{A}_{\exp}^{\Omega,\varphi}$ (35), where $\varphi = \varphi_3^{\tau,a,b}$, $\tau = 10$, $a = 0$, and b was selected as discussed in the previous paragraph.

VII. Concluding Remarks

In this paper, we have considered the image restoration problem in the case of a spatially varying blur and synthetic aperture imaging data. However, the presented material may be useful in other fields of applications. The main difficulties associated with this ill-posed inverse problem were overcome through the use of concave stabilizers that allow the preservation of discontinies in the image. The minimization of the resulting nonconvex cost functional was successfully carried out using the Metropolis dynamics with annealing.

Putting aside its ease of implementation, the Metropolis dynamics is computationally far less demanding than the Gibbs sampler. The problem of whether one is better than the other has not been completely solved yet. Partial answers can be found in [46] where it is shown that, for the Ising model, the Gibbs sampler should be preferred at high temperatures whereas the Metropolis dynamics is the best at low temperatures. Another interesting result is given in [47], where the authors establish that the Gibbs sampler and the Metropolis dynamics are asymptotically equivalent when the state space exhibits a lattice structure. Still, only the latter is feasible for inverse problems with large kernel supports which frequently arise in computed imaging.

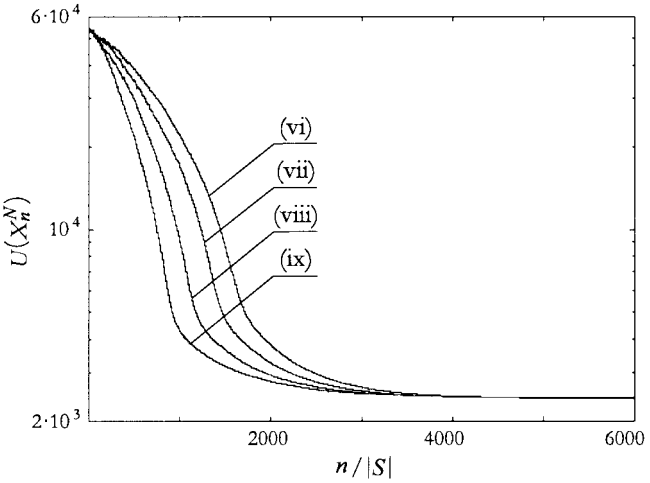


Fig. 9. Increasing concave transform of the cost functional. Convergence rates achieved with the algorithm $\mathcal{A}_{\exp}^{\Omega,\varphi}$ (35), $\varphi \circ U = \ln((b-a)^{\tau} - (b-U)^{\tau})$ $(a = 1466.3,\ b = 1.42 \cdot 10^6)$ for increasing values of τ: (vi) $\tau = 1$ (see also Fig. 8); (vii) $\tau = 15$; (viii) $\tau = 30$; (ix) $\tau = 40$.

The finite-time behavior of Metropolis-type annealing algorithms following an exponential cooling schedule appears to be particularly interesting. From Theorem 3 in Appendix B based on [26], for any initial temperature value and suitable positive constant α, the convergence measure $M_{\alpha}(N)$ (10) can be made as small as wanted by increasing the length N of the annealing chain provided that the final temperature value is small enough [i.e., $\beta_N^N \geq \alpha^{-1}\ln(\varepsilon^{-1})$]. Our main contribution was to show that some inexpensive acceleration techniques can significantly increase the performances of this class of algorithms while not altering their theoretical convergence properties. Observation of the convergence rates presented in Section VI shows that the final energy level $U(X_N^N)$ achieved with typical algorithms of the form $(\Lambda^{|S|}, U, q_{\Lambda}, \nu_0, (\beta_0 \exp(n\xi))_{1\leq n\leq N})$ can be reached about three times faster under restriction of $\Lambda^{|S|}$ to a locally bounded image space and increasing concave transform of the cost functional. Moreover, in view of these experimental results, the proposed approaches for the selection of the initial and final temperature values turn out to be very efficient.

In terms of the number of parameters, the proposed algorithm $\mathcal{A}_{\exp}^{\Omega,\varphi}$ (35) appears to be complex compared with classical methods based on convex cost functions. To summarize things, let us recall that the extra-parameters are the level δ of the restricted image space, the length N of the annealing chain, the ratios $\chi_{\max}, \chi_{\min}$ defining the cooling schedule, and the constants coming with the distortion function φ. Among these, $\delta = \lfloor 5 \cdot |\Lambda|/256 \rfloor$ is adequate for many applications [29],

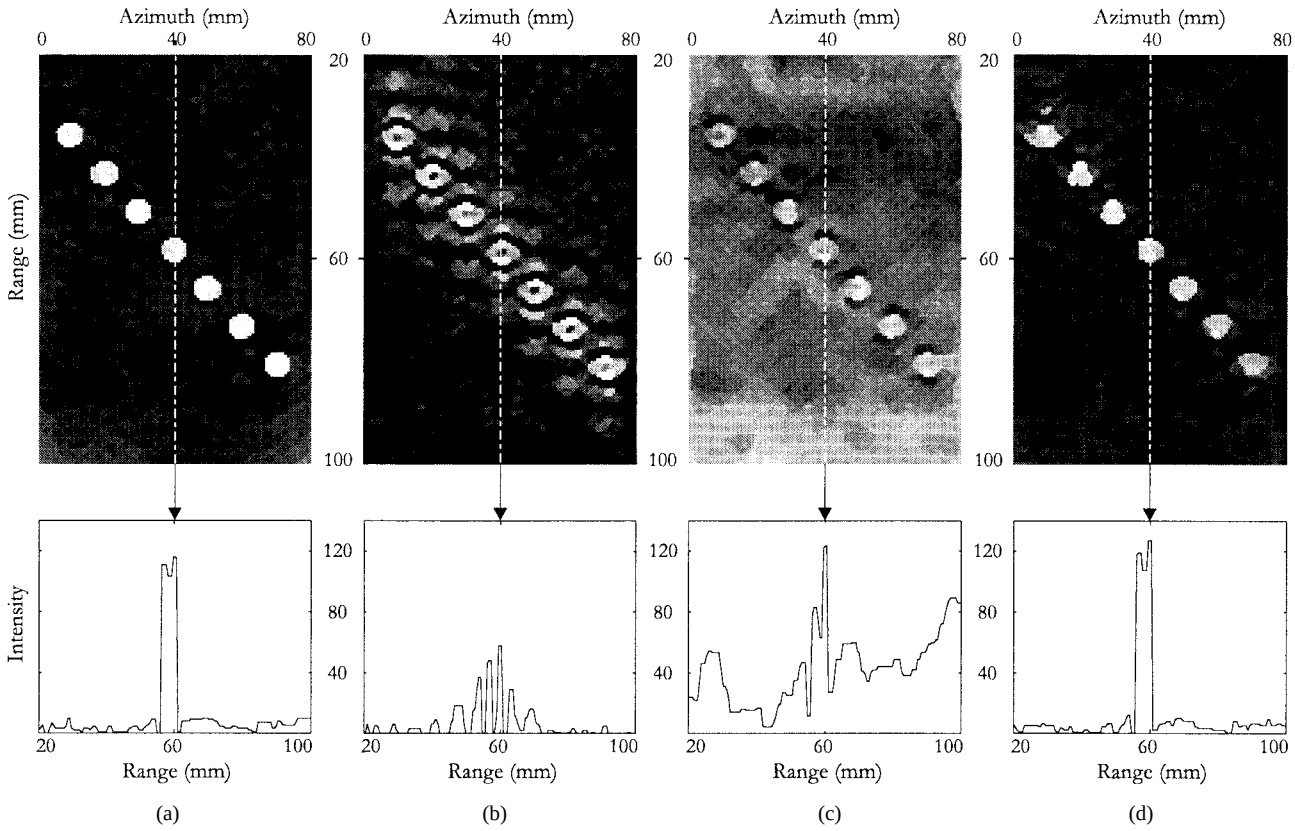


Fig. 10. Restoration results for the simulated B-scan image (see Fig. 4). (a)–(c) Space-variant restoration: (a) exponential cooling schedule and locally bounded image space [algorithm $\mathcal{A}^{\Omega}_{\exp}$ (34)]; (b) local minimum of the energy landscape $(\Omega_5^{|S|}, U, q_5)$; (c) logarithmic cooling schedule and total image space [algorithm $\mathcal{A}^{\Lambda}_{\ln}$ (32)]. (d) Space-invariant restoration using the mid-range PSF: exponential cooling schedule and locally bounded image space.

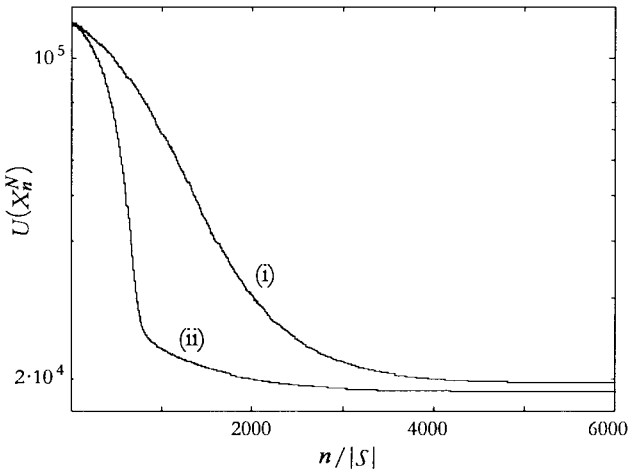


Fig. 11. Convergence rates: (i) exponential cooling schedule and total image space [algorithm $\mathcal{A}^{\Lambda}_{\exp}$ (33)]; (ii) exponential cooling schedule, locally bounded image space and increasing concave transform of the cost functional [algorithm $\mathcal{A}^{\Omega,\varphi}_{\exp}$ (35), $\varphi \circ U = \ln(b^{\tau} - (b-U)^{\tau})$, $b = 6.42 \cdot 10^5$, $\tau = 10$].

N is fixed by the available amount of computing time, and the choice $(\chi_{\max}, \chi_{\min}) = (0.8, 5 \cdot 10^{-4})$ remains appropriate in any case since $\chi_{\max}$ and $\chi_{\min}$ are independent of the energy landscape. Therefore, we are only left with the choice of φ. As suggested by Theorem 2 and our experimental results, one should systematically resort to $\varphi_3^{\tau,a,b}$ (20). Though it is always possible to set $\tau = 1$ and $a = 0$ for simplicity, $\tau = 10$ is quite fair, $a = |S'|\sigma^2$ is suitable when the noise variance is known, and b can be easily selected as discussed in Section VI-A. Based on all the above comments, it turns out that there is no particular need to calibrate the algorithm in order to observe good performance.

Finally, an important topic that remains to be addressed is the choice of the smoothing parameter λ of the cost functional, the issue of the selection of the scaling parameter Δ being generally less critical. When the minimization is to be performed on $\Lambda^{|S|}$, Geman and Reynolds [5] derived an explicit formula for choosing λ in the space-invariant case. Preliminary studies indicate that their approach can be adapted to restricted image spaces and shift-variant PSF's at the expense of numerical computations, a tradeoff which may indeed be worthwhile since *a priori* information about the discontinuities to be recovered (i.e., expected number, orientation, and amplitude) can then be easily introduced.

Appendix A
Cycle Decomposition and Difficulty of the Energy Landscape

Let us define on E the communication relation $\overset{v}{\leftrightarrow}$ at level $v \in \mathbb{R}$ by $x \overset{v}{\leftrightarrow} y$ if either $x = y$ and $U(x) \leq v$ or there exists a finite family $(s_k)_{0 \leq k \leq p}$ of configurations such that $x = s_0$, $y = s_p$, $q(s_k, s_{k+1}) > 0$ for all $0 \leqslant k < p$ and $U(s_k) \leq v$ for all $0 \leq k \leq p$. It follows from the symmetry of q that $\overset{v}{\leftrightarrow}$ is an equivalence relation on E. Let $\mathcal{C}_v(E, U, q)$ be the partition of E formed by the set of all the equivalence classes of $\overset{v}{\leftrightarrow}$, we

define the set of cycles of the energy landscape (E,U,q) by

$$\mathcal{C}(E,U,q)=\bigcup_{v\in\mathbb{R}}\mathcal{C}_v(E,U,q). \tag{36}$$

An immediate consequence of the above definition is that the inclusion relation on $\mathcal{C}(E,U,q)$ defines a tree whose leaves are the points of E and whose root is E itself. Hence, two cycles in E are either disjoint sets or have full intersection. For any cycle $\Gamma\in\mathcal{C}(E,U,q)\backslash\{E\}$ we define the boundary of Γ by

$$B(\Gamma)=\{z\in E\backslash\Gamma \mid \exists x\in\Gamma, q(x,z)>0\} \tag{37}$$

its energy by

$$U(\Gamma)=\min_{z\in\Gamma}U(z) \tag{38}$$

its depth by

$$H_U(\Gamma)=\min_{z\in B(\Gamma)}(U(z)-U(\Gamma))^+ \tag{39}$$

and its difficulty by

$$d_U(\Gamma)=H_U(\Gamma)/(U(\Gamma)-U(E)). \tag{40}$$

Using the above definitions, the difficulty of the energy landscape (E,U,q) at level $\alpha\in\mathbb{R}_+^*$ writes

$$D_\alpha(E,U,q)=\max\Big\{d_U(\Gamma)|\Gamma\in\mathcal{C}(E,U,q)\backslash\{E\}, U(\Gamma)\geq U(E)+\alpha\Big\}. \tag{41}$$

Appendix B
Exponential Cooling Schedules

We give here a limited form of [26, Th. 8.1] and we show that, as a straightforward consequence, carefully chosen finite time exponential cooling schedules make it possible to obtain $M_\alpha(N)$ of order $N^{-D_\alpha^{-1}(E,U,q)}$ for N sufficiently large and α small enough. Let us put $D_\alpha=D_\alpha(E,U,q)$ and define, for any subset A of E, $\mathcal{M}(A)$ to be the partition of A formed by the maximal elements of $\mathcal{P}(A)\cap\mathcal{C}(E,U,q)$, where $\mathcal{P}(A)$ is the power set of A.

Theorem 3: For any energy landscape (E,U,q) with symmetric and irreducible communication kernel, there exists positive constants B and K such that for any positive constants β_0, $\varepsilon<1$, $\alpha\leq\min\{U(\Gamma)-U(E)\mid\Gamma\in\mathcal{M}(E\backslash E_{\min})\}$,[1] for any initial distribution ν_0 and for any finite-time cooling schedule $(\beta_n^N)_{1\leq n\leq N}$ such that

$$\beta_n^N=\beta_0\exp(n\xi), \tag{42}$$

$$\xi\leq B\varepsilon^{D_\alpha}(\ln(\varepsilon^{-1}))^{-1}, \tag{43}$$

$$N\geq\xi^{-1}(\ln(\ln(\varepsilon^{-1}))-\ln(\alpha\beta_0)) \tag{44}$$

the finite-time annealing algorithm $(X_n^N)_{0\leq n\leq N}$ associated with $(E,U,q,\nu_0,(\beta_n^N)_{1\leq n\leq N})$ satisfies

$$\mathrm{P}\Big(U(X_N^N)\geq U(E)+\alpha) \leq K\varepsilon\exp(\beta_0 D_\alpha^{-1}\max_{\Gamma\in\mathcal{M}(E\backslash E_{\min})}H_U(\Gamma)\Big). \tag{45}$$

[1]The original form of the theorem authorizes in particular any positive value for α by appealing to a quantity $\bar{D}_\alpha$ which can be shown to be equal to D_α if $\alpha\leq\min\{U(\Gamma)-U(E)\mid\Gamma\in\mathcal{M}(E\backslash E_{\min})\}$.

Let us fix an arbitrary positive value for β_0 and choose $\xi=N^{-1}(\ln N-\ln(\alpha\beta_0))$. It is easy to check that (43) and (44) hold if $\varepsilon\geq\varepsilon_0(N)\doteq\max\{\exp(-N),\varepsilon_N\}$, where ε_N is such that

$$\frac{\varepsilon_N^{D_\alpha}}{\ln(\varepsilon_N^{-1})}=\frac{1}{BN}\ln\left(\frac{N}{\alpha\beta_0}\right).$$

Hence, for any $\varepsilon\in(0,1)$, there exists $N_0(\varepsilon)<\infty$ such that the conditions (43) and (44) are fulfilled $\forall\, N\geq N_0(\varepsilon)$. Using inequalities (11) and (45), we can then write

$$k_1N^{-D_\alpha^{-1}}\leq M_\alpha(N)\leq k_\alpha\varepsilon_0(N)$$

for some positive constants k_1, k_α linked to the energy landscape. Thus

$$-D_\alpha^{-1}\leq\lim_{N\to\infty}\frac{\ln M_\alpha(N)}{\ln N}\leq\lim_{N\to\infty}\frac{\ln\varepsilon_0(N)}{\ln N}$$

where the upper bound reduces to $\lim_{N\to\infty}(\ln N)^{-1}\ln\varepsilon_N=-D_\alpha^{-1}$ and it follows that $M_\alpha(N)$ is of order $N^{-D_\alpha^{-1}}$ for N sufficiently large or, equivalently, as $\varepsilon\to 0$.

Appendix C
Markov Kernel on a Locally Bounded Image Space

Consider a configuration $x\in\Omega_\delta^{|S|}$ and let y be the new candidate generated from x by affecting a gray level value l to a randomly selected site s. From (15), an immediate condition for y to be in $\Omega_\delta^{|S|}$ is

$$l\in\Lambda\cap\{\min_{t\in\mathcal{N}_s(S)}x_t-\delta,\cdots,\max_{t\in\mathcal{N}_s(S)}x_t+\delta\}.$$

However, this condition does not appear to be sufficient, since the new value l also provides bounds for the neighbors of s. By applying (15) to a given site $t\in\mathcal{N}_s(S)$, we get

$$\min\{l,\min_{u\in\mathcal{N}_t(S)\backslash\{s\}}x_u\}-\delta \leq x_t\leq\max\{l,\max_{u\in\mathcal{N}_t(S)\backslash\{s\}}x_u\}+\delta$$

and it follows that if $x_t>\max\{x_u;u\in\mathcal{N}_t(S)\backslash\{s\}\}+\delta$ $(<\min\{x_u;u\in\mathcal{N}_t(S)\backslash\{s\}\}-\delta)$, then one must verify $l\geq x_t-\delta\geq\min_{t\in\mathcal{N}_s(S)}x_t-\delta$ $(\leq x_t+\delta\leq\max_{t\in\mathcal{N}_s(S)}x_t+\delta)$. Taking this observation into account for all $t\in\mathcal{N}_s(S)$ leads straight to the following mathematical description. Define for any site $s\in S$ the sets $J_{\sup}^\delta(x,s)=\{t\in\mathcal{N}_s(S)\mid x_t>x_u+\delta\ \forall u\in\mathcal{N}_t(S)\backslash\{s\}\}$ and $J_{\inf}^\delta(x,s)=\{t\in\mathcal{N}_s(S)\mid x_t<x_u-\delta\ \forall\, u\in\mathcal{N}_t(S)\backslash\{s\}\}$. A necessary and sufficient condition for y to be in $\Omega_\delta^{|S|}$ is

$$l\in\Lambda\cap\{\gamma_{\min}^\delta(x,s),\cdots,\gamma_{\max}^\delta(x,s)\}\doteq\omega_s^\delta(x)$$

where

$$\gamma_{\min}^\delta(x,s)=\begin{cases}\max_{t\in J_{\sup}^\delta(x,s)}x_t-\delta, & \text{if } J_{\sup}^\delta(x,s)\neq\emptyset,\\ \min_{t\in\mathcal{N}_s(S)}x_t-\delta, & \text{otherwise}\end{cases}$$

and

$$\gamma_{\max}^\delta(x,s)=\begin{cases}\min_{t\in J_{\inf}^\delta(x,s)}x_t+\delta, & \text{if } J_{\inf}^\delta(x,s)\neq\emptyset\\ \max_{t\in\mathcal{N}_s(S)}x_t+\delta, & \text{otherwise.}\end{cases}$$

The communication kernel q_δ on $\Omega_\delta^{|S|}$ is then defined by

$$q_\delta(x,y)=\begin{cases}(|S|\cdot|\omega_s^\delta(x)|)^{-1}, & \text{if } y_t=x_t\ \forall\, t\neq s,\\ & \quad y_s\in\omega_s^\delta(x)\backslash\{x_s\},\\ 0, & \text{if } |\{s\mid x_s\neq y_s\}|>1,\\ 1-\displaystyle\sum_{z\neq x}q_\delta(x,z), & \text{if } y=x.\end{cases} \tag{46}$$

Appendix D
Proof of Theorem 1

The proof of the original theorem [30] can be found in [31]. Using our notations, $\mathcal{C}(E,\varphi\circ U,q)=\mathcal{C}(E,U,q)$ follows directly from the definitions in Appendix A when φ is strictly increasing. Then, it suffices to show that $d_{\varphi\circ U}(\Gamma)<d_U(\Gamma)$ for any cycle $\Gamma\in\mathcal{C}(E,U,q)\backslash\{E\}$ such that $U(\Gamma)>U(E)$ and $H_U(\Gamma)>0$. Let $x=\arg\min_{z\in B(\Gamma)}U(z)$, we have $U(x)>U(\Gamma)$ and

$$d_{\varphi\circ U}(\Gamma)=\frac{\varphi\circ U(x)-\varphi\circ U(\Gamma)}{\varphi\circ U(\Gamma)-\varphi\circ U(E)}.$$

Since φ is strictly concave in an open interval $(\mathcal{A},\mathcal{B})$ covering the range of U, we have $\varphi(u_2)-\varphi(u_1)<\varphi'(u_1)(u_2-u_1)$ for all $u_1,\ u_2\in(\mathcal{A},\mathcal{B})$, $u_1\neq u_2$. Thus

$$\varphi\circ U(x)-\varphi\circ U(\Gamma)<\varphi'\circ U(\Gamma)\Big(U(x)-U(\Gamma)\Big)$$
$$\varphi\circ U(\Gamma)-\varphi\circ U(E)>\varphi'\circ U(\Gamma)\Big(U(\Gamma)-U(E)\Big)$$

such that

$$d_{\varphi\circ U}(\Gamma)<\frac{U(x)-U(\Gamma)}{U(\Gamma)-U(E)}=d_U(\Gamma).$$

Appendix E
Proof of Theorem 2

One has to verify that, under condition (19), $d_{\varphi_j\circ U}(\Gamma)<d_{\varphi_i\circ U}(\Gamma)$ for any cycle Γ such that

$$\max_{z\in E}U(z)>\min_{z\in B(\Gamma)}U(z)>U(\Gamma)>U(E).$$

Letting $\theta=U(\Gamma)$, $v=U(E)$ and $w=\min_{z\in B(\Gamma)}U(z)$, this brings us to study the sign of the function

$$\psi(\theta)=\frac{\varphi_i(w)-\varphi_i(\theta)}{\varphi_i(\theta)-\varphi_i(v)}-\frac{\varphi_j(w)-\varphi_j(\theta)}{\varphi_j(\theta)-\varphi_j(v)}$$

in (v,w) or, equivalently, the sign of

$$\tilde\psi(\theta)=(\varphi_i(w)-\varphi_i(\theta))(\varphi_j(\theta)-\varphi_j(v))-(\varphi_j(w)-\varphi_j(\theta))(\varphi_i(\theta)-\varphi_i(v))$$

which derivative writes

$$\tilde\psi'(\theta)=\varphi_j'(\theta)(\varphi_i(w)-\varphi_i(v))-\varphi_i'(\theta)(\varphi_j(w)-\varphi_j(v)).$$

From (19), we have $\varphi_i''(\theta)\varphi_j'(\theta)-\varphi_j''(\theta)\varphi_i'(\theta)>0$ for all $\theta\in(v,w)$, such that the function φ_i'/φ_j' is strictly monotonic increasing in (v,w). Using the generalized mean value theorem, we deduce that there exists an unique $\theta_0\in(v,w)$ with

$$\left(\frac{\varphi_i'}{\varphi_j'}\right)(\theta_0)=\frac{\varphi_i(w)-\varphi_i(v)}{\varphi_j(w)-\varphi_j(v)}.$$

Thus, $\tilde\psi'(\theta_0)=0$, $\tilde\psi'(\theta)>0$ for all $\theta\in(v,\theta_0)$ and $\tilde\psi'(\theta)<0$ for all $\theta\in(\theta_0,w)$. Since $\tilde\psi(v)=\tilde\psi(w)=0$, we have $\tilde\psi(\theta)>0$ and hence $\psi(\theta)>0$ for all $\theta\in(v,w)$. This completes the proof.

References

[1] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, 1984.

[2] J. Besag, "On the statistical analysis of dirty pictures (with discussion)," *J. Roy. Stat. Soc. B*, vol. 48, pp. 259–302, 1986.

[3] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 2024–2036, 1989.

[4] M. I. Sezan and A. M. Tekalp, "Survey of recent developments in digital image restoration," *Opt. Eng.*, vol. 29, pp. 393–404, 1990.

[5] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 367–383, 1992.

[6] A. J. Patti, M. K. Özkan, A. M. Tekalp, and M. I. Sezan, "New approaches for space-variant image restoration," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Minneapolis, MN, Apr. 1993, vol. V, pp. 261–264.

[7] M. K. Özkan, A. M. Tekalp, and M. I. Sezan, "POCS-based restoration of space-varying blurred images," *IEEE Trans. Image Processing*, vol. 3, pp. 450–454, 1994.

[8] S. Koch, H. Kaufman, and J. Biemond, "Restoration of spatially varying blurred images using multiple model-based extended Kalman filters," *IEEE Trans. Image Processing*, vol. 4, pp. 520–523, 1995.

[9] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, pp. 932–946, 1995.

[10] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Processing Mag.*, vol. 14, pp. 24–41, 1997.

[11] T. Poggio and V. Torre, "Ill-posed problems and regularization analysis in eary vision," Artif. Intell. Lab., A. I. Memo 773, Mass. Inst. Technol., Cambridge, 1984.

[12] M. Bertero, "Regularization methods for linear inverse problems," in *Inverse Problems*, G. Talenti, Ed. Berlin, Germany: Springer-Verlag, 1986, pp. 52–112.

[13] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing*, vol. 6, pp. 298–311, 1997.

[14] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, pp. 84–93, 1990.

[15] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs priors," *IEEE Trans. Med. Imag.*, vol. 9, pp. 439–446, 1990.

[16] C. Bouman and K. Sauer, "A generalized gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Processing*, vol. 2, pp. 296–310, 1993.

[17] R. L. Stevenson, B. E. Schmitz, and E. J. Delp, "Discontinuity preserving regularization of inverse visual problems," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, pp. 455–469, 1994.

[18] S. Geman and D. E. McClure, "Bayesian image analysis: An application to single photon emission tomography," in *Proc. Annu. Meeting Amer. Stat. Assoc.*, Las Vegas, NV, Aug. 1985, pp. 12–18.

[19] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.

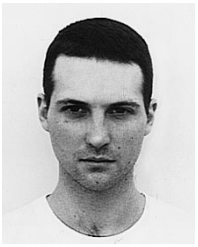
[20] D. Geiger and F. Girosi, "Parallel and deterministic algorithms from MRF's: Surface reconstruction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 401–412, 1991.

[21] A. H. Delaney and Y. Bresler, "Globally convergent edge-preserving regularized reconstruction: An application to limited-angle tomography," *IEEE Trans. Image Processing*, vol. 7, pp. 204–221, 1998.

[22] B. Hajek, "Cooling schedules for optimal annealing," *Math. Oper. Res.*, vol. 13, pp. 311–329, 1988.

[23] T. Chiang and Y. Chow, "On the convergence rate of annealing processes," *SIAM J. Contr. Optim.*, vol. 26, pp. 1455–1470, 1988.

[24] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, pp. 1087–1092, 1953.
[25] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.
[26] O. Catoni, "Rough large deviation estimates for simulated annealing: Application to exponential schedules," *Ann. Prob.*, vol. 20, pp. 1109–1146, 1992.
[27] C. Yang, *Stochastic Methods for Image Restoration*, Ph.D. dissertation, Univ. Mass., Amherst, Sept. 1991.
[28] D. Geman, G. Reynolds, and C. Yang, "Stochastic algorithms for restricted image spaces and experiments in deblurring," in *Markov Random Fields: Theory and Practice*, R. Chellappa and A. Jain, Eds. New York: Academic, 1993, pp. 39–68.
[29] C. Yang, "Efficient stochastic algorithms on locally bounded image space," *Comput. Vis., Graph., Image Process.: Graph. Models Image Process.*, vol. 55, pp. 494–506, 1993.
[30] R. Azencott, "Sequential simulated annealing: Speed of convergence and acceleration techniques," in *Simulated Annealing: Parallelization Techniques*, R. Azencott, Ed. New York: Wiley, 1992, pp. 1–10.
[31] _____, "A common large deviations mathematical framework for sequential annealing and parallel annealing," in *Simulated Annealing: Parallelization Techniques*, R. Azencott, Ed. New York: Wiley, 1992, pp. 11–23.
[32] T. Herbert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Med. Imag.*, vol. 8, pp. 194–202, 1989.
[33] S. Z. Li, "On discontinuity-adaptive smoothness priors in computer vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 576–586, 1995.
[34] O. Catoni, "Large deviations and cooling schedules for the simulated annealing algorithm (in French)," *C. R. Acad. Sci. Paris Sér. I Math.*, vol. 307, pp. 535–538, 1988.
[35] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*. Berlin, Germany: Springer-Verlag, 1984.
[36] P. J. M. van Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Practice*. Dordrecht, The Netherlands: Reidel, 1987.
[37] S. Kirkpatrick, C. D. Gellat, and M. P. Vecchi, "Optimization by simulated annealing," Res. Rep. RC 9355, IBM T. J. Watson Res. Ctr., Yorktown Heights, NY, 1982.
[38] M. Soumekh, "A system model and inversion for synthetic aperture radar imaging," *IEEE Trans. Image Processing*, vol. 1, pp. 64–76, 1992.
[39] P. de Herring, K. U. Simmer, E. Ochieng-Ogolla, and A. Wasiljeff, "A deconvolution algorithm for broadband synthetic aperture data processing," *IEEE J. Ocean. Eng.*, vol. 19, pp. 73–83, 1994.
[40] R. Bamler, "A comparison of range-doppler and wavenumber domain SAR focusing algorithms," *IEEE Trans. Geosci. Remote Sensing*, vol. 30, pp. 706–713, 1992.
[41] A. Moreira, J. Mittermayer, and R. Scheiber, "Extended chirp scaling algorithm for air- and space-borne SAR data processing in stripmap and ScanSAR imaging modes," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 1123–1136, 1996.
[42] C. Passman and H. Ermert, "*In vivo* imaging of the skin in the 100 MHz region using the synthetic aperture concept," in *Proc. IEEE Ultrasonics Symp.*, Seattle, WA, Nov. 1995, pp. 1287–1290.
[43] J. Ylitalo, "A fast ultrasonic aperture imaging method: Application to NDT," *Ultrasonics*, vol. 34, pp. 331–333, 1996.
[44] R. Benjamin, "Synthetic aperture antennas," *Microw. J.*, vol. 38, pp. 68–83, 1995.
[45] M. Hurn and C. Jennison, "An extension of Geman and Reynolds' approach to constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 657–662, 1996.
[46] A. Frigessi, P. di Stephano, C.-R. Hwang, and S.-J. Sheu, "Convergence rates of the Gibbs sampler, the Metropolis algorithm and other single-site updating dynamics," *J. R. Stat. Soc. B*, vol. 55, pp. 205–219, 1993.
[47] T.-S. Chiang and Y. Chow, "A comparison of simulated annealing of Gibbs sampler and Metropolis algorithms," in *Stochastic Models, Statistical Methods, and Algorithms in Image Analysis*, P. Barone, A. Frigessi, and M. Piccioni, Eds. Berlin, Germany: Springer-Verlag, 1992, pp. 117–124.



Marc C. Robini (M'99) received the M.Sc. degree in electrical engineering and the Ph.D. degree in signal and image processing from INSA-Lyon Scientific and Technical University, France, in 1993 and 1998, respectively.

He is currently "Maître de Conférences" at CREATIS (CNRS Research Unit, UMR 5515) and at the Electrical and Computer Engineering Department, INSA-Lyon. His research interests include inverse problems, multiresolution signal processing, and dynamic Monte Carlo methods.

Thierry Rastello received the electrical engineering diploma and the Ph.D degree in 1995 and 1998, respectively, from INSA-Lyon Scientific and Technical University, France.

His main research work deals with synthetic aperture ultrasonic processing applied to underwater and medical imaging systems.

Isabelle E. Magnin (M'85) received the ECAM engineering degree in 1977 and the Doctorat d'état ès Sciences degree in 1987 from INSA-Lyon Scientific and Technical University, France.

Since 1982, she has been a Researcher with the National Institute for Health and Medical Research (INSERM) at CREATIS, Villeurbanne, France. Her main interest concerns medical image processing, dynamic imaging, and 3-D reconstruction. She is also the Scientific manager of CREATIS, and Head of the Dynamic Imaging Group. She has more than 90 publications.