

A Stochastic Continuation Approach to Piecewise Constant Reconstruction

Marc C. Robini, *Member, IEEE*, Aimé Lachal, and Isabelle E. Magnin, *Member, IEEE*

Abstract—We address the problem of reconstructing a piecewise constant 3-D object from a few noisy 2-D line-integral projections. More generally, the theory developed here readily applies to the recovery of an ideal n -D signal ($n \geq 1$) from indirect measurements corrupted by noise. Stabilization of this ill-conditioned inverse problem is achieved with the Potts prior model, which leads to a challenging optimization task. To overcome this difficulty, we introduce a new class of hybrid algorithms that combines simulated annealing with deterministic continuation. We call this class of algorithms stochastic continuation (SC). We first prove that, under mild assumptions, SC inherits the finite-time convergence properties of generalized simulated annealing. Then, we show that SC can be successfully applied to our reconstruction problem. In addition, we look into the concave distortion acceleration method introduced for standard simulated annealing and we derive an explicit formula for choosing the free parameter of the cost function. Numerical experiments using both synthetic data and real radiographic testing data show that SC outperforms standard simulated annealing.

Index Terms—Continuation methods, inverse problems, signal reconstruction, simulated annealing.

I. INTRODUCTION

THE 3-D reconstruction problem discussed in this paper is to recover a piecewise constant distribution $x^0 = \{x_v^0; v \in V\}$ defined over a 3-D voxel lattice V from a few 2-D line-integral projections. Let $\{d^k; k = 1, \dots, K\}$, $d^k = \{d_s^k; s \in G^k\}$, be such a set of projections, where G^k denotes a rectangular pixel lattice. We consider the standard linear image formation model

$$d^k = \mathcal{H}^{(k)}(x^0) + \eta^k \quad (1)$$

where $\mathcal{H}^{(k)}$ is a linear map from $\mathbb{R}^{|V|}$ to $\mathbb{R}^{|G^k|}$ and where the noise process $\eta^k = \{\eta_s^k; s \in G^k\}$ consists of independent and identically distributed zero-mean Gaussian random variables with variance σ_k^2 . Although this type of ill-conditioned

inverse problem has been extensively studied for the purpose of full- and limited-angle tomography, voxel-based reconstruction from a very small number of projections has been only occasionally considered. The available approaches [1]–[3] lead to a class of nonconvex, usually multimodal, optimization problems solved by suboptimal procedures. The central theme of this work is to provide nearly optimal solutions to such problems in a stochastic framework. Before going further, let us mention that an alternative to voxel-based reconstruction is to estimate some geometric features of the object of interest by means of a deformable template model [4]–[6]. The number of unknown parameters is then drastically reduced, but their relationship with the data is no longer linear, which does not make the task easier. Besides, the deformable model approach is limited to the case of a single, compact homogeneous object.

To find an estimate of the true configuration x^0 , we seek a minimum of a cost function (also called energy function) $U : \Lambda^{|V|} \rightarrow \mathbb{R}$, $\Lambda \subset \mathbb{R}$, defined by

$$U(x) = \lambda \Phi(x) + \sum_{k=1}^K \sigma_k^{-2} \left\| \mathcal{H}^{(k)}(x) - d^k \right\|_2^2 \quad (2)$$

where the stabilizer $\Phi : \Lambda^{|V|} \rightarrow \mathbb{R}$ measures how well its argument matches our *a priori* knowledge about x^0 , and where the “smoothing parameter” $\lambda \in \mathbb{R}_+$ governs the tradeoff between stabilization and fidelity to the data. The justification for this choice lies within the regularization framework [7]–[9] or its bayesian interpretation [10]–[12]. We adopt the widespread prior model

$$\Phi(x) = \sum_{\mathcal{E} \subset V} \phi(\mathcal{D}\mathbf{e}(x)) \quad (3)$$

where \mathcal{D} is a first-order discrete derivative operator and $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is even, increasing in \mathbb{R}_+ and such that $\phi(0) = 0$. The summation is over pairs of voxels of the 26-nearest neighbor system: $\mathcal{D}\mathbf{e}(x) = 0$ if $|\mathcal{E}| \neq 2$ and for all $\{v, w\} \subset V$

$$\mathcal{D}_{\{v,w\}}(x) = \begin{cases} \frac{x_v - x_w}{\delta(v,w)}, & \text{if } 0 < \delta(v,w) \leq \sqrt{3} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

$$\delta(v,w) := ((i_v - i_w)^2 + (j_v - j_w)^2 + (k_v - k_w)^2)^{1/2}$$

where (i_v, j_v, k_v) is the set of integer indices labelling voxel v in lattice V . There are of course more intricate possibilities to define \mathcal{D} (see, e.g., [13] and [14] and references therein), but the extra computational load is not worth the effort. Indeed, there is theoretical evidence that the above defined operator is entirely appropriate for the reconstruction of piecewise constant objects [15].

Manuscript received March 22, 2007; revised June 1, 2007. This work was performed in accordance with the scientific trends of the national research group GDR ISIS of the CNRS. This paper was presented in part at the 2006 IEEE International Symposium on Signal Processing and Information Technology, Vancouver, BC, Canada, and at the 2006 IEEE Nuclear Science Symposium, San Diego, CA. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Thomas S. Denney, Jr.

M. C. Robini and I. E. Magnin are with the Center for Research and Applications in Image and Signal Processing (CREATIS), CNRS Research Unit UMR5520 and INSERM Research Unit U630, INSA Lyon, 69621 Villeurbanne Cedex, France (e-mail: marc.robini@creatis.insa-lyon.fr).

A. Lachal is with the Camille Jordan Institute (ICJ), CNRS Research Unit UMR5208, INSA Lyon, 69621 Villeurbanne cedex, France (e-mail: aime.lachal@insa-lyon.fr).

Digital Object Identifier 10.1109/TIP.2007.904975

Regularization functionals similar to (3) have been used in a variety of ill-posed problems (e.g., [1], [16]–[23]). The crucial choice is the function ϕ . A common example is $\phi(t) = t^2$, which corresponds to standard quadratic regularization. However, this choice is unsatisfactory because large gradients are unduly penalized, thus precluding the formation of sharp discontinuities. Many ϕ -functions have been proposed in the literature to overcome this limitation; they fall into two categories: convex [19], [24]–[27] and nonconvex [16]–[18], [28], [29]. Convex ϕ -functions ensure the convexity of U (assuming Λ is an interval) and reduce smoothing in the vicinity of discontinuities, while nonconvex ϕ -functions yield sharper edges at the expense of increased optimization difficulty. We consider the “0–1” function

$$\phi(t) = \begin{cases} 1, & \text{if } t \neq 0 \\ 0, & \text{if } t = 0 \end{cases} \quad (5)$$

which is particularly suited to piecewise constant reconstruction [15]. When Λ is finite, this choice of ϕ gives the Potts prior model, also called the multilevel logistic model [30]. The Potts model also arises in the context of minimum description length estimation [31], [32]; it measures somehow the length or the area of the imaginary boundary between regions of different intensities. Other forms of regularization functionals intended to favor the formation of piecewise constant objects involve the total variation seminorm [33], [34] and Besov norms [35]. Total variation regularization tends to smooth out boundaries [36] and reduces contrast in a way that is inversely proportional to scale and directly proportional to the smoothing parameter [37]. In Besov-type regularization, the prior term $\Phi(x)$ is a weighted ℓ^p -penalty on the wavelet coefficients of x , with $p \in [1, 2)$, which amounts to assuming that the wavelet coefficients are statistically independent and follow a generalized Gaussian distribution with shape parameter p [38], [39]. However, for piecewise smooth functions, the histogram distribution of the wavelet coefficients is very sharply peaked around zero, and, thus, the generalized Gaussian model is more accurate if p is in $(0, 1)$ and is rather close to zero. In this case, the difficulty of the associated optimization problem is comparable to that observed with the Potts model.

Faced with the challenge of minimizing the cost function U defined by (2)–(5), a first approach is to resort to deterministic continuation methods such as mean field annealing [40] or graduated nonconvexity (GNC) [41], which have reasonable computational load but produce suboptimal solutions. A second option is to use stochastic relaxation with annealing, which has the remarkable property to be asymptotically optimal when properly tuned [42]–[44] but is widely reported to have a very slow convergence rate. In fact, because of this embarrassing aspect, the stochastic approach is seldom if ever used for 3-D voxel reconstruction.

The first contribution of this paper is rather theoretical: by appealing to generalized simulated annealing (GSA) theory [45], [46] we show that it is possible to incorporate a continuation sequence into a simulated annealing framework to combine the best of both worlds. This leads to a class of annealing algorithms with time-dependent energy function we call stochastic continuation (SC) algorithms. The advantage over deterministic con-

tinuation is that SC inherits the finite-time convergence properties of GSA established in [46] and the advantage over standard simulated annealing is that appropriate control of the evolution of the difficulty of the explored energy landscape leads to faster practical convergence rates. In addition, we look into the concave distortion acceleration method introduced in [47], [48] for standard simulated annealing, which also provides an inexpensive potential speed-up for SC.

Our second contribution is more practical: we focus on the application of SC to 3-D reconstruction by considering the minimization of the cost function U defined by (2)–(5). This involves the construction of a suitable continuation sequence, the definition of the state space, the selection of the cooling schedule, and the peripheral issue of choosing a value for the free parameter λ . Regarding the last point, we borrow some ideas from [18] to prove a lower bound on λ for the true volume x^0 to be a coordinate-wise minimum of U with prescribed probability when x^0 belongs to a certain class of configurations. Experiments about 3-D reconstruction from both synthetic data and real radiographic testing data show the benefits of SC over standard annealing and validate the practical utility of the proposed method for selecting λ . In connection with other work of ours [49], some of the results we obtain also demonstrate the feasibility of reconstructing the 3-D shape of a defect from a few radiographs in the context of nondestructive inspection of thick metal components.

The paper is organized as follows. In Section II, we briefly review GSA theory and we provide our results about the convergence of SC. Section III is devoted to the application of SC to 3-D reconstruction together with the issue of selecting λ . Experimental results are presented in Section IV, followed by concluding remarks.

II. ANNEALING WITH TIME-DEPENDENT ENERGY FUNCTION

Throughout this section, we consider the problem of finding a global minimum of a real-valued function U defined on a general but finite state space E . We denote the ground state energy $\min_{x \in E} U(x)$ by U_{\min} and we let E_{\min} be the set of global minima of U , that is, $E_{\min} = \{x \in E | U(x) = U_{\min}\}$.

Let us start from standard simulated annealing (SA) algorithms (we refer to [50] for a comprehensive introduction). Standard SA operates on an “energy landscape” (E, U, q) defined by a symmetric and irreducible Markov matrix $q : E \times E \rightarrow [0, 1]$ which specifies the allowed moves for generating a new candidate solution from the current one. For any $\beta \in \mathbb{R}_+^*$, let $P_\beta : E \times E \rightarrow [0, 1]$ be the Markov matrix defined by

$$P_\beta(x, y) = \begin{cases} q(x, y) \exp\left(-\beta(U(y) - U(x))^+\right), & \text{if } y \neq x \\ 1 - \sum_{z \neq x} P_\beta(x, z), & \text{if } y = x \end{cases}$$

where $a^+ := \max\{a, 0\}$. Given a nondecreasing positive real sequence $(\beta_n)_{n \in \mathbb{N}^*}$ called the cooling schedule, a (Metropolis-type) SA algorithm on (E, U, q) is a discrete-time nonhomogeneous Markov chain $(X_n)_{n \in \mathbb{N}}$ with transitions $P(X_n = y | X_{n-1} = x) = P_{\beta_n}(x, y)$. It is well known [42]–[44] that, for suitably adjusted logarithmic cooling sched-

ules, this class of algorithms is asymptotically optimal in the sense that

$$\lim_{n \rightarrow \infty} \min_{x \in E} \mathbb{P}(X_n \in E_{\min} | X_0 = x) = 1.$$

However, these results are not of much practical interest since logarithmic schedules yield extremely slow convergence. More interestingly, the convergence measure

$$\mathcal{M}(n) = \max_{x \in E} \mathbb{P}(X_n \notin E_{\min} | X_0 = x)$$

cannot decrease faster than an upper bounded power of n^{-1} . Indeed, from [51]

$$\exists \alpha \in (1, +\infty), \quad \forall n \geq 1, \\ \alpha^{-1} n^{-1/D_U} \leq \inf_{\beta_n \geq \dots \geq \beta_1 > 0} \mathcal{M}(n) \leq \alpha n^{-1/D_U} \quad (6)$$

where the constant D_U , called the difficulty of the energy landscape, is the maximum ratio of the depth of nonglobal local minima to their energy level above the ground state energy. More specifically

$$D_U = \max_{x \in E \setminus E_{\min}} \min_{y \in E_{\min}} \frac{H_U(x, y) - U(x)}{U(x) - U_{\min}}$$

where the ‘‘communication height function’’ $H_U : E \times E \rightarrow \mathbb{R}$ is defined as follows:

$$H_U(x, y) = \min_{(\gamma_i)_{i=0}^m \in \Gamma_{xy}} \max_{0 \leq i \leq m} U(\gamma_i)$$

where Γ_{xy} is the set of q -admissible paths from x to y , that is, the set of paths $(\gamma_i)_{i=0}^m$ such that $\gamma_0 = x$, $\gamma_m = y$, and $q(\gamma_{i-1}, \gamma_i) > 0$ for all $i \in \{1, \dots, m\}$.

The close relationship between the convergence rate and the difficulty of the energy landscape suggests to speed up SA by replacing U by a sequence of functions $(U_n)_{n \in \mathbb{N}^*}$ converging pointwise to U and such that D_{U_n} increases with n . Intuitively, this should favor exploration of the state space and thus should reduce the influence of deep local basins of attraction. We have found that generalized simulated annealing (GSA) theory permits to derive useful convergence results for this variant of SA we call stochastic continuation (SC). Introduction to GSA is the subject of Section II-A and its application to SC is discussed in Section II-B.

A. Generalized Simulated Annealing

GSA theory was initially developed for studying parallelization of SA [52] but covers many stochastic optimization algorithms, as well. The processes of concern are each associated with a family $(Q_\beta : E \times E \rightarrow [0, 1])_{\beta \in \mathbb{R}_+^*}$ of Markov matrices satisfying a large deviation principle with speed β and rate function $V : E \times E \rightarrow \mathbb{R}_+ \cup \{+\infty\}$, namely

$$\exists a \in (1, +\infty), \quad \forall x, y \in E, \quad \forall \beta \in \mathbb{R}_+^* \\ a^{-1} e^{-\beta V(x, y)} \leq Q_\beta(x, y) \leq a e^{-\beta V(x, y)} \quad (7)$$

with the convention that $e^{-\beta(+\infty)} = 0$. The rate function V is assumed to be irreducible in the sense that for any $x, y \in E$, there exists a V -admissible path from x to y , that is, a path $(\gamma_i)_{i=0}^m$ such that $\gamma_0 = x$, $\gamma_m = y$, and $V(\gamma_{i-1}, \gamma_i) < +\infty$ for all $i \in \{1, \dots, m\}$.

Given such a family $(Q_\beta)_\beta$ and a cooling schedule $(\beta_n)_n$, a GSA algorithm is a Markov chain $(X_n)_{n \in \mathbb{N}}$ with transitions $\mathbb{P}(X_n = y | X_{n-1} = x) = Q_{\beta_n}(x, y)$. To see how this type of algorithm can be used to minimize $U : E \rightarrow \mathbb{R}$, we need to introduce the (generalized) communication height function $\tilde{H}_V : E \times E \rightarrow \mathbb{R}$ defined by

$$\tilde{H}_V(x, y) = \min_{(\gamma_i)_{i=0}^m \in \tilde{\Gamma}_{xy}} \max_{1 \leq i \leq m} (U(\gamma_{i-1}) + V(\gamma_{i-1}, \gamma_i))$$

where $\tilde{\Gamma}_{xy}$ denotes the set of V -admissible paths from x to y . Let π_β be the invariant probability measure of Q_β . Assuming further that V is such that \tilde{H}_V is symmetric, [46, Lemma 2.1] gives

$$\exists b \in \mathbb{R}_+^*, \quad \forall \beta \in \mathbb{R}_+^*, \quad \forall x \in E \\ b^{-1} e^{-\beta(U(x) - U_{\min})} \leq \pi_\beta(x) \leq b e^{-\beta(U(x) - U_{\min})}.$$

Consequently, as β goes to infinity, π_β tends to a limit distribution which gives positive mass to every element of E_{\min} and is zero elsewhere. It follows that, for sufficiently slowly increasing cooling schedules, we can expect the law of X_n to be close enough to π_{β_n} to achieve asymptotic optimality. As a matter of fact, it is proved in [52] that (6) holds with D_U replaced with

$$\tilde{D}_V = \max_{x \in E \setminus E_{\min}} \min_{y \in E_{\min}} \frac{\tilde{H}_V(x, y) - U(x)}{U(x) - U_{\min}}$$

and a theorem by Cot and Catoni [46, Th. 6.3] shows that there exists finite cooling schedules $(\beta_n)_{1 \leq n \leq N}$ such that

$$\ln(\mathcal{M}(N)) \underset{N \rightarrow \infty}{\sim} \ln(N^{-1/\tilde{D}_V}). \quad (8)$$

These schedules are piecewise constant exponential sequences of the form

$$\beta_n = \beta_{\min} \left(\frac{\beta_{\max}}{\beta_{\min}} \right)^{\frac{1}{\nu-1} (\lceil \frac{n}{N} \rceil - 1)} \quad (9)$$

where β_{\min} denotes the initial inverse temperature value and where the final inverse temperature value β_{\max} and the number ν of temperature steps are both functions of the horizon N .

B. Stochastic Continuation

Given a cooling schedule $(\beta_n)_n$, we call an SC algorithm a Markov chain $(X_n)_{n \in \mathbb{N}}$ with transitions $\mathbb{P}(X_n = y | X_{n-1} = x) = Q_{\beta_n}(x, y)$ defined by

$$Q_\beta(x, y) = \begin{cases} q(x, y) \exp\left(-\beta(U_\beta(y) - U_\beta(x))\right)^+, & \text{if } y \neq x \\ 1 - \sum_{z \neq x} Q_\beta(x, z), & \text{if } y = x \end{cases} \quad (10)$$

where q is a symmetric and irreducible Markov matrix on E and $(U_\beta)_{\beta \in \mathbb{R}_+^*}$ is a family of real valued functions on E —we use the

notation $(E, (U_\beta), q, (\beta_n))$ for short. Intuitively, a condition for this definition to make sense is that $\lim_{\beta \rightarrow \infty} U_\beta(x) = U(x)$ for all $x \in E$. Indeed, we have the following proposition whose proof is given in Appendix I.

Proposition 1: If E is finite and if $\lim_{\beta \rightarrow \infty} U_\beta(x) = U(x)$ for all $x \in E$, then there exists $\beta_0 \in \mathbb{R}_+$ such that for any $\beta \geq \beta_0$, the global minima of U_β belong to the set E_{\min} of global minima of U .

SC is a particular case of SA with time-dependent energy function, the behavior of which is studied in [53] and [54]. However, the results therein involve logarithmic schedules that are unusable in practice. The following theorem, whose proof is given in Appendix II, is more useful.

Theorem 1: Assume that E is finite and that

$$\diamond \forall x \in E, q(x, x) > 0 \quad (11)$$

$$\diamond \forall x \in E, \beta \in \mathbb{R}_+ \mapsto U_\beta(x) \text{ is a continuous map} \quad (12)$$

$$\diamond \forall x \in E, \sup_{\beta \in \mathbb{R}_+} \beta |U_\beta(x) - U(x)| < +\infty. \quad (13)$$

Then, the SC algorithm $(E, (U_\beta), q, (\beta_n))$ is a GSA algorithm with rate function

$$V_U(x, y) = \begin{cases} (U(y) - U(x))^+, & \text{if } q(x, y) > 0 \\ +\infty, & \text{otherwise} \end{cases}$$

and the associated communication height function \tilde{H}_{V_U} is equal to H_U , which is symmetric.

This result shows that, under mild assumptions, SC algorithms are GSA algorithms for which there exists cooling schedules of the form (9) such that $\mathcal{M}(N)$ is asymptotically equivalent to N^{-1/D_U} in the logarithmic scale (recall that, according to (6), $1/D_U$ is the optimal convergence speed exponent). Hence, the lower the value of D_U , the faster the convergence rate, and one can legitimately ask whether there are convenient ways to reduce the difficulty of the energy landscape without changing the set of solutions of the underlying minimization problem. To this end, the concave distortion idea proposed by Azencott [47], [48] for standard SA is potentially interesting, as Theorem 2 makes clear (see [23] for proof).

Theorem 2: Assume that E is finite and let I be an open interval covering the range of U . Then for any increasing, strictly concave, differentiable function $\psi : I \rightarrow \mathbb{R}$, the set of global minima of $\psi \circ U$ is the same as the set of global minima of U and $D_{\psi \circ U} < D_U$. Furthermore, if $\psi_1 : I \rightarrow \mathbb{R}$ and $\psi_2 : I \rightarrow \mathbb{R}$ are increasing, twice-differentiable functions such that

$$\left(-\frac{\psi_2''}{\psi_2'}\right)(u) < \left(-\frac{\psi_1''}{\psi_1'}\right)(u) \quad \text{for all } u \in I$$

then $D_{\psi_1 \circ U} < D_{\psi_2 \circ U}$.

It follows that for any suitable concave transform ψ , SC algorithms of the form $(E, (\psi \circ U_\beta), q, (\beta_n))$ are expected to converge faster than those of type $(E, (U_\beta), q, (\beta_n))$. Moreover, larger values of $-\psi''/\psi'$ should be encouraged. Different transforms are compared in [23] within the framework of image reconstruction by standard SA. In our experiments here, we will use the function

$$\psi(u) = \ln(u - \theta), \quad \theta \in \left[0, \inf_{(x, \beta) \in E \times \mathbb{R}_+} U_\beta(x)\right) \quad (14)$$

which can be implemented without difficulty and whose free parameter θ is easy to set.

III. APPLICATION TO 3-D RECONSTRUCTION

Recall from the introduction that 3-D reconstruction is formulated in terms of finding a global minimum of the energy function U defined by (2)–(5). For this purpose, the theory presented in Section II assumes a finite state space. Therefore, we take the domain E of U to be the set $\Lambda^{|\mathcal{V}|}$ (or a subset of it) with

$$\Lambda = \{al + b; l = 0, \dots, \mathcal{L} - 1\} \\ \mathcal{L} \in \mathbb{N} \setminus \{0, 1\}, \quad (a, b) \in \mathbb{R}_+^* \times \mathbb{R}. \quad (15)$$

Let us stress that this does not mean that the voxel values in the true configuration x^0 are known beforehand. In fact, minimizing U over $\Lambda^{|\mathcal{V}|}$ amounts to search for the best approximation of x^0 in $\Lambda^{|\mathcal{V}|}$ with a level of accuracy determined by the step size parameter a . Also note that when considering line-integral projection data such as in our experiments, b is naturally set to zero and it is easy to ensure that $\max_{v \in \mathcal{V}} x_v^0 \leq a(\mathcal{L} - 1)$ since we know at least the order of magnitude of the maximum possible voxel value.

Because the chief trouble comes from our choice of ϕ , it is natural to focus on SC algorithms in which the family $(U_\beta)_\beta$ is obtained by replacing ϕ in U with some function parameterized by β . More specifically, $\forall \beta \in \mathbb{R}_+, \forall x \in E$

$$U_\beta(x) = \lambda \sum_{\mathfrak{e} \subset \mathcal{V}} \phi_{\rho(\beta)}(\mathcal{D}\mathfrak{e}(x)) + \sum_{k=1}^K \sigma_k^{-2} \left\| \mathcal{H}^{(k)}(x) - d^k \right\|_2^2 \\ =: \mathcal{U}_{\rho(\beta)}(x) \quad (16)$$

where $\rho : \mathbb{R}_+ \rightarrow [0, 1)$ is continuous, increasing, and such that $\lim_{\beta \rightarrow \infty} \rho(\beta) = 1^-$, and where $(\phi_r : \mathbb{R} \rightarrow \mathbb{R})_{r \in [0, 1)}$ is a family of continuous functions satisfying $\lim_{r \rightarrow 1^-} \phi_r(t) = \phi(t)$ for all t .

Roughly speaking, the difficulty of the task of minimizing U_β should increase monotonically with β , which amounts to saying that the ‘‘concavity’’ of ϕ_r (i.e., $\sup_{t \in \mathbb{R}_+^*} |\phi_r''(t)|$ if ϕ_r is twice differentiable) should increase monotonically with r . It is also desirable that for any β , the relaxed energy U_β remains close to the target energy U . Hence, we propose to consider functions of the form

$$\phi_r(t) = \frac{|t/\Delta_r|}{(1 + |t/\Delta_r|)^r} \quad (17)$$

where the scale parameter Δ_r is determined so as to satisfy $\phi_r(c) = 1$ for some constant $c \in \mathbb{R}_+^*$ independent of r , that is, $\Delta_r = c/u_r$ where u_r is the unique fixed point of the contraction map $f_r : u \in \mathbb{R}_+ \mapsto (1 + u)^r \in \mathbb{R}_+$. Examples of such functions are given in Fig. 1 (in our experiments in Section IV, the constant c is set to the amplitude of the voxel value range, that is, $c = a(\mathcal{L} - 1)$).

To achieve the objective of minimizing U , the first step is to check whether the resulting family of relaxed energies defines appropriate SC algorithms. The following proposition, whose proof is given in Appendix III, shows that this is indeed the case.

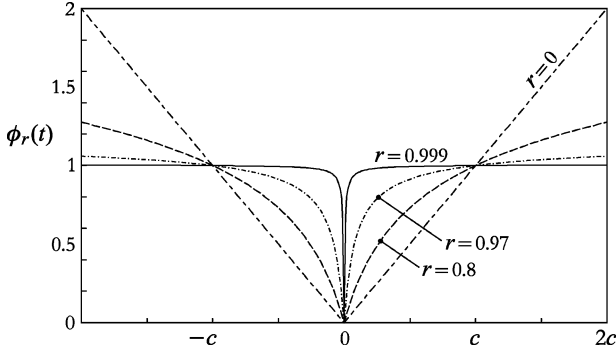


Fig. 1. Examples of functions in the family $(\phi_r)_r$, defined by (17).

Proposition 2: Let $(U_\beta)_\beta$ be defined by (16) and (17) and let U be the target energy defined by (2)–(5). Then for any $x \in E$, we have $\lim_{\beta \rightarrow \infty} U_\beta(x) = U(x)$.

The second step is to be able to guarantee that these SC algorithms belong to the class of GSA algorithms. From Theorem 1, it is sufficient to ensure that (11), (12) and (13) hold. Assumption (11) does not pose any problem; hence, Proposition 3 does the job (see Appendix IV for proof).

Proposition 3: Let $(U_\beta)_\beta$ be defined by (16) and (17). Then (12) holds unconditionally and (13) is satisfied if

$$\sup_{\beta \in \mathbb{R}_+} \beta(1 - \rho(\beta))^{1/2} < +\infty. \quad (18)$$

From now on, we only consider SC algorithms of the form $(E, (\omega \circ \mathcal{U}_{\rho(\beta)}), q, (\beta_n))$, where $E \subseteq \Lambda^{|\mathcal{V}|}$, $(\mathcal{U}_{\rho(\beta)})_\beta$ is defined by (16) and (17), $\omega = \text{id}_{\mathbb{R}}$ or $\omega = \psi$ (14), and q is irreducible, symmetric, and such that $q(x, x) > 0$ for all x . Considering Theorems 1 and 2 together with Propositions 2 and 3, condition (18) ensures that these algorithms are GSA algorithms that are “log-optimal” [in the sense of (8)] for properly tuned cooling schedule of the form (9). Nevertheless, efficient implementation requires wise choices regarding the communication graph (E, q) , the function ρ , the parameters of the cooling schedule, and the parameter θ in ψ . This is the subject of Section III-A, where we will see in particular that condition (18) can be relaxed somewhat. Section III-B is devoted to the remaining problem of finding a suitable value for the smoothing parameter λ .

A. Implementation of the Proposed Class of SC Algorithms

1) *Communication Graph:* If we set $E = \Lambda^{|\mathcal{V}|}$, a natural communication mechanism consists in generating a new candidate y by assigning a random value to a randomly selected voxel in the current configuration x , that is

$$q(x, y) = \begin{cases} (|\mathcal{V}\mathcal{L}|)^{-1}, & \text{if } |\{v \in \mathcal{V} | y_v \neq x_v\}| \leq 1 \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

Since we consider the reconstruction of piecewise homogeneous objects, we expect that each voxel v in the computed solution has a neighboring voxel whose value is close to that of v . Obviously, the communication process defined by (19) does not take this characteristic into account, which makes it less and less appropriate as the current configuration gets smoother (or equivalently, as β_n increases). As suggested in [55] in the context of 2-D reconstruction, this limitation can be overcome by

restricting $\Lambda^{|\mathcal{V}|}$ to a locally bounded space, that is, a set $\Omega_\ell^{|\mathcal{V}|}$ which consists of the elements x of $\Lambda^{|\mathcal{V}|}$ such that

$$\forall v \in \mathcal{V}, \quad \min_{w \in \mathcal{N}(v)} x_w - a\ell \leq x_v \leq \max_{w \in \mathcal{N}(v)} x_w + a\ell$$

where $\ell \in \{1, \dots, \mathcal{L} - 1\}$ and \mathcal{N} is a predefined neighborhood system on \mathcal{V} . For example, taking the six-nearest neighbor system together with a reasonably small level ℓ (e.g., $\ell \leq \mathcal{L}/10$) is in full accordance with the formation of piecewise homogeneous solutions. The key point is that the Markov matrix q_ℓ on $\Omega_\ell^{|\mathcal{V}|}$ defined by

$$q_\ell(x, y) = \begin{cases} (|\mathcal{V}\mathcal{L}\Omega_\ell^v(x)|)^{-1}, & \text{if } y \in \Omega_\ell^v(x) \\ 0, & \text{otherwise} \end{cases}$$

$$\Omega_\ell^v(x) := \left\{ y \in \Omega_\ell^{|\mathcal{V}|} | \forall w \in \mathcal{V} \setminus \{v\}, y_w = x_w \right\}$$

is symmetric and irreducible (we refer to [55] for proof and for implementation details). So, we set $(E, q) = (\Omega_\ell^{|\mathcal{V}|}, q_\ell)$.

2) *Choice of the Function ρ :* According to proposition 3, the function ρ should satisfy $1 - \rho(\beta) \in O(1/\beta^2)$. Still, it may be preferable to let ρ increase more slowly. As a matter of fact, condition (18) can be relaxed by imposing $\lim_{\beta \rightarrow \infty} \rho(\beta) = 1 - \varepsilon$ with $\varepsilon \in (0, 1)$, which amounts to consider a new target energy, namely $\mathcal{U}_{1-\varepsilon}$ as defined by (16). There are two arguments in support of this alteration. First, practical cooling schedules have finite horizon N , and, thus, computer simulations stop at $\mathcal{U}_{\rho(\beta_N)}$, where $\rho(\beta_N) < 1$. Second, since we consider a finite state space, we have the following result which is an immediate consequence of Proposition 1 (it suffices to take $0 < \varepsilon \leq 1 - \rho(\beta_0)$).

Proposition 4: If $\varepsilon \in (0, 1)$ is small enough, then the global minima of $\mathcal{U}_{1-\varepsilon}$ belong to the set of global minima of U .

Proposition 5 below gives a condition on ρ for SC algorithms with target energy $\mathcal{U}_{1-\varepsilon}$ to be GSA algorithms (see Appendix V for proof). It basically states that the rate at which ρ approaches its limit can be $O(1/\beta)$.

Proposition 5: Let $\varepsilon \in (0, 1)$ and let $(U_\beta)_\beta$ be defined by (16) and (17) with $\lim_{\beta \rightarrow \infty} \rho(\beta) = 1 - \varepsilon$. If

$$\sup_{\beta \in \mathbb{R}_+} \beta(1 - \varepsilon - \rho(\beta)) < +\infty$$

then (13) holds for U replaced by $\mathcal{U}_{1-\varepsilon}$.

This leads us naturally to choose ρ of the form

$$\rho(\beta) = a_0 \left(1 - \frac{a_1}{\beta^\kappa + a_2} \right) \quad (20)$$

with $\kappa \geq 1$, $a_0 = 1 - \varepsilon$, $a_1 > 0$ and $a_2 \geq 0$. In practice, given $\varepsilon' \in (0, 1)$, we will require that $\rho(\beta_N) = 1 - \varepsilon'$ at the end of the simulation of the annealing chain. Starting with $\rho(\beta_1) = 0$, this gives

$$a_0 = (1 - \varepsilon') \frac{(\beta_N)^\kappa + a_2}{(\beta_N)^\kappa - (\beta_1)^\kappa} \text{ and } a_1 = (\beta_1)^\kappa + a_2. \quad (21)$$

3) *Selection of the Cooling Schedule:* We consider cooling sequences $(\beta_n)_{1 \leq n \leq N}$ of type (9) which stem from the theory developed in [46]. Yet, this theory does not give any information regarding the choice of the cooling parameters. Typically, the horizon N is fixed by the amount of computing resources that

can be allocated. Also, experimental evidence suggests that the performance of the proposed class of GSA algorithms is not sensitive to the choice of the number of steps ν (we take $\nu = \lfloor N/(20|V|) \rfloor$). In fact, the issue at stake here is to find suitable values for the initial and final inverse temperatures $\beta_1 = \beta_{\min}$ and $\beta_N = \beta_{\max}$. We select these parameters in accordance with the following standard criteria.

- 1) Most candidate moves should be accepted at the beginning of the simulation of the annealing chain, that is, β_1 should be chosen so that $\sum_{z \in E \setminus \{x\}} Q_{\beta_1}(x, z)$ is close to one for all x .
- 2) By the end of the simulation of the annealing chain, most candidate moves that lead to an increase in energy should be rejected. More formally, β_N should be determined so that $\sum_{z \in E: U_{\beta_N}(z) > U_{\beta_N}(x)} Q_{\beta_N}(x, z)$ is close to zero for any x .

We refer to [23] for information about how to perform these tasks.

4) *Choice of the Parameter θ* : For the concave transform ψ (14) to be well defined, the parameter θ must be less than $\inf_{(x, \beta) \in E \times \mathbb{R}_+} U_{\beta}(x)$. Of course, we can always set $\theta = 0$, but Theorem 2 indicates that the closer θ is to the ground state energy $\min_{x \in E} U(x)$, the less difficult the target energy landscape. To account for this, we emphasize that if the true configuration x^0 represents a piecewise constant object, then we should have

$$\forall \beta \in \mathbb{R}_+, \quad \min_{x \in E} U_{\beta}(x) > \sum_{k=1}^K \sigma_k^{-2} \left\| \mathcal{H}^{(k)}(x^0) - d^k \right\|_2^2.$$

The right-hand side of this inequality is the data fidelity term evaluated at x^0 . According to (1) and by the law of large numbers, we have

$$\left\| \mathcal{H}^{(k)}(x^0) - d^k \right\|_2^2 = \|\eta^k\|_2^2 \approx |G^k| \sigma_k^2$$

for all k . Hence, we propose to set

$$\theta = \sum_{k=1}^K |G^k|. \quad (22)$$

B. Selection of the Smoothing Parameter

The parameter λ , which balances stabilization and fidelity to the data, is the only free parameter of the target energy U (the projection operator $\mathcal{H}^{(k)}$ and the noise standard deviation σ_k are given for all k). Two well-known approaches to selecting an appropriate value for λ are generalized cross-validation [56] and the L-curve method [57]. In our case, however, these methods cannot be clearly justified and would lead to serious computational difficulties. Other possibilities include the Chi-square choice [58] and maximum likelihood estimation [12], [16], and we could also imagine to define a prior distribution for λ and estimate the pair (x^0, λ) using a maximum a posteriori criterion. Yet, owing to the recalcitrant nature of U , these approaches would also have heavy computational requirements. Alternatively, following [18], we look for a condition on λ to ensure with a given probability that the true volume x^0 is a coordinate-wise minimum of U when it belongs to a relevant class

of configurations; this approach does not require the finite state space assumption, so we consider that U is defined on $\mathbb{R}^{|V|}$.

Let $l^v \in \mathbb{R}^{|V|}$ be the configuration that is zero everywhere except at voxel v , where it assumes the value l . We say that $x \in \mathbb{R}^{|V|}$ is a coordinate-wise minimum of U if there exists a finite set $I \subset \mathbb{R}$ such that $\forall v \in V, \forall l \in \mathbb{R}^* \setminus I, U(x + l^v) > U(x)$. In this case, the probability to obtain a lower energy state from x by assigning a random value in a closed interval to some voxel is zero; hence, the approach is particularly suited to single-site updating algorithms such as those considered here. Let \mathcal{E} be the event that x^0 is a coordinate-wise minimum of the random variable U , that is

$$\mathcal{E} = \{ \exists I \subset \mathbb{R}, |I| < \infty \text{ and } \forall v \in V, \forall l \in \mathbb{R}^* \setminus I, U(x^0 + l^v) > U(x^0) \}. \quad (23)$$

Then, our goal is to find a condition on λ such that $P(\mathcal{E}) \geq 1 - \varepsilon$, $\varepsilon \in (0, 1)$, when x^0 belongs to a certain prototype configuration class. More precisely, we study the case where x^0 is made up of two constant media separated by a plane interface, which amounts to assume that for all $v \in V$

$$x^0(i_v, j_v, k_v) = \begin{cases} L \in \mathbb{R}_+^*, & \text{if } k_v \geq k_0 \\ 0, & \text{otherwise} \end{cases} \quad (24)$$

where k_0 is a fixed integer. It is important to be clear that substantial discrepancies with the situation considered in [18] compels us to revisit a portion of this work, though we follow the same line of reasoning. These differences include the fact that projection operators are not translation-invariant and the fact that the prior model (3), (4) involves a 3-D neighborhood system together with a different ϕ -function. For our part, we have the following simple result whose proof is sketched in Appendix VI.

Proposition 6: Let $\varepsilon \in (0, 1)$ and assume that x^0 is of the form (24). Then, $P(\mathcal{E}) \geq 1 - \varepsilon$ provided that $\lambda \geq 2\vartheta^2/17$, where ϑ is defined by

$$\text{erf}(\vartheta) := \frac{2}{\sqrt{\pi}} \int_0^{\vartheta} e^{-t^2} dt = (1 - \varepsilon)^{1/|V|}.$$

Remarks:

- 1) The only purpose of the configuration class defined by (24) is to derive a theoretical lower bound on λ that is easy to use in practice (note in passing that this bound depends neither on L nor on k_0). We emphasize again that our application of SC to 3-D reconstruction does not make any assumption about the voxel values in the true configuration.
- 2) The proof of Proposition 6 is based on the worst-case scenario in which all voxels are assumed to be adjacent to some interface. Besides, estimating λ with great accuracy is unnecessary since values with the same order of magnitude give roughly similar reconstruction results. Therefore, ϑ can be defined by a much smaller number than $|V|$ and our experiments below confirm that setting $\vartheta = 3$ is sufficient.
- 3) Proposition 6 can be generalized to operators other than (4). Indeed, it is readily seen from the proof that if

$$\delta\Phi := \inf_{\substack{v \in V, \\ l \in \mathbb{R}^* \setminus \{-L, L\}}} \sum_{\mathfrak{E} \subset V} (\phi(\mathcal{D}_{\mathfrak{E}}(x^0 + l^v)) - \phi(\mathcal{D}_{\mathfrak{E}}(x^0)))$$

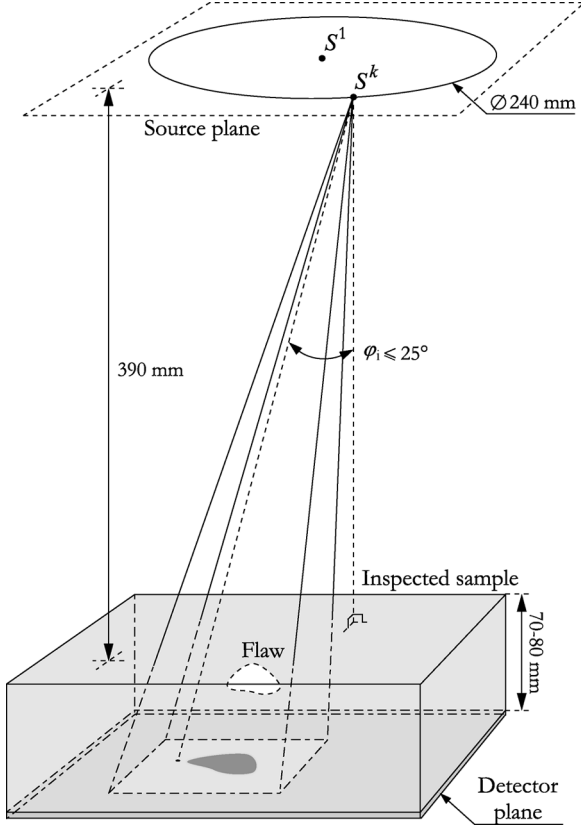


Fig. 2. Experimental data acquisition geometry (radiographic inspection of thick metal components).

is positive, then $P(\mathcal{E}) \geq 1 - \varepsilon$ provided $\lambda \geq 2\vartheta^2/\delta\Phi$.

- 4) It may be disturbing that the proposed parameter selection rule does not depend on noise characteristics. In fact, the regularization strength depends implicitly on the noise level since the data fidelity term in (2) is a quadratic distance weighted by the inverse of the noise variance. This dependence is particularly clear when each noise process η^k has the same standard deviation σ . In this case, minimizing (2) is equivalent to minimizing

$$\sigma^2 U(x) = \lambda \sigma^2 \Phi(x) + \sum_{k=1}^K \left\| \mathcal{H}^{(k)}(x) - d^k \right\|_2^2.$$

The actual smoothing parameter is, therefore, $\lambda \sigma^2$, and, hence, the regularization strength vanishes as σ goes to zero.

IV. EXPERIMENTAL RESULTS

Two sets of experiments are presented here. The first set concerns simulated projection data while the second set deals with the real radiographic testing data examined in [49]. Put very briefly (we refer to [49] for details), the radiographic testing data consists of noisy 2-D projections of some defects embedded in thick ferrous specimens. The associated raw radiographs were obtained in accordance with the geometry depicted in Fig. 2: the source positions S^1, \dots, S^K are in a plane parallel to the detector plane and S^2, \dots, S^K lie on a circle whose center S^1

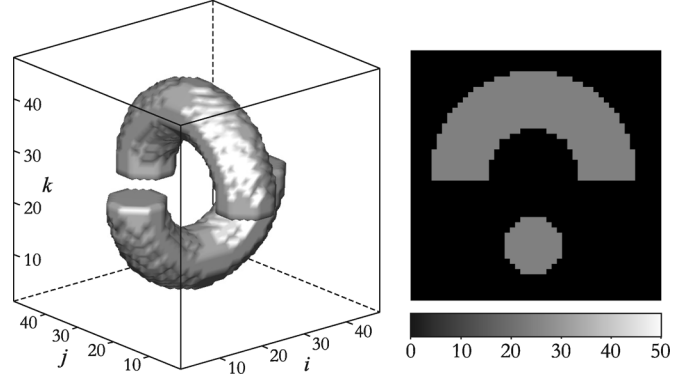


Fig. 3. Three-dimensional test object: isosurface representation and jk cross-section at $i = 24$.

and whose diameter are fixed by control requirements (in particular, the angle of incidence φ_i cannot exceed 25°).

In any situation, the objective is to minimize the target cost function U defined by (2)–(5), where λ is chosen to be equal to the lower bound given in proposition 6 with $\vartheta = 3$ (i.e., $\lambda \approx 1.06$). We investigate the behavior of the subclass of GSA algorithms introduced in Section III. More specifically, we consider SC algorithms of type

$$\mathcal{A}(\omega) := \left(\Omega_\ell^{|\mathcal{V}|}, (\omega \circ \mathcal{U}_{\rho(\beta)}), q_\ell, (\beta_n) \right) \quad (25)$$

with either $\omega = \text{id}_{\mathbb{R}}$ or $\omega = \psi$ (14), (22) and where the communication graph $(\Omega_\ell^{|\mathcal{V}|}, q_\ell)$ is defined by $\mathcal{L} = 51$ quantization levels together with the six-nearest neighbor system and $\ell = 5$. The cooling schedule $(\beta_n)_n$ is an exponential sequence of type (9) with $(N, \nu) = (10^4 |\mathcal{V}|, 500)$ and the function ρ that defines the family $(\mathcal{U}_{\rho(\beta)})_\beta$ (16), (17) is of the form (20) and (21) with $\varepsilon' = 10^{-3}$ and $(\kappa, a_2) = (1, 0)$ unless otherwise stated. The reconstruction results associated with $\mathcal{A}(\omega)$ are systematically compared with those produced by the standard¹ SA algorithm

$$\mathcal{A}_0 := \left(\Omega_\ell^{|\mathcal{V}|}, U, q_\ell, (\beta'_n) \right) \quad (26)$$

where $(\beta'_n)_n$ is a conventional exponential cooling schedule, that is, a sequence of type (9) with $\nu = N$. Both algorithms start from random configurations and their initial and final inverse temperature values are selected by means of the procedures proposed in [23].

A. Reconstruction From Synthetic Data

We consider the synthetic test object consisting of two half-tori whose discrete representation is shown in Fig. 3. The corresponding true configuration x^0 is defined over a 48^3 voxel lattice V and for each voxel v in V , $x_v^0 = a(\mathcal{L}-1)/2+b$ if v belongs to either half-torus, 0 otherwise. In this set of experiments, the set Λ (15) of admissible voxel values is defined by $(a, b) = (1, 0)$; so, $\Lambda = \{0, \dots, 50\}$ and $x_v^0 \in \{0, 25\}$. Five 128×128 projections of this numerical phantom were simulated according to model (1) using the same data acquisition geometry as in Fig. 2, where S^1 was placed vertically above the center of the object.

¹ \mathcal{A}_0 is not truly “standard”: Standard SA algorithms operate on the full state space, use logarithmic cooling schedules, and perform significantly worse than \mathcal{A}_0 [23].

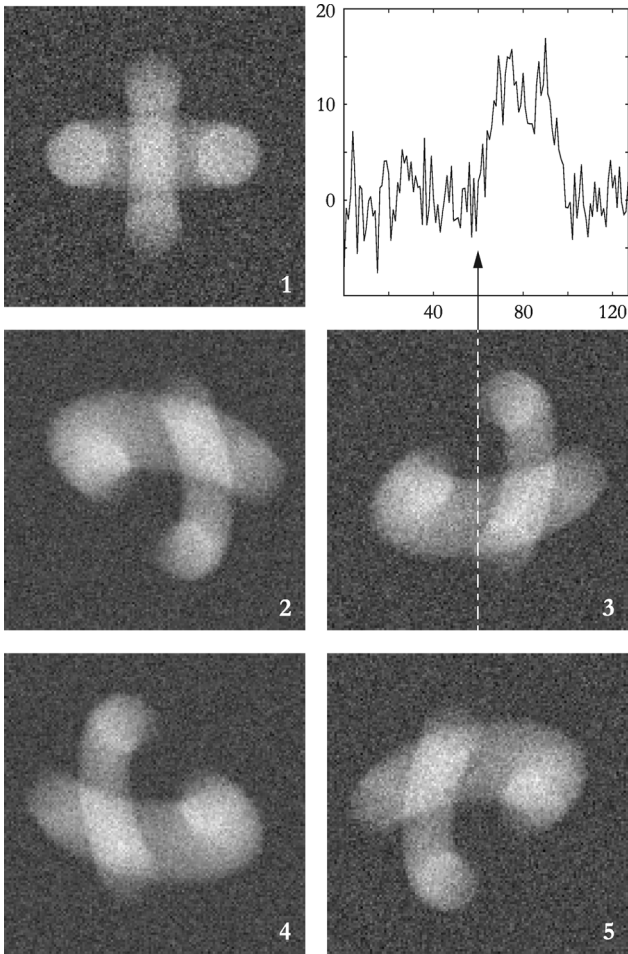


Fig. 4. Simulated line-integral projection data associated with the object shown in Fig. 3.

These projections are shown in Fig. 4. The variance σ_k^2 of the Gaussian white noise component η^k in projection number k was determined by specifying the decibel level of the signal-to-noise ratio: $(\text{SNR})_{\text{dB}} = 10 \log_{10}(\zeta_k^2/\sigma_k^2)$, where ζ_k^2 is the variance of the exact data $\mathcal{H}^{(k)}(x^0)$. We added 10-dB noise to projections number 2 and 4, 8-dB noise to projections number 3 and 5, and projection number 1 was corrupted by 6-dB noise. Note that the 2-D regions of support of each half-torus always overlap, and, thus, the existence of two separate connected 3-D regions cannot be predicted *a priori*.

Two error measures are used to assess reconstruction quality, namely the root mean square error (RMSE) and the percentage of misclassified voxels (MVP). Given an estimate x of x^0 , the latter is defined as follows:

$$\text{MVP} = 100|V|^{-1} \sum_{v \in V} \mathbb{1}_{\{|x_v - x_v^0| \geq 12.5\}}.$$

Fig. 5(a) shows the reconstruction achieved when running the standard SA algorithm \mathcal{A}_0 (26): a poor local minimum. The associated energy level is approximately $1.38 \cdot 10^5$ and the distance to the true configuration is emphasized by large error values: RMSE = 12.6, MVP = 58.8%. It should be stressed that increasing the length N of the annealing chain

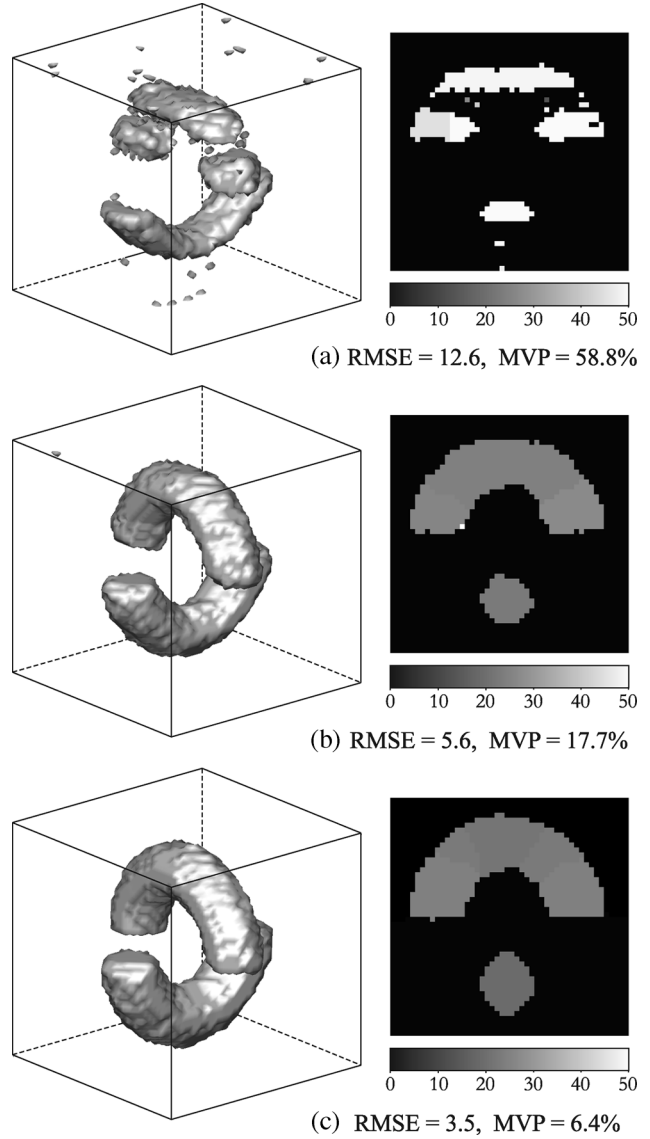


Fig. 5. Reconstruction from the data shown in Fig. 4: (a) optimization by standard SA; (b) optimization by SC; (c) optimization by SC with concave transform of the energy function.

does not yield any significant improvement. For example, setting $N = 5 \cdot 10^5|V|$ (i.e., allowing fifty times more iterations) produces an estimate with energy value of about $1.35 \cdot 10^5$, RMSE = 12.1 and MVP = 56.9%. The fact is that the energy landscape $(\Omega_\ell^{|V|}, U, q_\ell)$ is “perverse” in the sense that it contains local minima with deep basins of attraction.

By contrast, the SC algorithm $\mathcal{A}(\text{id}_{\mathbb{R}})$ (25) does not get stuck in the non relevant cycle where \mathcal{A}_0 ends up. The estimate produced by $\mathcal{A}(\text{id}_{\mathbb{R}})$ is shown in Fig. 5(b). It has only a slightly lower energy level ($1.35 \cdot 10^5$) than does the output of \mathcal{A}_0 , but is much closer to the true configuration, both qualitatively and quantitatively (RMSE = 5.6, MVP = 17.7%). Using SC with logarithmic transform of the energy function (i.e., algorithm $\mathcal{A}(\psi)$), we obtain the reconstruction shown in Fig. 5(c) which is even more faithful to the original object. As a matter of fact, $\mathcal{A}(\psi)$ offers substantial benefits over $\mathcal{A}(\text{id}_{\mathbb{R}})$, and, hence, a fortiori over \mathcal{A}_0 : the final energy level is $1.25 \cdot 10^5$, RMSE = 3.5

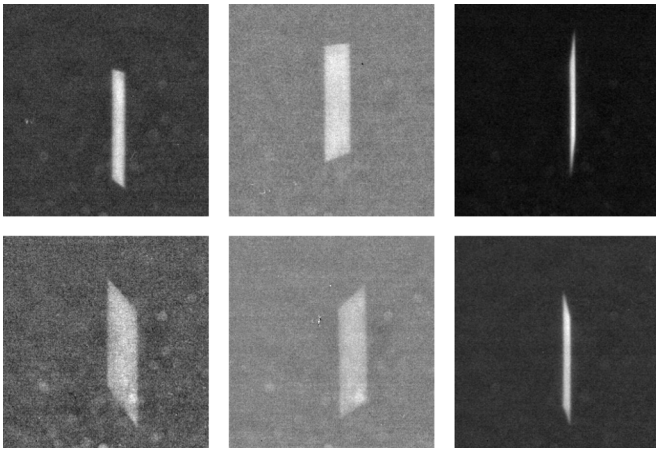


Fig. 6. Radiographs of an open notch of size $20 \times 0.3 \times 20$ mm embedded in a 7-cm-thick steel specimen.

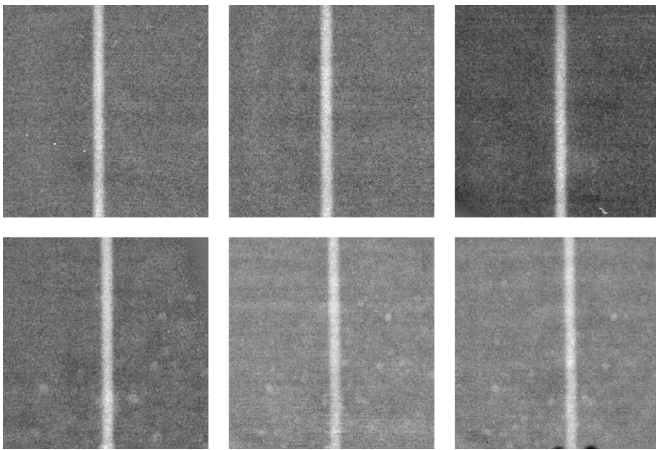


Fig. 7. Radiographs of a 2-mm-diameter hole drilled into an 8-cm-thick steel specimen.

and $MVP = 6.4\%$. Let us conclude this subsection with two remarks about the choice of the parameters of the function ρ (20), (21).

- 1) The results obtained by choosing κ greater than 1 are less spectacular, which justifies the additional effort to prove Propositions 4 and 5. As an example, setting $\kappa = 2$, $\mathcal{A}(\psi)$ gives a solution with an energy level of $1.37 \cdot 10^5$, $RMSE = 4.6$ and $MVP = 11.4\%$.
- 2) The quality of the reconstructions can be further improved by adjusting the parameter a_2 . For instance, by running $\mathcal{A}(\psi)$ with $\kappa = 1$ and $a_2 = 1$, we obtain an energy value of $1.24 \cdot 10^5$, $RMSE = 2.8$ and $MVP = 4.0\%$. However, the performance may deteriorate if a_2 is chosen too large and it is doubtful whether any theoretically sound selection rule could be drawn.

B. Reconstruction From Experimental Data

We now consider the reconstruction of real defects from the data sets depicted in Figs. 6 and 7. The radiographs in Fig. 6 show a vertically oriented electro-eroded notch of size $20 \times 0.3 \times 20$ mm embedded in a 7-cm-thick austenitic stainless steel casting. The defect revealed by the data in Fig. 7 is a 2-mm-diameter hole drilled into a similar steel specimen and whose

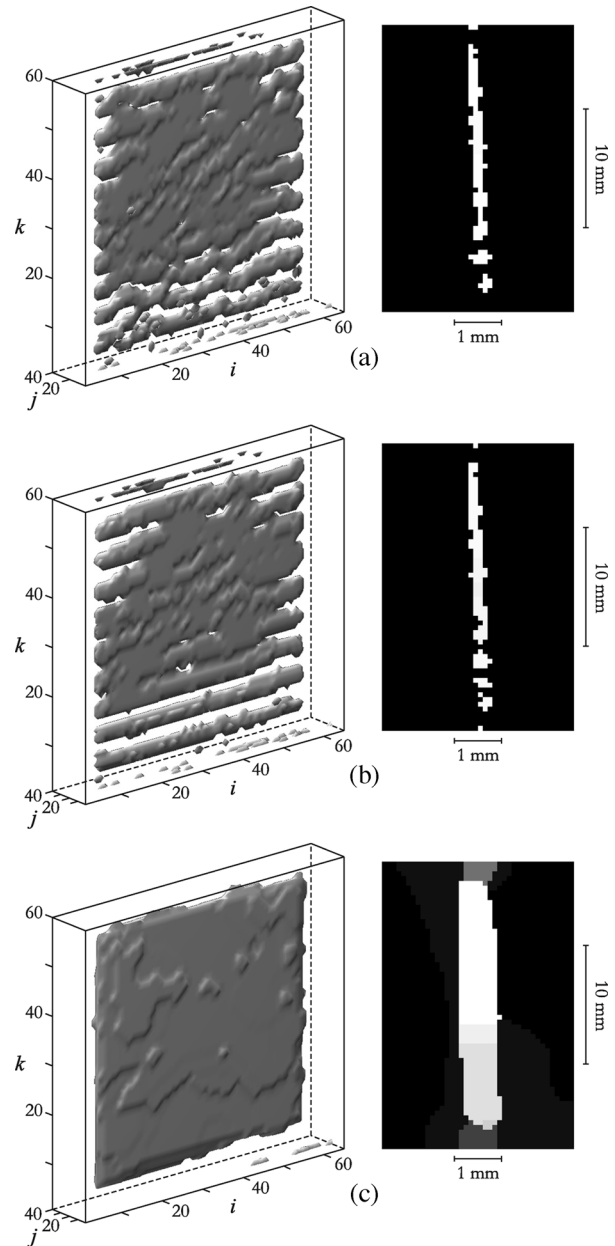


Fig. 8. Reconstruction from the data shown in Fig. 6 (isosurface representation and jk cross-section at $i = 32$): (a) standard SA with $10^4|V|$ iterations; (b) standard SA with $5 \cdot 10^5|V|$ iterations; (c) SC with $10^4|V|$ iterations.

axis is located at a depth of 10 mm on the source side. In both figures, the upper left projection corresponds to source position S^1 in Fig. 2. All projections are of size 222×215 and have spatial resolution of 0.2 mm.

1) *Reconstruction of the Notch:* We take the lattice V to be of size $65 \times 40 \times 60$ with each voxel representing a $0.4 \times 0.1 \times 0.4$ mm³ rectangular parallelepiped (the use of anisotropic voxels is dictated by the dimensions of the 3-D region of support computed from the smallest rectangles encompassing the defect signals in each projection). Fig. 8(a) displays the reconstruction obtained using the standard SA algorithm \mathcal{A}_0 . This estimate has an energy level of $2.44 \cdot 10^5$ and does not adequately capture the shape of the real defect. Again, increasing the length of the annealing chain does not help much: allowing

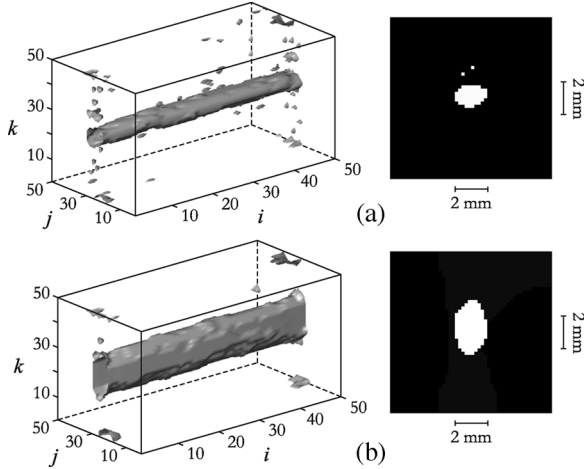


Fig. 9. Reconstruction from the data shown in Fig. 7 (isosurface representation and jk cross-section at $i = 25$): (a) optimization by standard SA; (b) optimization by SC.

fifty times more iterations leads to the result in Fig. 8(b) whose energy value is $2.40 \cdot 10^5$. The solution produced by the SC algorithm $\mathcal{A}(\psi)$ appears in Fig. 8(c). It has a markedly lower energy level ($2.05 \cdot 10^5$) than does the output of SA and it is also much closer to the notch geometry. The larger dimensions are accurately estimated, but the thickness of the reconstructed notch is 0.8 mm instead of 0.3 mm. This overestimation is essentially due to source positioning errors which can be up to 3 mm in each direction.

2) *Reconstruction of the Hole*: Here, the lattice V is 50^3 and each voxel is of size $0.4 \times 0.2 \times 0.2$ mm. The reconstructions achieved by running \mathcal{A}_0 and $\mathcal{A}(\psi)$ are respectively shown in Fig. 9(a) and (b). The associated energy levels ($1.36 \cdot 10^5$ and $1.31 \cdot 10^5$) confirm the superiority of SC over SA, although not as strikingly as in the previous example. It turns out that neither solution is entirely satisfactory: the radial section of the reconstructed hole is either compressed or stretched along the vertical direction. In fact, the 25° limit placed on the angle of incidence translates into increased unstability along the source-film direction, which is undeniably the main limitation attached to the particular acquisition geometry of the considered radiographic inspection problem.

V. CONCLUSION

Three-dimensional reconstruction with a Potts prior leads to a challenging optimization task. The two main approaches to tackle this problem, namely deterministic continuation methods and SA, have complementary advantages and disadvantages: continuation methods are computationally attractive but produce suboptimal solutions, whereas SA is asymptotically optimal but converges very slowly.

We devised a class of hybrid algorithms, SC algorithms, that are interesting in two respects. First, we showed that SC inherits the convergence properties of GSA and is thus more theoretically grounded than deterministic continuation methods. Second, controlling the evolution of the difficulty of

the energy landscape by a GNC-like continuation sequence can lead substantially better performance than SA. This makes SC potentially attractive for a wide range of signal processing applications.

In our application to piecewise constant 3-D reconstruction, we designed a specific SC algorithm that outperforms SA without requiring more computational efforts. We also provided a rigorously justified, yet simple and efficient, selection scheme for the stabilization parameter involved in the target cost function. The reconstruction results obtained from real data ultimately validate the image formation model as well as the processing protocol proposed in [49] in the context of radiographic inspection of thick metal components.

APPENDIX I

PROOF OF PROPOSITION 1

$$\begin{aligned} \text{Let } \xi \in E_{\min}. \text{ For any } x \in E \setminus E_{\min}, \text{ we have} \\ U_\beta(x) - U_\beta(\xi) &= (U_\beta(x) - U(x)) + (U(x) - U(\xi)) \\ &\quad - (U_\beta(\xi) - U(\xi)) \\ &\xrightarrow{\beta \rightarrow \infty} U(x) - U(\xi) > 0 \end{aligned}$$

and, thus, there exists $\beta_0(x) \in \mathbb{R}_+$ such that $U_\beta(x) > U_\beta(\xi)$ for all $\beta \geq \beta_0(x)$. Since E is assumed to be finite, we have

$$\beta_0 := \sup_{x \in E \setminus E_{\min}} \beta_0(x) < +\infty.$$

Consequently, for any $\beta \geq \beta_0$, the set $E \setminus E_{\min}$ is disjoint from the set of global minima of U_β .

APPENDIX II

PROOF OF THEOREM 1

We first have to prove (7) for $V = V_U$ (note that $V_U(x, y) < +\infty \iff q(x, y) > 0$ and thus V_U is irreducible). Recalling that E is a finite set, it suffices to show that for any $x, y \in E$, $Q_\beta(x, y) = 0$ if $q(x, y) = 0$ and

$$\begin{aligned} \exists a(x, y) \in (1, +\infty), \quad \forall \beta \in \mathbb{R}_+^* \\ (a(x, y))^{-1} \leq Q_\beta(x, y) e^{\beta V_U(x, y)} \leq a(x, y) \quad (27) \end{aligned}$$

if $q(x, y) > 0$. The case when $q(x, y) = 0$ is straightforward: we have $x \neq y$ by (11), and, hence, $Q_\beta(x, y) = 0$ by (10). Assume that $q(x, y) > 0$. If $y = x$, then $V_U(x, x) = 0$ and

$$Q_\beta(x, x) \geq 1 - \sum_{z \neq x} q(x, z) = q(x, x) > 0$$

so that (27) is satisfied. If $y \neq x$, the function $\beta \in \mathbb{R}_+ \mapsto Q_\beta(x, y)$ is continuous [by (12)] and positive; therefore, we simply need to verify that

$$\lim_{\beta \rightarrow \infty} Q_\beta(x, y) e^{\beta V_U(x, y)} > 0$$

or equivalently that

$$\lim_{\beta \rightarrow \infty} W_\beta(x, y) < +\infty$$

with $W_\beta(x, y) := \beta(U_\beta(y) - U_\beta(x))^+ - \beta(U(y) - U(x))^+$. If $U(y) < U(x)$, then for sufficiently large β , $U_\beta(y) < U_\beta(x)$ and thus $W_\beta(x, y) = 0$. If $U(y) = U(x)$, we have

$$\begin{aligned} W_\beta(x, y) &\leq \beta |U_\beta(y) - U_\beta(x)| \\ &\leq \beta |U_\beta(y) - U(y)| + \beta |U_\beta(x) - U(x)| \\ &\leq B(x) + B(y) \end{aligned}$$

where $B(z) := \sup_{\beta \in \mathbb{R}_+} \beta |U_\beta(z) - U(z)| < +\infty$ by (13). Finally, if $U(y) > U(x)$, then for sufficiently large β , $U_\beta(y) > U_\beta(x)$ and consequently $W_\beta(x, y) \leq B(x) + B(y)$. This completes the proof that $(E, (U_\beta), q, (\beta_n))$ is a GSA algorithm.

Let $x, y \in E$ and let $(\gamma_i)_{i=0}^m$ be a V_U -admissible path from x to y . For any $i \in \{1, \dots, m\}$, we have

$$\begin{aligned} U(\gamma_{i-1}) + V_U(\gamma_{i-1}, \gamma_i) &= U(\gamma_{i-1}) + (U(\gamma_i) - U(\gamma_{i-1}))^+ \\ &= \max\{U(\gamma_{i-1}), U(\gamma_i)\}. \end{aligned}$$

Therefore

$$\tilde{H}_{V_U} = \min_{(\gamma_i)_{i=0}^m \in \tilde{\Gamma}_{xy}} \max_{0 \leq i \leq m} U(\gamma_i).$$

Since $V_U(x, y) < +\infty \iff q(x, y) > 0$, we have $\tilde{\Gamma}_{xy} = \Gamma_{xy}$, and, hence, $\tilde{H}_{V_U} = H_U$. The symmetry of H_U follows directly from the symmetry of q .

APPENDIX III PROOF OF PROPOSITION 2

We need to show that for any $t \in \mathbb{R}$, $\lim_{\beta \rightarrow \infty} \phi_{\rho(\beta)}(t) = \phi(t)$ or equivalently

$$\lim_{r \rightarrow 1^-} \phi_r(t) = \begin{cases} 1, & \text{if } t \neq 0 \\ 0, & \text{if } t = 0. \end{cases}$$

The case when $t = 0$ is trivial. Let $t \in \mathbb{R}^*$ and let $r \in [0, 1)$. The derivative of $f_r, f'_r : u \in \mathbb{R}_+ \mapsto r/(1+u)^{1-r}$, is nonnegative and decreasing. Therefore

$$\begin{aligned} f'_r(u_r)u_r &\leq f_r(u_r) - f_r(0) \leq f'_r(0)u_r \\ &\iff \frac{ru_r}{1+u_r}u_r \leq u_r - 1 \leq ru_r \\ &\iff \left(\frac{1}{1-r}\right)^{1/2} \leq u_r \leq \frac{1}{1-r}. \end{aligned} \quad (28)$$

In particular, $\lim_{r \rightarrow 1^-} u_r = +\infty$ and, thus

$$\begin{aligned} \phi_r(t) &= \left(\frac{|t|u_r}{c}\right)^{1-r} \left(1 + \frac{c}{|t|u_r}\right)^{-r} \underset{r \rightarrow 1^-}{\sim} \left(\frac{|t|u_r}{c}\right)^{1-r} \\ &\underset{r \rightarrow 1^-}{\sim} u_r^{1-r} = \frac{(1+u_r)^r}{u_r^r} \underset{r \rightarrow 1^-}{\longrightarrow} 1. \end{aligned}$$

APPENDIX IV PROOF OF PROPOSITION 3

Let $x \in E$. To prove that the map $\beta \in \mathbb{R}_+ \mapsto U_\beta(x)$ is continuous, it suffices to show that for any $t \in \mathbb{R}$, $r \in [0, 1) \mapsto \phi_r(t)$ is continuous. This amounts to establishing the continuity of $r \in [0, 1) \mapsto u_r$, where u_r denotes the unique fixed point of $f_r : u \in \mathbb{R}_+ \mapsto (1+u)^r$. It is easy to check that for any

$r \in [0, 1)$, f_r is contractive with contraction factor r . Let $r_0 \in [0, 1)$. For all $r \in [0, 1)$, we have

$$\begin{aligned} |u_r - u_{r_0}| &= |f_r(u_r) - f_{r_0}(u_{r_0})| \\ &\leq |f_r(u_r) - f_r(u_{r_0})| + |f_r(u_{r_0}) - f_{r_0}(u_{r_0})| \\ &\leq r|u_r - u_{r_0}| + |f_r(u_{r_0}) - f_{r_0}(u_{r_0})| \end{aligned}$$

and, hence, $|u_r - u_{r_0}| \leq (1-r)^{-1}|f_r(u_{r_0}) - f_{r_0}(u_{r_0})|$. The right-hand side of this last inequality tends to zero as r tends to r_0 , which shows that $r \in [0, 1) \mapsto u_r$ is continuous at r_0 .

Now let us prove that (18) implies (13). For any $\beta \in \mathbb{R}_+$ and any $x \in E$

$$|U_\beta(x) - U(x)| \leq \lambda \sum_{\mathfrak{e} \subset V} |\phi_{\rho(\beta)}(\mathcal{D}\mathfrak{e}(x)) - \phi(\mathcal{D}\mathfrak{e}(x))|.$$

Consequently, since V is finite, a sufficient condition for (13) is that

$$\forall t \in \mathbb{R}_+^*, \quad \sup_{\beta \in \mathbb{R}_+} \beta |\phi_{\rho(\beta)}(t) - \phi(t)| < +\infty. \quad (29)$$

Set $\tau = t/c$ and $r = \rho(\beta)$. As $\lim_{r \rightarrow 1^-} u_r = +\infty$ [see (28)], we have

$$\begin{aligned} |\phi_{\rho(\beta)}(t) - \phi(t)| &= \left| \frac{\tau u_r}{(1 + \tau u_r)^r} - 1 \right| \\ &\underset{r \rightarrow 1^-}{\sim} \left| (\tau u_r)^{1-r} - \left(1 + \frac{1}{\tau u_r}\right)^r \right| \\ &= \left| \exp((1-r) \ln(\tau u_r)) - 1 - \frac{r}{\tau u_r} \right. \\ &\quad \left. + \underset{r \rightarrow 1^-}{o} \left(\frac{1}{u_r}\right) \right|. \end{aligned}$$

Using (28) again

$$\begin{cases} (1-r) \ln(\tau u_r) \leq (1-r) \underbrace{\left(\ln \tau + \ln \left(\frac{1}{1-r} \right) \right)}_{\underset{r \rightarrow 1^-}{\sim} ((1-r)^{1/2})} \\ \frac{1}{u_r} \leq (1-r)^{1/2} \end{cases}$$

and it follows that

$$\begin{aligned} |\phi_{\rho(\beta)}(t) - \phi(t)| &= \frac{r}{\tau u_r} + \underset{r \rightarrow 1^-}{o} \left((1-r)^{1/2} \right) \\ &= (1-r)^{1/2} \underset{r \rightarrow 1^-}{O}(1). \end{aligned}$$

Hence, (18) implies (29).

APPENDIX V PROOF OF PROPOSITION 5

A sufficient condition for (13) is that

$$\forall t \in \mathbb{R}_+^*, \quad \sup_{\beta \in \mathbb{R}_+} \beta |\phi_{\rho(\beta)}(t) - \phi_{1-\varepsilon}(t)| < +\infty.$$

Therefore, setting $r = \rho(\beta)$ and $\alpha = 1 - \varepsilon$, it suffices to show that

$$\forall t \in \mathbb{R}_+^*, \quad \phi_r(t) - \phi_\alpha(t) = \underset{r \rightarrow \alpha^-}{O}(\alpha - r).$$

For any $t > 0$, we have

$$\phi_r(t) - \phi_\alpha(t) = O_{r \rightarrow \alpha^-} (1 - g_r(\tau))$$

where $\tau := t/c$ and $g_r : \mathbb{R}_+ \rightarrow \mathbb{R}$ is defined by

$$\forall \tau \in \mathbb{R}_+, \quad g_r(\tau) = \frac{u_r(1 + \tau u_\alpha)^\alpha}{u_\alpha(1 + \tau u_r)^r}.$$

Let $r \in [0, \alpha]$. We have $g_r(0) = u_r/u_\alpha < 1$ and it is easy to show that g_r is increasing. Consequently, if $\tau \in [0, 1]$, then

$$0 \leq 1 - g_r(\tau) \leq 1 - g_r(0) = (u_\alpha - u_r)/u_\alpha \quad (30)$$

and if $\tau > 1$, then

$$\begin{aligned} 0 &\leq g_r(\tau) - 1 \\ &< \frac{(1 + \tau u_\alpha)^\alpha}{(1 + \tau u_r)^r} - 1 \\ &= (1 + \tau u_\alpha)^{\alpha-r} \left(1 + \tau \frac{u_\alpha - u_r}{1 + \tau u_r}\right)^r - 1 \\ &= \left(1 + O_{r \rightarrow \alpha^-}(\alpha - r)\right) \left(1 + O_{r \rightarrow \alpha^-}(u_\alpha - u_r)\right) - 1. \end{aligned} \quad (31)$$

As the derivative of f_r is nonnegative and decreasing, we have for any $r \in [0, \alpha]$

$$\begin{aligned} u_\alpha - u_r &= (f_\alpha(u_\alpha) - f_r(u_\alpha)) + (f_r(u_\alpha) - f_r(u_r)) \\ &\leq (f_\alpha(u_\alpha) - f_r(u_\alpha)) + f'_r(0)(u_\alpha - u_r) \end{aligned}$$

and thus, since $f'_r(0) = r$

$$\begin{aligned} u_\alpha - u_r &\leq (1 - r)^{-1} (f_\alpha(u_\alpha) - f_r(u_\alpha)) \\ &= O_{r \rightarrow \alpha^-} (1 - (1 + u_\alpha)^{r-\alpha}) \\ &= O_{r \rightarrow \alpha^-} (\alpha - r). \end{aligned}$$

Using this last result in (30) and (31) gives $1 - g_r(\tau) = O_{r \rightarrow \alpha^-}(\alpha - r)$ for all $\tau \in \mathbb{R}_+$, which ends the proof.

APPENDIX VI

PROOF OF PROPOSITION 6

The following lemma provides a lower bound for the energy difference $U(x^0 + l^v) - U(x^0)$. We then use this result to show that $\lambda \geq 2\vartheta^2/17$ is a sufficient condition for the event \mathcal{E} (23) to have probability at least $1 - \varepsilon$.

Lemma 1: Let $Z(\eta) = \{Z_v(\eta); v \in V\}$ be the zero-mean, Gaussian random process defined by

$$\forall v \in V, \quad Z_v(\eta) = \sum_{k=1}^K \sigma_k^{-2} \sum_{s \in G^k} \mathcal{H}_{v,s}^{(k)} \eta_s^k$$

where $\mathcal{H}_{v,s}^{(k)}$ denotes the contribution of voxel v to pixel s in the k th projection. Then for any $v \in V$ and any $l \in \mathbb{R}^* \setminus \{-L, L\}$

$$U(x^0 + l^v) - U(x^0) \geq l^2 \zeta_v - 2lZ_v(\eta) + 17\lambda$$

where $\zeta_v = \sum_{k=1}^K \sigma_k^{-2} \sum_{s \in G^k} (\mathcal{H}_{v,s}^{(k)})^2$.

Proof: Let $v \in V$ and let $l \in \mathbb{R}^* \setminus \{-L, L\}$. We have

$$U(x^0 + l^v) - U(x^0) = \delta_{v,l} \mathcal{H}(\eta) + \lambda \delta_{v,l} \Phi$$

with

$$\begin{aligned} \delta_{v,l} \mathcal{H}(\eta) &= \sum_k \sigma_k^{-2} \left(\left\| \mathcal{H}^{(k)}(x^0 + l^v) - d^k \right\|_2^2 \right. \\ &\quad \left. - \left\| \mathcal{H}^{(k)}(x^0) - d^k \right\|_2^2 \right) \\ &= \sum_k \sigma_k^{-2} \left(\left\| \mathcal{H}^{(k)}(l^v) - \eta^k \right\|_2^2 - \left\| \eta^k \right\|_2^2 \right) \\ &= \sum_k \sigma_k^{-2} \sum_{s \in G^k} \left(\left(\mathcal{H}^{(k)}(l^v)_s - \eta_s^k \right)^2 - (\eta_s^k)^2 \right) \\ &= \sum_k \sigma_k^{-2} \sum_{s \in G^k} \left(l^2 \left(\mathcal{H}_{v,s}^{(k)} \right)^2 - 2l \mathcal{H}_{v,s}^{(k)} \eta_s^k \right) \\ &= l^2 \zeta_v - 2lZ_v(\eta) \end{aligned}$$

and

$$\begin{aligned} \delta_{v,l} \Phi &= \sum_{\mathfrak{E} \subset V} \left(\phi(D\mathfrak{E}(x^0 + l^v)) - \phi(D\mathfrak{E}(x^0)) \right) \\ &= \sum_{w \in \mathcal{N}_{26}(v)} \left(\phi(x_v^0 + l - x_w^0) - \phi(x_v^0 - x_w^0) \right) \end{aligned}$$

where \mathcal{N}_{26} denotes the 26-nearest neighbor system on V . If v is not adjacent to the interface, then $x_w^0 = x_v^0$ for all $w \in \mathcal{N}_{26}(v)$ and thus $\delta_{v,l} \Phi = 0$. If v is adjacent to the interface, then there are 17 voxels $w_1, \dots, w_{17} \in \mathcal{N}_{26}(v)$ such that $x_{w_1}^0 = \dots = x_{w_{17}}^0 = x_v^0$ and 9 voxels $w'_1, \dots, w'_9 \in \mathcal{N}_{26}(v)$ such that $x_{w'_1}^0 = \dots = x_{w'_9}^0 = x_v^0 \pm L$, and it follows that $\delta_{v,l} \Phi = 17l$. This completes the proof of the lemma. \square

Set $R_v(\lambda) = \inf_{l \in \mathbb{R}_+^* \setminus \{L\}} (1/2)(l\zeta_v + 17\lambda/l)$ and note that if $|Z_v(\eta)| < R_v(\lambda)$, then $l^2 + 17\lambda > 2lZ_v(\eta)$ for all $l \in \mathbb{R}^* \setminus \{-L, L\}$. Therefore, from Lemma 1

$$\mathcal{E} \supset \{\forall v \in V, |Z_v(\eta)| < R_v(\lambda)\}$$

and, hence, $P(\mathcal{E}) \geq P(\bigcap_{v \in V} \{|Z_v(\eta)| < R_v(\lambda)\})$.

Then, since for any $v \in V$, $Z_v(\eta)$ is a normal random variable with zero mean and variance ζ_v

$$\begin{aligned} P(\mathcal{E}) &\geq \prod_{v \in V} P(|Z_v(\eta)| < R_v(\lambda)) \\ &= \prod_{v \in V} \left(\frac{2}{\sqrt{\pi}} \int_0^{R_v(\lambda)} \frac{1}{\sqrt{2\zeta_v}} \exp\left(-\frac{t^2}{2\zeta_v}\right) dt \right) \\ &= \prod_{v \in V} \operatorname{erf}\left(R_v(\lambda)/\sqrt{2\zeta_v}\right). \end{aligned}$$

It is easy to check that $R_v(\lambda) = \sqrt{17\lambda\zeta_v}$. Consequently

$$P(\mathcal{E}) \geq \left(\operatorname{erf}\left(\sqrt{17\lambda/2}\right) \right)^{|V|}$$

which shows that $\lambda \geq 2\vartheta^2/17$ is a sufficient condition to ensure that $P(\mathcal{E}) \geq 1 - \varepsilon$.

REFERENCES

- [1] K. Sauer, J. Sachs, Jr., and C. Klifa, "Bayesian estimation of 3-D objects from few radiographs," *IEEE Trans. Nucl. Sci.*, vol. 41, no. 5, pp. 1780–1790, Sep. 1994.

- [2] F. Reira, F. Peyrin, and J.-M. Dinten, "Three-dimensional regularized binary image reconstruction from three two-dimensional projections using a randomized ICM algorithm," *Int. J. Imag. Syst. Technol.*, vol. 9, no. 2–3, pp. 135–146, 1998.
- [3] B. Chalmond, F. Coldefy, and B. Lavyssière, "Tomographic reconstruction from non-calibrated noisy projections in non-destructive evaluation," *Inv. Probl.*, vol. 15, pp. 399–411, 1999.
- [4] B. Chalmond, F. Coldefy, and B. Lavyssière, "3D curve reconstruction from noisy projections," in *Wavelets, Images, and Surface Fitting*, P.-J. Laurent, A. L. Méhauté, and L. Schumaker, Eds. Natick, MA: A. K. Peters, 1994, pp. 113–119.
- [5] X. L. Battle, G. S. Cunningham, and K. M. Hanson, "Tomographic reconstruction using 3D deformable models," *Phys. Med. Biol.*, vol. 43, no. 4, pp. 983–990, 1998.
- [6] A. Mohammad-Djafari and C. Soussen, "Reconstruction of compact homogeneous 3-D objects from their projections," in *Discrete Tomography: Foundations, Algorithms and Applications*, G. T. Herman and A. Kuba, Eds. Boston, MA: Birkhäuser, 1999, pp. 317–342.
- [7] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Amer. Statist. Assoc.*, vol. 82, no. 397, pp. 76–89, 1987.
- [8] M. Bertero, T. A. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proc. IEEE*, vol. 76, no. 8, pp. 869–889, Aug. 1988.
- [9] H. W. Engl, M. Hanke, and A. Neubauer, *Regularisation of inverse problems*. Dordrecht, The Netherlands: Kluwer, 1996, vol. 375, Mathematics and its Applications.
- [10] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Jun. 1984.
- [11] J. Besag, "On the statistical analysis of dirty pictures (with discussion)," *J. Roy. Statist. Soc. B*, vol. 48, no. 3, pp. 259–302, 1986.
- [12] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 2024–2036, Dec. 1989.
- [13] L. Wang, T.-T. Wong, P. H. Heng, and J. C. Y. Cheng, "Template-matching approach to edge detection of volume data," in *Proc. Int. Workshop Medical Imaging and Augmented Reality*, Shatin, Hong Kong, Jun. 2001, pp. 286–291.
- [14] H. Farid and E. P. Simoncelli, "Differentiation of discrete multidimensional signals," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 496–508, Apr. 2004.
- [15] M. Nikolova, "Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares," *SIAM J. Multi-scale Model. Simul.*, vol. 4, no. 3, pp. 960–991, 2005.
- [16] S. Geman and D. E. McClure, "Statistical methods for tomographic image reconstruction," *Bull. Int. Statist. Inst.*, vol. 52, pp. 5–21, 1987.
- [17] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from poisson data using Gibbs priors," *IEEE Trans. Med. Imag.*, vol. 8, no. 2, pp. 194–202, Feb. 1989.
- [18] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.
- [19] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 296–310, Mar. 1993.
- [20] R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Trans. Image Process.*, vol. 3, no. 3, pp. 233–242, Mar. 1994.
- [21] A. H. Delaney and Y. Bresler, "Globally convergent edge-preserving regularized reconstruction: An application to limited-angle tomography," *IEEE Trans. Image Process.*, vol. 7, no. 2, pp. 204–221, Feb. 1998.
- [22] M. Nikolova, J. Idier, and A. Mohammad-Djafari, "Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF," *IEEE Trans. Image Process.*, vol. 7, no. 4, pp. 571–585, Apr. 1998.
- [23] M. C. Robini, T. Rastello, and I. E. Magnin, "Simulated annealing, acceleration techniques and image restoration," *IEEE Trans. Image Process.*, vol. 8, no. 10, pp. 1374–1387, Oct. 1999.
- [24] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, no. 1, pp. 84–93, Jan. 1990.
- [25] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs priors," *IEEE Trans. Med. Imag.*, vol. 9, no. 4, pp. 439–446, Apr. 1990.
- [26] R. L. Stevenson, B. E. Schmitz, and E. J. Delp, "Discontinuity preserving regularization of inverse visual problems," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 3, pp. 455–469, Mar. 1994.
- [27] S. Z. Li, "On discontinuity-adaptive smoothness priors in computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 6, pp. 576–586, Jun. 1995.
- [28] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.
- [29] M. Nikolova, "Markovian reconstruction using a GNC approach," *IEEE Trans. Image Process.*, vol. 8, no. 9, pp. 1204–1220, Sep. 1999.
- [30] S. Z. Li, *Markov Random Field Modeling in Computer Vision*. New York: Springer, 1995.
- [31] Y. G. Leclerc, "Constructing stable descriptions for image partitioning," *Int. J. Comput. Vis.*, vol. 3, pp. 73–102, 1989.
- [32] N. Saito, "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion," in *Wavelets in Geophysics*, E. Foufoula-Georgiou and P. Kumar, Eds. New York: Academic, 1994, pp. 299–324.
- [33] D. Dobson and F. Santosa, "Recovery of blocky images from noisy and blurred data," *SIAM J. Appl. Math.*, vol. 56, no. 4, pp. 1181–1198, 1996.
- [34] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, no. 1, pp. 227–238, 1996.
- [35] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [36] M. Nikolova, "Regularisation functions and estimators," in *Proc. IEEE Int. Conf. Image Processing*, Lausanne, Switzerland, Sep. 1996, vol. 2, pp. 457–460.
- [37] D. Strong and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization," *Inv. Probl.*, vol. 19, no. 6, pp. 165–187, 2003.
- [38] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized gaussian and complexity priors," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 909–919, May 1999.
- [39] M. Belge, M. E. Kilmer, and E. L. Miller, "Wavelet domain image restoration with adaptive edge-preserving regularization," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 597–608, Apr. 2000.
- [40] D. Geiger and F. Girosi, "Parallel and deterministic algorithms from MRF's: Surface reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 5, pp. 401–412, May 1991.
- [41] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.
- [42] B. Hajek, "Cooling schedules for optimal annealing," *Math. Oper. Res.*, vol. 13, no. 2, pp. 311–329, 1988.
- [43] T.-S. Chiang and Y. Chow, "On the convergence rate of annealing processes," *SIAM J. Control Optim.*, vol. 26, no. 6, pp. 1455–1470, 1988.
- [44] 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.
- [45] O. Catoni, "Simulated annealing algorithms and Markov chains with rare transitions," in *Séminaire de probabilités, XXXIII*. New York: Springer, 1999, vol. 1709, Lecture Notes in Math., pp. 69–119.
- [46] C. Cot and O. Catoni, "Piecewise constant triangular cooling schedules for generalized simulated annealing algorithms," *Ann. Appl. Probab.*, vol. 8, no. 2, pp. 375–396, 1998.
- [47] 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.
- [48] R. Azencott, "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.
- [49] M. C. Robini and I. E. Magnin, "Radiographic testing of anomalies in thick metal components: Fitting the standard line-integral model," *IEEE Trans. Nucl. Sci.*, to be published.
- [50] G. Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*. Berlin, Germany: Springer, 2003, vol. 27, Applications of mathematics.
- [51] O. Catoni, "Rough large deviation estimates for simulated annealing: Application to exponential schedules," *Ann. Probab.*, vol. 20, no. 3, pp. 1109–1146, 1992.
- [52] A. Trouvé, "Massive parallelization of simulated annealing," (in French) Ph.D. dissertation, Univ. Paris XI, Paris, France, Jan. 1993.
- [53] A. Frigerio and G. Grillo, "Simulated annealing with time-dependent energy function," *Math. Z.*, vol. 213, pp. 97–116, 1993.
- [54] M. Löwe, "Simulated annealing with time-dependent energy function via Sobolev inequalities," *Stoch. Process. Appl.*, vol. 63, no. 2, pp. 221–233, 1996.

- [55] C. Yang, "Efficient stochastic algorithms on locally bounded image space," *CVGIP: Graph. Models Image Process.*, vol. 55, no. 6, pp. 494–506, 1993.
- [56] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.
- [57] C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Rev.*, vol. 34, no. 4, pp. 561–580, 1992.
- [58] A. M. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterington, "A study of methods of choosing the smoothing parameter in image restoration by regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 4, pp. 326–339, Apr. 1991.



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

In 1999, he joined the INSA-Lyon where he is currently an Associate Professor at the "First Cycle" Department and a Research Associate at the Center for Research and Applications in Image and Signal Processing (CREATIS), CNRS Research Unit UMR5520 and INSERM Research Unit U630.

His current research interests include inverse problems, multiresolution image processing, and dynamic Monte Carlo methods.



Aimé Lachal received the M.Sc. degree in mathematics and the Ph.D. degree in probability from the Claude Bernard University of Lyon, France, in 1988 and 1992, respectively.

In 1990, he joined the INSA-Lyon, where he is currently an Associate Professor in the "First Cycle" Department and a Research Associate at the Camille Jordan Institute (ICJ), CNRS Research Unit UMR5208. His current research deals with stochastic analysis and, more specifically, with certain Markov processes including Brownian motion.



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

Since 1982, she has been a Researcher with the French National Institute for Health and Medical Research (INSERM). She is the Head of the Center for Research and Applications in Image and Signal Processing (CREATIS), CNRS Research Unit UMR5520 and INSERM Research Unit U630, Lyon. Her main interests concern medical image

processing, dynamic imaging, and inverse problems.