Markovian Reconstruction Using a GNC Approach

Mila Nikolova

***Abstract*— This paper is concerned with the reconstruction of images (or signals) from incomplete, noisy data, obtained at the output of an observation system. The solution is defined in *maximum a posteriori* (MAP) sense and it appears as the global minimum of an energy function joining a convex data-fidelity term and a Markovian prior energy. The sought images are composed of nearly homogeneous zones separated by edges and the prior term accounts for this knowledge. This term combines general nonconvex potential functions (PF's) which are applied to the differences between neighboring pixels.**

The resultant MAP energy generally exhibits numerous local minima. Calculating its local minimum, placed in the vicinity of the maximum likelihood estimate, is inexpensive but the resultant estimate is usually disappointing. Optimization using simulated annealing is practical only in restricted situations. Several deterministic suboptimal techniques approach the global minimum of special MAP energies, employed in the field of image denoising, at a reasonable numerical cost. The latter techniques are not directly applicable to general observation systems, nor to general Markovian prior energies.

This work is devoted to the generalization of one of them, the graduated nonconvexity (GNC) algorithm, in order to calculate nearly-optimal MAP solutions in a wide range of situations. In fact, GNC provides a solution by tracking a set of minima along a sequence of approximate energies, starting from a convex energy and progressing toward the original energy. In this paper, we develop a common method to derive efficient GNC-algorithms for the minimization of MAP energies which arise in the context of *any* observation system giving rise to a convex data-fidelity term and of Markov random field (MRF) energies involving *any* nonconvex and/or nonsmooth PF's. As a side-result, we propose how to construct pertinent initializations which allow us to obtain meaningful solutions using local minimization of these MAP energies.

Two numerical experiments—an image deblurring and an emission tomography reconstruction—illustrate the performance of the proposed technique.

***Index Terms*— Continuation methods, image reconstruction, inverse problems, MAP estimation, nonconvex optimization, regularization.**

I. Introduction

WHENEVER the *objects* we need to visualize—anomalies inside a medium, anatomical images, astronomical images, petroleum deposits, etc.—cannot be observed directly, the effects of some physical phenomena which characterize them are measured. The observation relation $\mathcal{A}(\boldsymbol{x}) \odot \boldsymbol{n} = \boldsymbol{y}$ models the link between the unknown object $\boldsymbol{x}$ (a signal, an image, a three-dimensional object) and the measurements $\boldsymbol{y}$, which are corrupted by a noise process $\boldsymbol{n}$ intervening by an operation denoted $\odot$. This relation is assumed discrete, while the notations used in this paper are mainly for images.

Recovering image $\boldsymbol{x}$ from data $\boldsymbol{y}$ amounts to invert the observation relation. The log-likelihood $-\mathcal{L}(\boldsymbol{x}) \propto \log p(\boldsymbol{y}|\boldsymbol{x})$ measures the fidelity of an image $\boldsymbol{x}$ to data $\boldsymbol{y}$. Function $\mathcal{L}$ is supposed *convex*—a numerical requirement met by many observation models [12], [27], [23]. In the popular *linear-Gaussian* (LG) model $\boldsymbol{y} = \mathcal{A}\boldsymbol{x} + \boldsymbol{n}$, $\boldsymbol{n}$ is a Gaussian noise field while observation operator $\mathcal{A}$ can represent a point spread function (PSF) accounting for optical blurring, a distortion wavelet in seismic imaging and nondestructive evaluation, a Radon transform in X-ray tomography, a Fourier transform in diffraction tomography, or it can be the identity operator $\mathcal{A} = \mathcal{I}$ if the observation is direct. Up to a constant factor,

$$(\text{LG}) \quad \mathcal{L}(\boldsymbol{x}) = ||\mathcal{A}\boldsymbol{x} - \boldsymbol{y}||^2. \tag{1}$$

In both emission and transmission computed tomography (ECT and TCT, respectively), the observed photon counts $\boldsymbol{y}$ have a Poissonian distribution [9], [18], [39]. Their mean is determined using projection operators $\{\boldsymbol{h}_j^T, j = 1, 2, \cdots, N\}$ and constant ρ. In both cases, $\mathcal{L}$ is convex (see Section VII-B) and it reads

$$\begin{aligned} (\text{ECT}) \quad & \mathcal{L}(\boldsymbol{x}) = \rho \boldsymbol{x}^T \sum_j \boldsymbol{h}_j - \sum_j y_j \ln(\boldsymbol{h}_j^T \boldsymbol{x}), \\ (\text{TCT}) \quad & \mathcal{L}(\boldsymbol{x}) = \sum_j \rho \exp(-\boldsymbol{h}_j^T \boldsymbol{x}) + y_j \boldsymbol{h}_j^T \boldsymbol{x}. \end{aligned} \tag{2}$$

Quite often, $\boldsymbol{x}$ is underdetermined in $\boldsymbol{y}$ and the inverse problem is *ill-posed* [43], [12]. The recovery of the unknown image must rely on both, data and prior knowledge about it. *Maximum a posteriori* (MAP) estimation, or regularization, is a flexible means allowing the recovery of objects $\hat{\boldsymbol{x}}$ which exhibit *a priori* expected features [17], [2], [12]. Estimate $\hat{\boldsymbol{x}}$ is the minimizer of an energy $\mathcal{E}$, which balances, using parameter β, closeness to data and confidence in prior guesses embodied in prior energy Φ

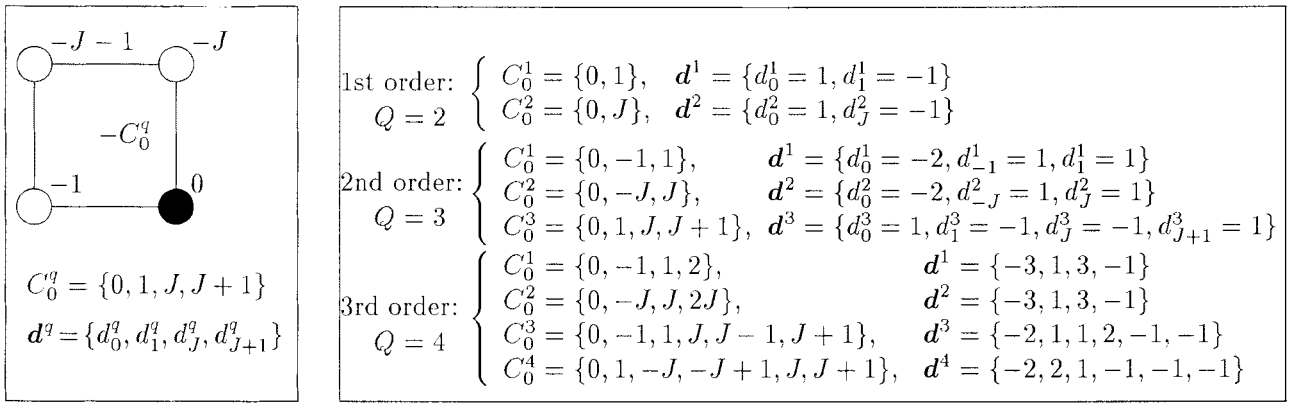
$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \mathcal{E}(\boldsymbol{x}), \quad \text{where} \quad \mathcal{E}(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}) + \beta\Phi(\boldsymbol{x}). \tag{3}$$

Following [2], [8], [18], [20], and [31], Φ is the energy of a Markov random field (MRF) of the form $\Phi(\boldsymbol{x}) = \Sigma_C \varphi(\boldsymbol{d}_C \boldsymbol{x})$. *Potential function* (PF) φ controls the interaction between the pixels involved in any clique C as a function of their *difference* $\boldsymbol{d}_C \boldsymbol{x}$. An important class of images are composed of homogeneous zones, separated by edges. Edges contain crucial visual information, so PF φ should favor their recovery.

Manuscript received January 7, 1997; revised June 19, 1998. This work was presented in part at the 1994 IEEE International Conference on Image Processing, Austin, TX. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stephen E. Reichenbach.

The author is with the UFR Mathématiques et Informatique—Université René Descartes, 75270 Paris Cédex 06, France (e-mail: nikolova@math-info.univ-paris5.fr).

Publisher Item Identifier S 1057-7149(99)06820-7.



(a) (b)

Fig. 1. Difference operators. (a) A pattern clique C_0^q; various difference patterns can be associated to it. (b) Difference patterns $\boldsymbol{d}^q$—their indices are elements of the relevant pattern cliques C_0^q.

Convex PF's are often used, mainly because energy $\mathcal{E}$ is then convex and its optimization is always tractable. Among them, generalized Gaussian, Huber and log-cosh [8], [24] PF's permit to partially preserve edges. Other such PF's are derived in [30]. Still, MAP energies defined using nonconvex PF's permit to recover truly homogeneous zones separated by sharp edges [18], [20], [31], [34], [45], [41]. The resultant $\mathcal{E}$ generally has numerous local minima and its global optimization is a difficult task. Several techniques were conceived to deal with *particular* observation systems and often with *particular* prior energies (Section III). Simulated annealing (SA) needs either an observation operator with an extremely restricted support[1] [19], [26], [45] or an LG model with $\mathcal{A}$ shift-invariant [21]. The iterated conditional modes (ICM) algorithm [1] finds a local minimum of $\mathcal{E}$ in the vicinity of the maxumum likelihood (ML) estimate, but the latter is not a pertinent initialization unless the inverse problem is well-posed and the noise is weak. Another suboptimal deterministic approximation of SA was proposed in [10]. Several deterministic techniques, proposed for the denoising and the segmentation of images, address the LG model with $\mathcal{A} = \mathcal{I}$ in (1) and a truncated quadratic or a "0-1" PF's (cf., Appendix A): the mean field annealing [3], [4], [14], the multigrid minimization [7], and the graduated non convexity (GNC) algorithm [5], [6], [29]. All these techniques are suboptimal in practice, while their extension to general settings is not straightforward. Reciprocally, practical use of nonconvex PF's is subject to the feasibility of the optimization stage.

The objective of this work is to enable *practical* use of various PF's in a broad range of situations. To this end, we focus on the GNC algorithm. Basically, GNC can be seen as a continuation technique [44] which substitutes a sequence of local minimizations along a sequence of approximate (*relaxed*) energies $\mathcal{E}_{r_k}$, where $(r_k)_{k=0}^K$ is an increasing positive relaxation sequence, for the global minimization of $\mathcal{E}$. The first relaxed energy $\mathcal{E}_{r_0}$ is convex and the last one fits $\mathcal{E}$. The main contribution of our paper consists in the generalization of the original GNC algorithm in order to compute MAP estimates corresponding to

(g1) MRF energies defined using *any* nonconvex, and possibly nonsmooth, PF φ (Section II);

(g2) *any* observation model giving rise to a convex data-fidelity term $\mathcal{L}$ (Section V).

The success of a GNC optimization is closely dependent on the pertinence of the approximation involved in the relaxed energies: this important problem is addressed in Section IV. Conditions for finding a convex initial relaxed energy are established for both the well-posed and the ill-posed cases (Section V). Considerations about the relaxation sequence and the running of GNC are exposed in Section VI.

A side result of this work is to propose a systematic way to calculate initializations for which a local minimization of $\mathcal{E}$ provides a meaningful solution. Such a strategy yields an *improved ICM* (Section VI-D) which can be applied in general situations, including the cases of ill-posed inverse problems.

Although mathematically suboptimal, generalized GNC performs efficiently in a broad range of situations and leads to nearly optimal solutions. Two numerical experiments—the deconvolution of an image and the inversion of ECT synthetic data—illustrate its performance, which is compared with the main alternative approaches (Section VII). Concluding remarks are given in Section VIII.

II. Prior Model

A. Markovian Models on Differences

Let $\boldsymbol{x}$ be an $(I \times J)$ image whose $M = IJ$ sites are arranged in lattice $\mathcal{S}$. Identifying $\mathbb{R}^{I \times J}$ and $\mathbb{R}^M$, we write henceforth $\boldsymbol{x} \in \mathbb{R}^M$; similarly, $\boldsymbol{y} \in \mathbb{R}^N$. We define over $\mathcal{S}$ several families $\mathcal{C}^1, \cdots, \mathcal{C}^Q$ of linked cliques: any q-clique $C_m^q \in \mathcal{C}^q$ is indexed by the pixel m to which it corresponds. The q-cliques result from the translation of a pattern clique $-C_0^q$ over $\mathcal{S}$, so that $C_m^q = m - C_0^q$ for $m \in \mathcal{S}^{(q)}$, where $\mathcal{S}^{(q)} = \{m \in \mathcal{S}: C_m^q \subseteq \mathcal{S}\}$ is the interior of $\mathcal{S}$ with respect to (w.r.t.) $\mathcal{C}^q$. Pattern $C_0^q = \{0, p_1^q, \cdots, p_{P^q-1}^q\}$ is an ordered set of positive and negative indices and includes 0 and P^q is its cardinality [Fig. 1(a)].

[1] The *support of an observation operator* is the set of the pixels of $\boldsymbol{x}$ which contribute to the obtention of any datum y_j.

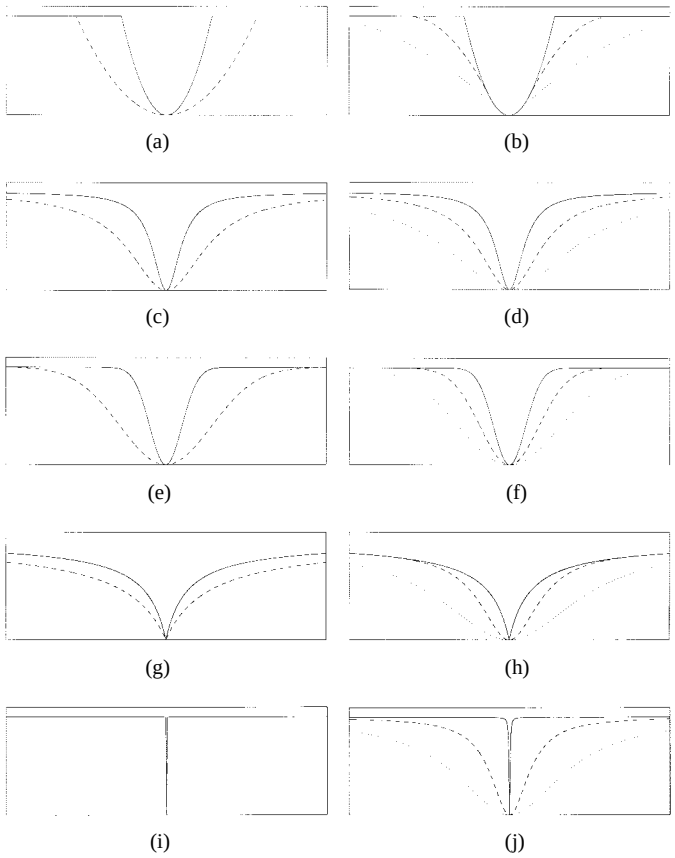


Fig. 2. PF's: originals (left column) and relaxed versions (right column). The PF's are plotted for two values of α. The relevant relaxed PF's are plotted for three values of r. (a)–(b) Truncated quadratic PF. (c)–(d) Lorentzian PF. (e)–(f) Gaussian PF. (g)–(h) Concave PF. (i)–(j) Kronecker PF.

We consider the popular class of MRF's which are defined over differences between neighbors [2], [8], [20]. A difference pattern $\boldsymbol{d}^q = \{d_j^q \neq 0 \text{ if } j \in C_0^q, d_j^q = 0 \text{ otherwise}, \Sigma_j\, d_j^q = 0\}$ is associated to each family $\mathcal{C}^q$. The q-differences are obtained by translating $\boldsymbol{d}^q$ over $\mathcal{S}^{(q)}$. The q-difference $\boldsymbol{d}_m^q \boldsymbol{x}$ relevant to m is

$$\boldsymbol{d}_m^q \boldsymbol{x} = (\boldsymbol{x} * \boldsymbol{d}^q)(m) = \sum_{i \in C_m^q} x_i\, d_{m-i}^q = \sum_{j \in C_0^q} x_{m-j}\, d_j^q. \quad (4)$$

The resultant set $(\boldsymbol{d}^q)_{q=1}^Q$ can in particular correspond to a finite differences operator [Fig. 1(b)].

The contribution of each q-clique to the prior energy Φ is weighted by a PF φ^q and a coefficient γ^q

$$\Phi(\boldsymbol{x}) = \sum_{q=1}^{Q} \gamma^q \Phi^q(\boldsymbol{x}), \quad \text{where}$$
$$\Phi^q(\boldsymbol{x}) = \sum_{m \in \mathcal{S}^{(q)}} \varphi^q(\boldsymbol{d}_m^q \boldsymbol{x}), \qquad q = 1, \cdots, Q. \quad (5)$$

Often, the same PF is applied to all cliques, $\varphi^q = \varphi$, for any q. When unambiguous, we write φ for φ^q.

B. Potential Functions Recovering Edges

Various types of edges, and of homogeneous regions, are observed in real-world images. They can be modeled using appropriate difference patterns $\boldsymbol{d}^q$, so that the edges are the locations of the largest differences $|\boldsymbol{d}_m^q \boldsymbol{x}|$, while the homogeneous zones correspond to differences which are close to zero. Thus, if $\boldsymbol{d}^q$ is first-order (e.g., $\boldsymbol{d}_m^1 \boldsymbol{x} = x_m - x_{m-1}$), the edges are jumps and the homogeneous zones are nearly constant; if $\boldsymbol{d}^q$ is second-order (cf., Fig. 1), the edges are creases while the homogeneous zones are nearly planar, etc. The edges contain essential visual information. Justly, the use of nonconvex PF's in (5) allows to obtain MAP estimates exhibiting homogeneous regions separated by sharp transitions [18], [20], [31], [34], [41], [45]. We focus on strictly nonconvex PF's φ, satisfying several general conditions:

- (f1) φ is symmetric, $\varphi(t) = \varphi(-t)$ and we set $\varphi(0) = 0$;
- (f2) φ is almost everywhere (a.e.) twice differentiable (i.e., $\mathcal{C}^1$-continuous);
- (f3) φ is monotone increasing on $[0, +\infty[$—its first derivative $\dot{\varphi}$, defined a.e., is $\dot{\varphi}(t) \geq 0$ for $t \geq 0$;
- (f4) there exists $u \geq 0$ such that $\dot{\varphi}$ decreases toward zero for $t > u$ and $\lim_{|t| \to \infty} \ddot{\varphi}(t) = 0$; in other words, φ has strictly concave parts while its concavity vanishes asymptoptically.

Thanks to (f4), the large differences in the original image are allowed to remain large in the estimate as well. PF φ reaches its maximum concavity $\eta < 0$ at $\pm T$ (which is $|T| \geq u$)

$$\eta = \inf_{t \in \mathbb{R}} \lim_{\epsilon \to 0} \frac{\varphi(t+\varepsilon) + \varphi(t-\varepsilon) - 2\varphi(t)}{\varepsilon^2}, \quad \eta < 0 \quad \text{and}$$
$$T = |\arg \eta|. \quad (6)$$

If φ is twice differentiable at T, then $\eta = \min_t \ddot{\varphi}(t) = \ddot{\varphi}(\pm T)$.

An estimation using (3) and (f1–f4) involves an *implicit* line process [20], [45], [34]. Energy $\mathcal{E}$ is generally multimodal and its local minima correspond to alternative configurations of the edges in the solution. More precisely, it can be remarked that these minima are separated by strictly nonconvex zones (and possibly by maxima) which contain differences of the form $|\boldsymbol{d}_m^q \boldsymbol{x}| = T^q$ (see Fig. 4). In consequence, differences which are $|\boldsymbol{d}_m^q \hat{\boldsymbol{x}}| \approx T^q$ appear quite rarely, or never, in an estimate. Pixels involved in a difference $|\boldsymbol{d}_m^q \hat{\boldsymbol{x}}| < T^q$ belong to the same q-region, while a q-edge sets their adherence to different regions if $|\boldsymbol{d}_m^q \hat{\boldsymbol{x}}| > T^q$. In practice, differences of $\hat{\boldsymbol{x}}$ are either $|\boldsymbol{d}_m^q \hat{\boldsymbol{x}}| \ll T^q$ or $|\boldsymbol{d}_m^q \hat{\boldsymbol{x}}| \gg T^q$ and T^q can be viewed as a threshold for the recovering of large q-differences, i.e., of q-edges.

Several among the most widely used nonconvex PF's are given in Appendix A. These are usually bounded and we set $\varphi(t) \leq 1$ for any t. The methodology presented in this paper is easy to extend to general nonconvex and/or nonsmooth PF's.

III. Optimization of Multimodal Criteria and GNC

Calculating an estimate $\hat{\boldsymbol{x}}$ of the form (3), (f1–f4) needs a M-dimensional global optimization of $\mathcal{E}$. However, the latter cannot be performed exactly except in several very special cases [5], [22]. Grossly speaking, two strategies can be followed for the minimization of $\mathcal{E}$ which are either to find a pertinent initialization and compute the nearest local minimum of $\mathcal{E}$, or to undertake a global minimization.

A. Direct Local Minimization

When a plausible initial guess about the sought image is easy to obtain, a local minimum of $\mathcal{E}$ in its vicinity can be substituted for the global minimum of $\mathcal{E}$. Such a strategy is simple to implement and can give rise to fast algorithms, but it is highly sensitive to the pertinence of the initialization.

The original ICM algorithm [1] takes the ML estimate as initialization, so it can be used only in the situations where the inverse problem is well-posed (i.e., $\boldsymbol{x}$ is well determined in $\boldsymbol{y}$) and if the noise is weak. Indeed, the ML solution of an ill-posed inverse problem is unstable in the sense that it exhibits an arbitrarily large amplification of the observation noise $\boldsymbol{n}$ [12]. As an alternative, a convolution back-projection of TCT data is used as starting point in [39]. We propose an improvement of ICM in Section VI-D.

B. Global Minimization Using Simulated Annealing

Optimization using simulated annealing (SA) is based on the fact that the distribution $p_\tau(\boldsymbol{x}|\boldsymbol{y}) = \exp[-\mathcal{E}(\boldsymbol{x})/\tau]/Z_\tau$, where τ means temperature, decreases toward zero as $\tau \to 0$ for objects $\boldsymbol{x}$ different from the global minima $\hat{\boldsymbol{x}}$. Family $p_\tau(\boldsymbol{x}|\boldsymbol{y})$ is processed in order to construct a Markov chain which converges to the set of the global minima of $\mathcal{E}$ as long as τ decreases slowly from an initial high value toward zero. Such algorithms were proposed initially for quantified objects $\boldsymbol{x}$ and later for real-valued images and signals. The Markov chain can be constructed in different ways. Some methods use stochastic gradient maximization of $p_\tau(\boldsymbol{x}|\boldsymbol{y})$ [11], [15], [13]. In others, $p_\tau(\boldsymbol{x}|\boldsymbol{y})$ is sampled using Metropolis dynamics [16], [26], [32] or using Gibbs dynamics [17], [19], [45]. The latter of them generates realizations of $p(x_i|x_j, j \neq i, \boldsymbol{y})$, so it is well suited if $\boldsymbol{x}|\boldsymbol{y}$ is a MRF with local interactions: but these become global if the support of the observation system is large, which makes this SA [17], [19], [26] impractical [21], [45], [35]. A recent form of SA [21] eludes this constraint for a LG model (1) where $\mathcal{A}$ is shift-invariant.

Asymptotical convergence toward the global minima of $\mathcal{E}$ can be ensured if 1) $\mathcal{E} \to \infty$ when $||\boldsymbol{x}|| \to \infty$ and $\mathcal{E}$ is smooth (for $\boldsymbol{x}$ real-valued), and 2) if temperature τ decreases according to a theoretical schedule. But $\mathcal{E}$, as given in (3) and (5), (f1–f4), usually does not satisfy 1) while schedule 2) is too slow to be followed in practice. Any numerical solution thus found is almost surely suboptimal [20], [22].

C. Global Minimization Using Deterministic Relaxation

An appealing approach for the global minimization of $\mathcal{E}$ is the following: a family of approximate (relaxed) energies is constructed by reducing the nonconvexity of $\mathcal{E}$. So is reached a relaxed energy having only a few local minima, one of which is then calculated. Starting from it, each relaxed energy is minimized locally, by descent in the current attraction valley, while the energies are progressing toward the original $\mathcal{E}$. Such techniques have been proposed for the optimization of several particular MAP energies.

In [7], $\mathcal{E}$ involves truncated quadratic PF's and a LG model (1) with $\mathcal{A} = \mathcal{I}$, and its nonconvexity is reduced by coarsening the grid of $\boldsymbol{x}$. The calculation cost at early stages is thus drastically reduced; however, it is difficult to check whether the initial energy has a unique minimum. In mean field annealing (MFA) [3], [4], [14], [41], an edge-process is replaced by its mean effect at varying temperatures. MFA depends on the initialization while its extention beyond the LG case with $\mathcal{A} = \mathcal{I}$ seems difficult. Another way to approximate $\mathcal{E}$ is to slightly deform the nonconvex zones of the PF—which underlies the GNC algorithm.

D. GNC Relaxation

The original GNC algorithm was proposed in [5] for the denoising and the segmentation of images and signals from data $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$ in the LG model, whereas $\mathcal{E}$ involves truncated quadratic PF's. A similar algorithm was used in [40]. This approach was applied to "0-1" PF's in [29] and later in [38].

The general GNC approach is sketched below. Consider a family of *relaxed energies* $\mathcal{E}_r$, dependent on a parameter $r \in]0, 1[$, and such that

- (e1) relaxed energies $\mathcal{E}_r$ are $\mathcal{C}^1$-continuous w.r.t. $\boldsymbol{x}$ and continuous w.r.t. r;
- (e2) the concavity of $\mathcal{E}_r$ is *relaxed* (i.e., reduced), and it vanishes monotonously when r decreases;
- (e3) moreover, there exists $r_0 > 0$ such that $\mathcal{E}_r$ is *convex* for any $r \leq r_0$ while $\lim_{r \to 1} \mathcal{E}_r(\boldsymbol{x}) = \mathcal{E}(\boldsymbol{x})$.

So, $\mathcal{E}_r$ has a unique minimum for $r \leq r_0$. When r increases, minima progressively appear in $\mathcal{E}_r$.

Consider an increasing relaxation sequence $(r_k)_{k=0}^K$ and the relevant relaxed energies $\mathcal{E}_{r_k}$, indexed by $k = 0, 1, 2, \cdots, K$. If $\mathcal{E}$ is $\mathcal{C}^1$-continuous, we take $r_K = 1$ and $\mathcal{E}_1 = \mathcal{E}$. Otherwise, the family $(\mathcal{E}_{r_k})_{k=0}^K$ cannot involve any element equal to $\mathcal{E}$; then we take $r_K \lessapprox 1$ such that $\mathcal{E}_{r_K} \approx \mathcal{E}$.

The GNC minimization starts from calculating the unique minimum $\boldsymbol{x}^{(r_0)}$ of $\mathcal{E}_{r_0}$. Afterwards, for each r_k a minimum of $\mathcal{E}_{r_k}$—an intermediate solution $\boldsymbol{x}^{(r_k)}$—is calculated by local descent in the vicinity $\mathcal{V}(\boldsymbol{x}^{(r_{k-1})})$ of the previously obtained minimum

$$\boldsymbol{x}^{(r_k)} = \arg \min_{\boldsymbol{x} \in \mathcal{V}(\boldsymbol{x}^{(r_{k-1})})} \mathcal{E}_{r_k}(\boldsymbol{x}) \quad \text{for } k = 1, 2, \cdots, K. \quad (7)$$

The ultimate solution $\boldsymbol{x}^{(r_K)}$, its closeness to the global minimum $\hat{\boldsymbol{x}}$, are determined by the sequence of relaxed energies. It is reasonable to require that for any r, relaxed energy $\mathcal{E}_r$ closely approximates the original energy $\mathcal{E}$.

IV. Relaxation of the Prior Energy

The nonconvexity of $\mathcal{E}$ is due to the strict nonconvexity of Φ. Hence, relaxing $\mathcal{E}$ amounts to relax Φ. Approximations Φ_r are obtained by relaxing the strictly nonconvex regions of φ:

$$\Phi_r(\boldsymbol{x}) = \sum_{q=1}^{Q} \gamma^q \Phi_r^q(\boldsymbol{x}), \quad \text{where}$$
$$\Phi_r^q(\boldsymbol{x}) = \sum_{m \in \mathcal{S}^{(q)}} \varphi_r^q(\boldsymbol{d}_m^q \boldsymbol{x}), \qquad q = 1, \cdots, Q. \quad (8)$$

Moreover, we require that approximations Φ_r remain as close as possible to the original Φ. The relaxed PF's φ_r^q should satisfy the following conditions (we drop the superscript q):

(r1) φ_r are $\mathcal{C}^1$-continuous w.r.t. t, while for any t fixed φ_r are continuous w.r.t. r;

(r2) φ_r should stray the least from φ for each r and $\lim_{r\to 1}\varphi_r = \varphi$ (and $\varphi_1 = \varphi$ if φ is $\mathcal{C}^1$);

(r3) the maximum concavity of φ_r, $\eta_r = \inf_{t\in\mathbb{R}} \ddot{\varphi}_r(t) < 0$, occuring at $\pm T_r$, i.e., $\eta_r = \inf \lim_{t\to\pm T_r} \ddot{\varphi}_r(\pm T_r)$, is required to increase continuously and strictly monotonously toward 0 as $r \to 0$.

A. Relaxed Potential Functions

There are numerous ways to get a relaxed PF φ_r, corresponding to a given original PF φ.

1) Dilation: When φ is $\mathcal{C}^1$-continuous, its second derivative exists a.e. and it is finite, so the minimal value of the latter may be controlled using a dilation of the original PF: $\varphi_r(t) = \varphi(rt)$ for any $r \in (0, 1]$. Thus are relaxed the Lorentzian and the Gaussian PF's (cf., Appendix A).

2) Regularization: An elegant manner to obtain φ_r is to regularize φ by a family of $\mathcal{C}^1$-continuous functions ϱ_r converging to the Dirac distribution δ, which reads: $\varphi_r = \varphi * \varrho_r$ and $\lim_{r\to 1}\varrho_r = \delta$. Kronecker PF is relaxed thus (cf., Appendix A). In general, it is difficult to find regularization functions ϱ_r leading to an explicit and easy convolution product φ_r.

3) Splines: Another way to construct φ_r is to fit splines in the vicinity of the points where φ is not differentiable and nonconvex; this is always feasible. This is the technique proposed in [5] for the relaxation of the truncated quadratic PF; we relax the concave PF in this way (cf. Appendix A).

Given a PF, various relaxed PF's can be envisaged, but these may induce the GNC optimization to converge toward slightly different solutions. It is hence important to use approximations which ensure a better convergence of the GNC.

Given a PF φ, suppose that we have two relaxed PF's, φ_r and ψ_r, which reach the same maximum concavity η_0 for $r = r_0$ and $r = \rho_0$, respectively. Among φ_r and ψ_r, the approximation which is "closer" to φ is clearly the better. The sense of this closeness needs a further precision.

Recall that the local minima of $\mathcal{E}$ correspond to different configurations of the edges in the solution. Thus, a GNC optimization is aimed at finding the edge configuration corresponding to the global minimum of $\mathcal{E}$. Observe that a relaxed PF φ_r is at the same time a nonconvex PF which permits to recover edges with respect to threshold T_r (cf., Section II-B). Hence, the choice of a particular relaxed PF should be based on its behavior w.r.t. the recovery of edges. On the one hand, thresholds T_r should decrease monotonously when r increases toward one. This naturally corresponds to an edge recovering which starts with the large edges and progresses toward the small edges. On the other hand, the closeness of T_r to the original T should be surveyed. Thus, the better amongst the PF's φ_r and ψ_r is the one whose initial threshold is closer to the original T: we check whether the threshold relevant to φ_{r_0} or the threshold relevant to ψ_{ρ_0} is closer to T. These ideas are further developed in the context of an example in Appendix B.

B. Maximum Concavity of the Relaxed Prior Energy

Now we focus on the maximum nonconvexity[2] reached by a relaxed prior energy.

Definition 1: The maximum concavity $\underline{\sigma_r} < 0$, reached by a strictly nonconvex function Φ_r, is

$$\underline{\sigma_r} = \inf_{\boldsymbol{v}} \inf_{\boldsymbol{x}} \frac{\boldsymbol{v}^T[D^2\Phi_r(\boldsymbol{x})]\boldsymbol{v}}{||\boldsymbol{v}||^2}, \qquad \boldsymbol{x} \in \mathbb{R}^M, \quad \boldsymbol{v} \in \mathbb{R}^M \tag{9}$$

where D^2 stands for second order differential operator. Hessian $D^2\Phi_r(\boldsymbol{x})$ is given in Appendix C.

Below, we establish an inferior bound of $\underline{\sigma_r}$ which is often reached in practice.

Proposition 1: Let $\overline{\nu}^q$ and $\overline{\nu}$ be defined as it follows:

$$\overline{\nu}^q = \max_{\boldsymbol{v}} \frac{\gamma^q \sum_{m\in\mathcal{S}^{(q)}} (\boldsymbol{d}_m^q\boldsymbol{v})^2}{||\boldsymbol{v}||^2} \quad \text{and}$$

$$\overline{\nu} = \max_{\boldsymbol{v}} \frac{\sum_{q=1}^{Q}\gamma^q \sum_{m\in\mathcal{S}^{(q)}} (\boldsymbol{d}_m^q\boldsymbol{v})^2}{||\boldsymbol{v}||^2}.$$

Let $\underline{\eta_r} = \min_q \eta_r^q$.

The maximum concavity $\underline{\sigma_r}$ of the relaxed prior energy Φ_r satisfies $\underline{\sigma_r} \geq \max\{\underline{\eta_r}\overline{\nu}, \Sigma_{q=1}^{Q}\eta_r^q\overline{\nu}^q\}$.

From the definition of $\underline{\sigma_r}$, we can write that $\underline{\sigma_r}||\boldsymbol{v}||^2 \geq \boldsymbol{v}^T[D^2\Phi_r(\boldsymbol{x})]\boldsymbol{v}$. The latter term can also be put into the form $\boldsymbol{v}^T[D^2\Phi_r(\boldsymbol{x})]\boldsymbol{v} = \Sigma_q\,\Sigma_n\,\gamma^q(\boldsymbol{d}_n^q\boldsymbol{v})^2\ddot{\varphi}_r^q(\boldsymbol{d}_n^q\boldsymbol{x})$—cf., Appendix A. Let now $T_r^q = T_r$ and $\eta_r^q = \eta_r$ for any q—which corresponds to the usual situation when $\varphi_r^q = \varphi_r$ for any q. Suppose moreover that $(\boldsymbol{d}_m^q)_{q=1}^Q$ admits images $\tilde{\boldsymbol{x}}$ satisfying $\boldsymbol{d}_m^q\tilde{\boldsymbol{x}} = T_r$ for any $m \in \mathcal{S}^{(q)}$ and for any q. For instance, if $(\boldsymbol{d}^q)_{q=1}^Q$ is a first-order difference operator (Fig. 1), this hold for the image with elements $\tilde{x}_{i,j} = T_r(i+j)$, while if the difference operator is of second order, we can take $\tilde{x}_{i,j} = T_r(i+j)^2/2$. In such a case, $\boldsymbol{v}^T[D^2\Phi_r(\boldsymbol{x})]\boldsymbol{v} = \eta_r\,\Sigma_q\,\Sigma_n\,\gamma^q(\boldsymbol{d}_n^q\boldsymbol{v})^2$, and hence the bound given in Proposition 1 is reached, $\underline{\sigma_r} = \underline{\eta_r}\overline{\nu}$.

Magnitudes $\overline{\nu}$ and $\overline{\nu}^q$ can be closely approximated by substituting a *circulant* convolution for the usual convolution in (4). The circulant operator $\tilde{\boldsymbol{d}}^q$ approximating $\boldsymbol{d}^q$, reads

$$\begin{aligned}
\tilde{d}^q{}_m &= \gamma^q d_m^q, \quad \text{if } m \in C_0^q \quad \text{and } m \geq 0,\\
\tilde{d}_{M+m}^q &= \gamma^q d_m^q, \quad \text{if } m \in C_0^q \quad \text{and } m < 0,\\
\tilde{d}_m^q &= 0, \quad \text{otherwise}
\end{aligned}$$

and $\Sigma_n(\boldsymbol{d}_n^q\boldsymbol{v})^2 \approx \Sigma_n(\tilde{\boldsymbol{d}}_n^q\boldsymbol{v})^2$ for any $\boldsymbol{v}$. Then, $\overline{\nu}^q = \max_i\{|\mu_i^q|^2\colon \exists\boldsymbol{v} \Rightarrow \tilde{\boldsymbol{d}}_n^q\boldsymbol{v} = \mu_i^q v_n, n \in \mathcal{S}\}$. Let $\boldsymbol{d}$ be given by $\tilde{d}_m = \Sigma_{q=1}^{Q}\tilde{d}_m^q$; similarly, $\overline{\nu} = \max_i\{|\mu_i|^2\colon \exists\boldsymbol{v} \Rightarrow \tilde{\boldsymbol{d}}_n\boldsymbol{v} = \mu_i v_n, n \in \mathcal{S}\}$. Eigenvalues $(\mu_k)_{k=1}^M$ and $(\mu_k^q)_{k=1}^M$ are the coefficients of the discrete Fourier transform of $\tilde{\boldsymbol{d}}$ and $\tilde{\boldsymbol{d}}^q$, respectively [25].

[2] A strictly nonconvex $\mathcal{C}^1$-continuous function necessarily involves zones where it is strictly concave.

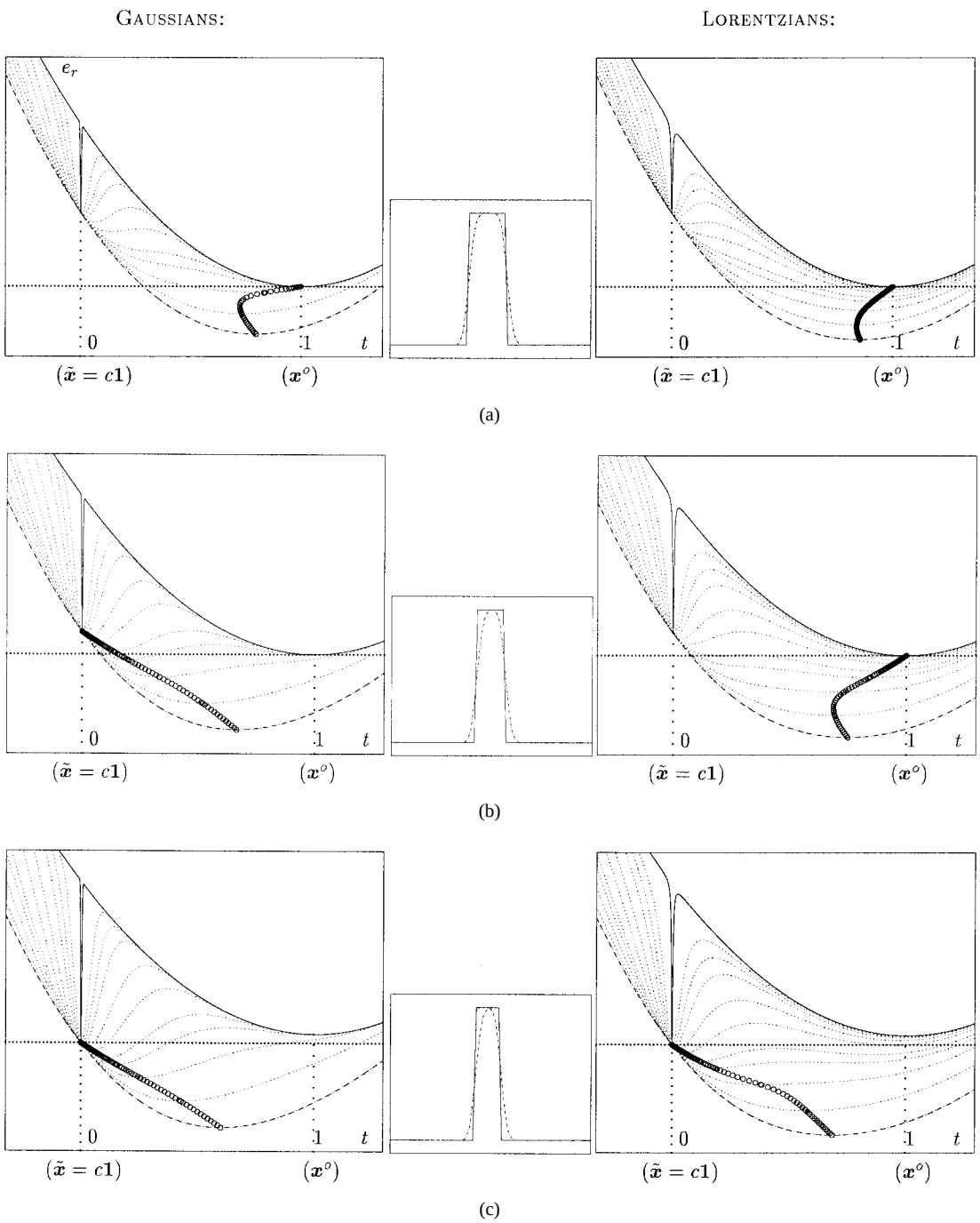


Fig. 3. Comparison of relaxations. Each row presents the reconstruction of the original signal $\boldsymbol{x}^o$ (——) from data $\boldsymbol{y} = \boldsymbol{a} * \boldsymbol{x}^o$ (-·-), plotted in the middle. The gates in (a)–(c) have a width of 12, 9, and 8 samples, respectively, while the length of $\boldsymbol{x}^o$ is 64. Energy $\mathcal{E}$ involves Kronecker PF's, defined over the first-order differences, while $\beta = 1.5$. This PF is relaxed using Gaussians (left column) and using Lorentzians (right column). Original energy $\mathcal{E}$ has two minima: the original gate $\boldsymbol{x}^o$ and a constant signal $\bar{\boldsymbol{x}} = c\mathbf{1}$. The figures depict $e_r(t) = \mathcal{E}_r[t\bar{\boldsymbol{x}} + (1-t)\boldsymbol{x}^o]$ for several values of r. The sequence of minima yielded by a slowly evolving GNC is marked with '$\circ$.' Explications are given in Section IX-B.

V. INITIAL CONVEXITY

GNC minimization starts from a relaxed energy $\mathcal{E}_{r_0}$ which has a unique minimum. A strictly nonconvex function can clearly be unimodal; but small changes in $\boldsymbol{y}$ can make it multimodal (see Fig. 4). Being unable to provide general necessary conditions for the unimodality of $\mathcal{E}_{r_0}$, we ensure instead its convexity for any $\boldsymbol{x}$. Following the example of [5], the initial convex approximation $\mathcal{E}_{r_0}$ is found by checking whether its Hessian is nonnegative definite for any $\boldsymbol{x} \in \mathbb{R}^M$:

find r such that

$$\boldsymbol{v}^T[D^2\mathcal{E}_r(\boldsymbol{x})]\boldsymbol{v} = \boldsymbol{v}^T[D^2\mathcal{L}(\boldsymbol{x})]\boldsymbol{v} + \beta\boldsymbol{v}^T[D^2\Phi_r(\boldsymbol{x})] \geq 0 \text{ for any } \boldsymbol{x} \text{ and } \boldsymbol{v}. \quad (10)$$

It is difficult to find the largest r_0 satisfying (10) for general $\mathcal{L}$ and Φ_r. But it is sufficient to take r_0 such that for $r < r_0$, the minimum convexity of the convex terms of $\mathcal{E}_r$ is larger than the maximum nonconvexity of the nonconvex terms of $\mathcal{E}_r$.

Definition 2: The minimum convexity $\underline{\mu} > 0$, reached by a convex function $\mathcal{L}$, is

$$\underline{\mu} = \inf_{\boldsymbol{v}} \inf_{\boldsymbol{x}} \frac{\boldsymbol{v}^T [D^2 \mathcal{L}(\boldsymbol{x})] \boldsymbol{v}}{||\boldsymbol{v}||^2}, \qquad \boldsymbol{x} \in \mathbb{R}^M, \qquad \boldsymbol{v} \in \mathbb{R}^M. \tag{11}$$

The Hessian corresponding to the LG model (1) is $D^2\mathcal{L}(\boldsymbol{x}) = 2\mathcal{A}^T\mathcal{A}$ so $\underline{\mu} = 2\lambda_{\min}$, where $\lambda_{\min}$ is the smallest eigenvalue of $\mathcal{A}^T\mathcal{A}$. In ECT and TCT, $\mathcal{L}$ is convex with $\underline{\mu} = 0$ (cf., Section VII-B).

The conjunction of (9), (10), Proposition 1 and Definition 2 yields

$$\begin{aligned} \boldsymbol{v}^T [D^2 \mathcal{E}_r(\boldsymbol{x})] \boldsymbol{v} &\geq \underline{\mu} ||\boldsymbol{v}||^2 + \beta \underline{\sigma_r} ||\boldsymbol{v}||^2 \\ &\geq \left(\underline{\mu} + \beta \max \left\{ \underline{\eta_r} \overline{\nu}, \sum_{q=1}^{Q} \eta_r^q \overline{\nu}^q \right\} \right) ||\boldsymbol{v}||^2 \\ &\qquad \text{for any } \boldsymbol{x} \text{ and } \boldsymbol{v}. \end{aligned} \tag{12}$$

If $\mathcal{E}_r$ is strictly nonconvex, the right-hand side (r.h.s.) of (12) is negative for r close to one; moreover, there exists $\boldsymbol{x}$ for which the left hand side (l.h.s.) of (12) is negative as well. Thanks to (r3), the sum in the r.h.s. of (12) increases from a negative value to zero as long as r is decreased to zero.

A. Strictly Convex Data-Fidelity Term

Consider first the resolution of a well-posed inverse problem, which means that $\mu > 0$. In this case, there exists r_0 for which the r.h.s. of (12) becomes positive.

Proposition 2: Let $\underline{\mu} > 0$ and $\mathcal{E}$ strictly nonconvex; consider the family of relaxed energies $(\mathcal{E}_r)_{r=r_0}^1$. The equation in r

$$\underline{\mu} + \beta \max \left\{ \underline{\eta_r} \overline{\nu}, \sum_{q=1}^{Q} \eta_r^q \overline{\nu}^q \right\} = 0 \tag{13}$$

is satisfied for a unique r_0. Any relaxed energy $\mathcal{E}_r$, corresponding to $r < r_0$, is strictly convex.

Conditions for initial convexity, established in [5], appear as a particular case of this proposition. The above condition is sufficient; it is also necessary when the both equalities in (12) can be met, i.e., when there exists a pair $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{v}})$ such that $\tilde{\boldsymbol{v}}^T [D^2 \mathcal{L}(\tilde{\boldsymbol{x}})] \tilde{\boldsymbol{v}} = \underline{\mu} ||\tilde{\boldsymbol{v}}||^2$ and $\tilde{\boldsymbol{v}}^T [D^2 \Phi_r(\tilde{\boldsymbol{x}})] \tilde{\boldsymbol{v}} = \underline{\sigma_r} ||\tilde{\boldsymbol{v}}||^2$, i.e., when the both bounds in Definitions 1 and 2 are reached for the same $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{v}}$. This occurs for any $\tilde{\boldsymbol{v}}$ in the case when $\mathcal{A} = \mathcal{I}$ in the LG model (1) and the prior energy is such that $\underline{\sigma_r} = \underline{\eta_r} \overline{\nu}$ (cf., Section IV-B).

Equation (13) may be solved numerically, if analytic resolution is difficult. For numerical reasons, the case $\underline{\mu} > 0$ presents a practical interest only when $\underline{\mu}$ is not too close to zero, since $\underline{\mu} \approx 0$ leads to $r_0 \approx 0$. The latter case then should be assimilated to the case $\underline{\mu} = 0$.

B. Nonstrictly Convex Data-Fidelity Term

The situation $\underline{\mu} = 0$ corresponds to ill-posed inverse problems [12], [43]. Such are the ECT and TCT models, and the frequent cases when $\mathcal{A}^T\mathcal{A}$ in (1) is singular or ill-conditioned. Data-fidelity $\mathcal{L}$ is only nonstrictly convex and the r.h.s. of (12) is negative for any $r \in]0, 1[$. This fact does not necessarily imply that $\mathcal{E}_r$ is strictly nonconvex for any r, although this is produced if a pair $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{v}})$ exists such that $\tilde{\boldsymbol{v}}^T [D^2 \mathcal{L}(\tilde{\boldsymbol{x}})] \tilde{\boldsymbol{v}} = 0$ and $\tilde{\boldsymbol{v}}^T [D^2 \Phi_r(\tilde{\boldsymbol{x}})] \tilde{\boldsymbol{v}} < 0$. So, a relaxed energy $\mathcal{E}_r$, corresponding to a singular LG model (1) and a regularization applied to the usual finite differences, is *never convex*, for any r, since the previous conditions are met for any $\tilde{\boldsymbol{v}} \in \operatorname{Ker}(\mathcal{A}^T\mathcal{A})$ and for $\tilde{\boldsymbol{x}}$ such that $|\boldsymbol{d}_m^q \tilde{\boldsymbol{x}}| = T_r$ for any (q, m). Note that this fact has not been accounted for by several authors who applied GNC to LG ill-posed problems [29], [38]: the initial relaxed energy they used was nonconvex and certainly multimodal.

A possible issue is then to *render convex* the initial relaxed prior energy Φ_{r_0}. Let us fix ρ close to zero, $\rho \in]0, 1[$ and consider $r_0 < \rho$. In the beginning of the GNC minimization—i.e., when $r \in [r_0, \rho[$—the relaxed prior energy Φ_ρ is kept constant while an auxiliary convex term Θ is appended to it:

$$\begin{aligned} \Phi_r(\boldsymbol{x}) &= \Phi_\rho(\boldsymbol{x}) + \kappa(r) \Theta(\boldsymbol{x}) \quad \text{with} \\ \kappa(r) &= \frac{\rho - r}{\rho - r_0} \quad \text{for } r \in [r_0, \rho[, \end{aligned} \tag{14}$$

$$\begin{aligned} \Theta(\boldsymbol{x}) &= \sum_{q=1}^{Q} \Theta^q(\boldsymbol{x}), \quad \text{where} \\ \Theta^q(\boldsymbol{x}) &= \sum_{m \in \mathcal{S}^{(q)}} \theta^q(\boldsymbol{d}_m^q \boldsymbol{x}). \end{aligned} \tag{15}$$

In order to render Φ_r convex for $r \leq r_0$, each Θ^q, $1 \leq q \leq Q$, has to compensate for the nonconvexity of the relevant Φ_ρ^q. At the same time, Θ should deform Φ_ρ the least. The auxiliary PF's θ^q satisfy:

(a1) θ are $\mathcal{C}^1$-continuous;

(a2) θ are symmetric, $\theta(t) = \theta(-t)$ and $\theta(0) = 0$;

(a3)

$$\begin{cases} \ddot{\theta}(t) = 0, & \text{if } \ddot{\varphi}_\rho(t) > 0 \text{ i.e., if } |t| < u_\rho, \\ \ddot{\theta}(t) = -\ddot{\varphi}_\rho(t), & \text{if } \ddot{\varphi}_\rho(t) \leq 0 \text{ i.e., if } |t| \geq u_\rho \end{cases}$$

(where we drop superscript q).

Observe that a relaxed PF φ_ρ, satisfying (r1–r3), is strictly convex over an interval, $|t| < u_\rho$, where $u_\rho > 0$. We set $\theta(t) = 0$ for $|t| < u_\rho$ and $\theta(t) = -\varphi_\rho(t) + c|t| + d$ for $|t| \geq u_\rho$. Smoothness at $\pm u_\rho$ requires that $\theta(u_\rho^+) = 0$ and $\dot{\theta}(u_\rho^+) = 0$, where $\theta(u^+)$ means $\lim_{t \to u, t > u} \theta(t)$. These conditions lead to $c = \dot{\varphi}_\rho(u_\rho^+)$ and $d = \varphi_\rho(u_\rho^+) - c u_\rho$, so finally

$$\theta(t) = \begin{cases} 0, & \text{if } |t| < u_\rho, \\ \varphi_\rho(u_\rho^+) - \varphi_\rho(t) + (|t| - u_\rho) \dot{\varphi}_\rho(u_\rho^+), & \text{if } |t| \geq u_\rho. \end{cases}$$

The obtained $\mathcal{E}_r$ (14)–(15) is convex for any r such that $r \leq r_0$. It is nonconvex for $r \in]r_0, \rho[$, while the auxiliary term vanishes at $r = \rho$. For $r \geq \rho$, relaxed prior energy Φ_r finds the original form (8).

If ever calculating $\underline{\mu}$ is problematic, the "nonstrict convexity" strategy can be applied: if ρ is smaller than the solution of (13), the both strategies (Sections V-A and V-B) yield the same solution.

VI. Minima Tracking Along the Relaxed Energies

Minimization by GNC needs to calculate a sequence of intermediate solutions $\boldsymbol{x}^{(r_k)}$ which are local minima of the relaxed energies $(\mathcal{E}_{r_k})_{k=1}^{Q}$. Its computational cost is hence determined by the efficiency of the local minimization algorithm which is used. The latter can be performed using either coordinate-wise descent (ICD, ICM) or directional descent—gradient descent, conjugate gradients, etc. [1], [37], [39].

A. Convergence Considerations

Convergence of a continuously evolving GNC toward the global minimum in several very simple cases (step- and gate-shaped signals, noise-free data $\boldsymbol{y} = \boldsymbol{x}$, truncated quadratic PF) has been demonstrated in [5]. These results can hardly be extended to other observation models, to other PF's and to general images. Instead, we have conducted extensive experiments in order to evaluate the performance of the proposed GNC minimization and to explain its functioning.

Relaxed energy $\mathcal{E}_r$ is obtained by subtracting to $\mathcal{E}$ a *varying bias* $\mathcal{B}_r(\boldsymbol{x}) = \mathcal{E}(\boldsymbol{x}) - \mathcal{E}_r(\boldsymbol{x})$. Bias $\mathcal{B}_r$ is small for images $\boldsymbol{x}$ containing a large number of nearly null differences, $\boldsymbol{d}_m^q\boldsymbol{x} \approx 0$, and it vanishes for large differences. However, $\mathcal{B}_r$ increases along with the number of differences of $\boldsymbol{x}$ which are $\{(m,q)\text{: } |\boldsymbol{d}_m^q\boldsymbol{x}| \in [T, T_r]\}$. Let a deep minimum of $\mathcal{E}$, say $\hat{\boldsymbol{x}}$, contain numerous differences which are close to zero; in $\mathcal{E}_r$, such a $\hat{\boldsymbol{x}}$ is surrounded by a "corona" where many differences are $|\boldsymbol{d}_m^q\boldsymbol{x}| \in [T, T_r]$—i.e., where $\mathcal{B}_r$ is large. Suppose that the global minimum $\hat{\boldsymbol{x}}$ of $\mathcal{E}$ is deep and consider a deep local minimum $\hat{\boldsymbol{x}}_1$ such that $\mathcal{E}(\hat{\boldsymbol{x}}) \lessapprox \mathcal{E}(\hat{\boldsymbol{x}}_1)$. We examine the two coronae where $\mathcal{B}_r$ is large, generated by $\hat{\boldsymbol{x}}$ and by $\hat{\boldsymbol{x}}_1$, respectively. Let r_k be such that $\mathcal{E}_{r_k}$ is unimodal over a region containing $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}_1$, as well as the past intermediate solution $\boldsymbol{x}^{(r_{k-1})}$. If *moreover* $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}_1$ are close to each other, i.e., $||\hat{\boldsymbol{x}} - \hat{\boldsymbol{x}}_1||$ is small, $\mathcal{E}_r(\hat{\boldsymbol{x}})$ involves an important bias coming from $\hat{\boldsymbol{x}}_1$ and reciprocally, $\mathcal{E}_r(\hat{\boldsymbol{x}}_1)$ involves an important bias due to $\hat{\boldsymbol{x}}$ (i.e., $\mathcal{B}_r(\hat{\boldsymbol{x}})$ and $\mathcal{B}_r(\hat{\boldsymbol{x}}_1)$ are large). It may then happen that $\mathcal{E}_{r_k}(\hat{\boldsymbol{x}}_1) < \mathcal{E}_{r_k}(\hat{\boldsymbol{x}})$, and that the actual intermediate solution $\boldsymbol{x}^{(r_k)}$ is closer to $\hat{\boldsymbol{x}}_1$ rather than to $\hat{\boldsymbol{x}}$. The sequence of minima $\boldsymbol{x}^{(r_k)}$ may then be entailed to evolve in the attraction valley of $\hat{\boldsymbol{x}}_1$ and finally, to provide $\hat{\boldsymbol{x}}_1$ as the ultimate solution. Normally, such a "confusion" of minima does not seriously deform the solution since it concerns only several details—because close minima differ only at several small edges. On the other hand, such a confusion does not occur if minima $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}_1$ are distanced from each other and/or their energies are quite different. This reasoning is corroborated by the experiments in Section VII-A: the images shown in Figs. 7(c), 9(b)–(c), correspond to close local minima and indeed, these solutions are quite similar.

An illustration of these considerations can be found in Fig. 3 and Appendix B.

The *adequacy* of an ultimate solution $\boldsymbol{x}^{(r_k)}$ obtained using GNC—its closeness to the global minimum $\hat{\boldsymbol{x}}$—greatly depends on both, the level of the convex initialization (cf., Section VI-B) and the discrete relaxation sequence $(r_k)_{k=0}^{K}$ (cf., Section VI-C). To simplify the presentation, it is assumed that $\varphi^q = \varphi$ for any q.

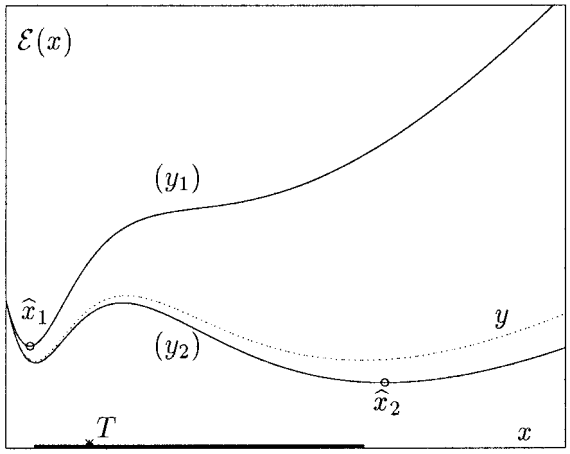


Fig. 4. Edge detection. Scalar energy $\mathcal{E}$ corresponds to TCT (2) regularized with a Lorentzian PF [Fig. 2(b)]: $\mathcal{E}(x) = \rho \exp(-hx) + yhx + \beta\alpha x^2/(1+\alpha x^2)$. If $y_1 < y$, the global minimum is close to zero, while for $y_2 > y$ it is far from zero. The global minimum of $\mathcal{E}$ cannot lie in the zone marked with (——) which contains the point $x = T$.

B. Initialization of GNC

Convexity conditions established in Sections V-A and V-B provide a bound r_0 ensuring that $\mathcal{E}_r$ is convex for $r \leq r_0$. It is hence sufficient to start GNC with $r \lessapprox r_0$, since $\mathcal{E}_r$ then has a unique minimum.

When $\mathcal{L}$ is strictly convex, r_0 is uniquely determined by Proposition 2. If $\mathcal{L}$ is only nonstrictly convex, the choice of ρ in (14) is guided by the following argument. Unlike the principal relaxation of Φ_r (8), the weight of the auxiliary terms Θ^q, $q = 1, \cdots, Q$ in $\mathcal{E}_r$ increases rapidly with $|\boldsymbol{d}_m^q\boldsymbol{x}|$, which penalizes the recovery of large differences during the early stages of a GNC minimization. The intermediate solutions can thus be attracted to evolve toward a smooth local minimum. For this reason, we choose ρ quite small—or equivalently, T_ρ quite large. The many experiments that we performed showed systematically that choosing T_ρ larger leads to deeper minima of $\mathcal{E}$. However, numerical instabilities may arise when ρ is too close to zero, because $\mathcal{E}_{r_0}$ may then exhibit almost flat regions where local descent becomes problematical. In our experiments (Section VII) we used $T_\rho = 40 - 80$ along with a concave PF.

C. Relaxation Sequence

If $\mathcal{E}$ has two close local minima, $\mathcal{E}$ varies sharply and has a maximum along the line linking these minima. Conversely, sharp variations in the shape of $\mathcal{E}$ are related to the presence of close minima; following Section VI-A, closeness of minima suggests the recovery of fine features in the estimate. During the early stages of a GNC optimization, parameter r is close to zero and the concavities of the relaxed PF's φ_r are weak. The relevant relaxed energies $\mathcal{E}_r$ have only a small number of local minima which are distanced from each other, while $\mathcal{E}_r$ vary weakly between them. Mainly the rough features of the solution are recovered during these early stages. As long

as r increases, T_r evolves toward the original T while $\mathcal{E}_r$ exhibits more and more close local minima. Thus, fine features are recovered along with the advancement of the GNC-optimization. It results that edges are detected throughout the whole GNC-minimization, which corresponds to a *progressive segmentation*, "from coarse to fine," of the reconstructed image, under the control of T_r. This is the reason why we focus on the evolution of $(T^q_{r_k})_{k=0}^K$ toward T and then calculate the relevant sequence $(r_k)_{k=0}^K$ using the expressions for T_r in Appendix A.

The auxiliary relaxation (when it is applied) introduces a preliminary stage in the GNC-optimization.

- During the auxiliary relaxation (14)–(15), the relaxed threshold remains constant, $T_{r_k} = T_\rho$ for $r_k \in [r_0, \rho]$. It can be observed that GNC is not particularly sensitive to the form of this early relaxation. We decrease κ *linearly* from $\kappa(r_0) = 1$ to $\kappa(\rho) = 0$ in several steps (3–5 in practice).

  If auxiliary relaxation is unnecessary, this stage is automatically omitted by setting $\rho = r_0$.
- The main relaxation, $r_k \geq \rho$, is adapted to the evolution of the thresholds T_{r_k}. Different schedules for this evolution can be envisaged. A slow initial decrease of T_{r_k} followed by an acceleration is justified by the fact that the main features of the solution are recovered during the early stages of the GNC minimization. In particular, the *exponential decrease*, $T_{r_k} = T_\rho + (T_{r_K} - T_\rho)[1 - e^{\tau(k-1)}]/[1-e^{\tau(K-1)}]$, where $\tau \in]0,1[$ controls the speed of this decrease, is revealed to give rise to important practical minimizations.

  Alternative schemes for the decrease of T_{r_k} are *linear* $T_{r_k} = T_\rho + k(T - T_\rho)/(K-1)$ or *logarithmical* $T_{r_k} = T_\rho + \ln(k+1)(T - T_\rho)/\ln(K)$. We found that the logarithmical scheme leads to more shallow minima then the linear scheme. These two schemes perform less well than the exponential scheme. In our experiments, we did not find exceptions.

We found also that a GNC relaxation in $K \approx 30$–40 steps permits a convenient evolution for T_{r_k}.

D. Pertinent Initialization for Direct Local Optimization

By construction, $\mathcal{E}_{r_0}$ is the convex energy "closest" to the original $\mathcal{E}$. Following Section VI-A, we can remark that $\mathcal{E}_{r_0}$ reflects the main features expressed in Φ, but in an "attentuated form"—because of the relaxation. We can say that $\boldsymbol{x}^{(r_0)}$ is the unique minimum of the "best" convex approximation of the original $\mathcal{E}$ and that it partially incorporates some features expressed in Φ. That is why we advocate that an initial GNC solution $\boldsymbol{x}^{(r_0)}$ can be used as starting point for ICM-like direct local minimization (Section III-A). In particular, this minimizer $\boldsymbol{x}^{(r_0)}$ is closer to $\hat{\boldsymbol{x}}$ than a ML estimate. Moreover, it is always available, even when the ML estimate is unstable.

For a well-posed problem, the initial convex energy $\mathcal{E}_{r_0}$ is uniquely determined using Proposition 2. If the problem is ill-posed, parameter ρ in (14)–(15) has to be chosen. Note that a GNC-suited initial solution—corresponding to Section VI-B—involves important noise effects [see Fig. 7(a)] and it is not adapted for a direct local minimization. In the latter case, the value of ρ should be larger. On the other hand, an initial solution where the noise effects are well smoothed is very likely to correspond to a smooth local minimum of $\mathcal{E}$. Experience confirms that T_ρ should yield an initial solution $\boldsymbol{x}^{(r_0)}$ which is only slightly under-regularized [see Fig. 9(a)–(c)]. The precise value of ρ is set experimentally (Section VII).

VII. Numerical Results

Reconstructions, presented in this Section, are defined using different MRF prior energies. The use of nonconvex PF's and GNC minimization is compared to the main alternative approaches.

A. Deconvolution of an Image

The original 72×72 image $\boldsymbol{x}$ in Fig. 5(a) is locally constant and is observed through a 9×9 PSF: $a(i,j) = \exp\{-0.3[(i-4)^2+(j-4)^2]\}$, $0 \leq i,j \leq 8$. Data $\boldsymbol{y} = \boldsymbol{x} * \boldsymbol{a} + \boldsymbol{n}$, given in Fig. 5(b), are corrupted by white Gaussian noise $\boldsymbol{n}$ with 10 dB SNR. The inverse problem is ill-posed and the ML estimate, Fig. 5(c), is unstable. The MRF prior energies used below involve the following elements:

$$\boldsymbol{d}^1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \boldsymbol{d}^{2^T}, \quad \boldsymbol{d}^3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \boldsymbol{d}^4 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$
$$\boldsymbol{d}^5 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = \boldsymbol{d}^{6^T}, \quad \boldsymbol{d}^7 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix};$$
$$\gamma^q = 1 \quad \text{for } 1 \leq q \leq 4, \quad \gamma^q = 0.045 \quad \text{for } 5 \leq q \leq 7,$$
$$\varphi^q = \varphi \quad \text{for } 1 \leq q \leq 7. \tag{16}$$

The estimates presented below are defined using either convex or nonconvex PF's. Among the former PF's, we applied a Huber PF and a generalized Gaussian (GG) PF [8], [24]:

$$\text{(Huber)} \quad \varphi(t) = t^2 \mathbb{I}(|t| < \alpha) + (\alpha^2 + 2\alpha|t-\alpha|) \cdot \mathbb{I}(|t| \geq \alpha), \tag{17}$$

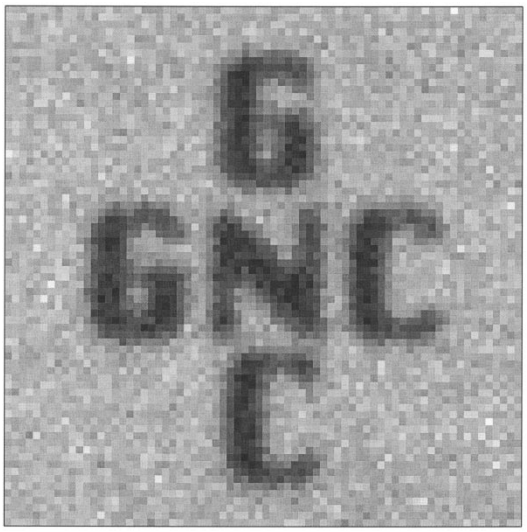
$$\text{(GG)} \quad \varphi(t) = |t|^\alpha, \qquad 1 < \alpha \leq 2 \tag{18}$$

where $\mathbb{I}(p) = 1$ if p is true and $\mathbb{I}(p) = 0$ otherwise. Parameters (α, β) are set *experimentally* in such a way as to reach the best reconstruction allowed by each PF.

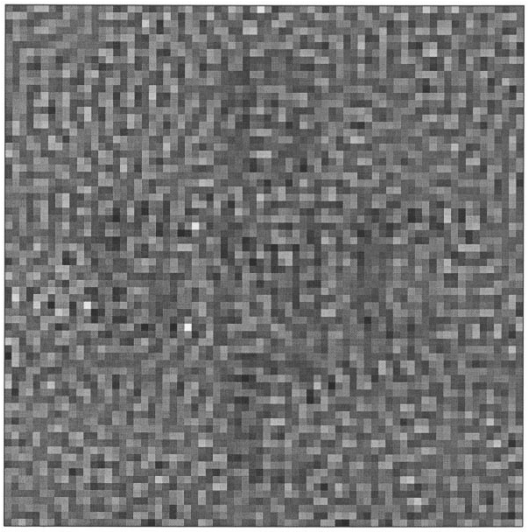
The image in Fig. 6(a) is restored using a quadratic PF (a Gaussian MRF) and $(\alpha = 2, \beta = 5)$ in (18), and it does not exhibit large differences. The next Fig. 6(b) shows a restoration obtained using a GG PF with $(\alpha = 1.1, \beta = 6)$. In Fig. 6(c), a Huber PF with $(\alpha = 0.2, \beta = 5)$ is used. In the reconstructed images, large differences may be distinguished but they are not neat.

The remaining reconstructions are defined using the *concave PF* with $(\alpha = 12, \beta = 16)$ and they are calculated using different techniques. The images in Fig. 7 illustrate the proposed GNC minimization with auxiliary relaxation. The initial solution $\boldsymbol{x}^{(r_0)}$, shown in Fig. 7(a), is underregularized and so is the intermediate solution $\boldsymbol{x}^{(\rho)}$ in Fig. 7(b). Both, the contours and the locally constant regions are correctly

(a)

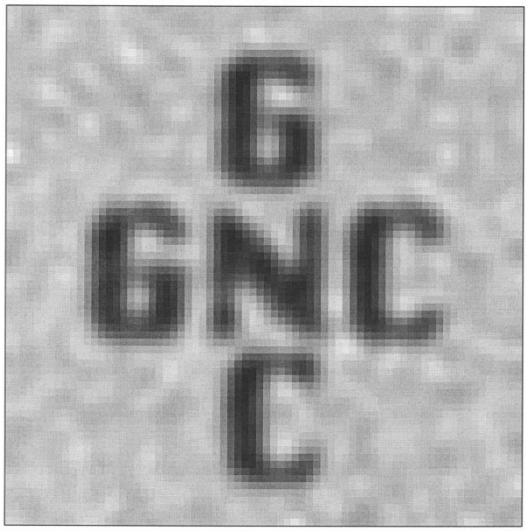


(b)



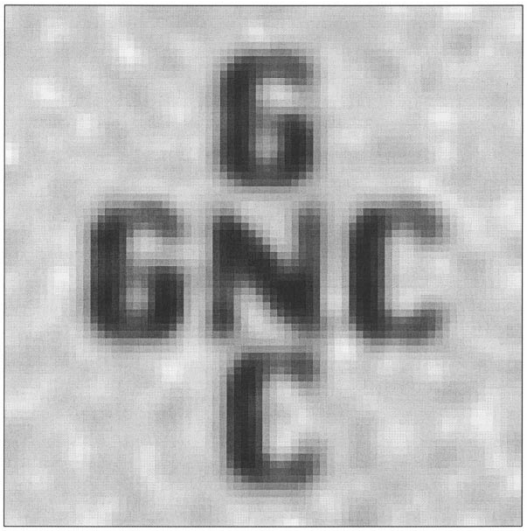
(c)

Fig. 5. Deconvolution example. (a) Original image. (b) Data—a blurred and noisy (10 dB SNR) version of the original image. (c) ML estimate.



(a)

(b)



(c)

Fig. 6. Convex MRF energies: $\gamma = [1, 1, 1, 1, 0.045, 0.045, 0.045]$. (a) Gaussian MRF $(\varphi(t) = t^2)$, $(\alpha = 2, \beta = 5)$. (b) GG MRF, $(\alpha = 1.1, \beta = 6)$. (c) Huber PF, $(\alpha = 0.2, \beta = 5)$.

retrieved in the ultimate GNC solution—Fig. 7(c). The latter is obtained using an exponential decrease of T_{r_k}, where $T_\rho = 80$, $T_{r_K} = 0.01$, $K = 30$, and $\tau = 1$, followed by an ultimate stage with $T_{r_{K+1}} = 0.003$—in order to closely fit $\varphi_{r_{K+1}} \approx \varphi$. Ultimate energy $\mathcal{E}(\hat{\boldsymbol{x}}) = 2.53\ 10^4$ is the lowest energy that we could reach by any minimization technique. The role of the initial threshold T_ρ is illustrated in Fig. 8(a): a larger $T_\rho = 20$ leads to a more shallow minimum—$\mathcal{E}(\hat{\boldsymbol{x}}) = 2.73\ 10^4$—which is slightly different visually. GNC optimization with T_{r_k} decreasing linearly and logarithmically starting from $T_\rho = 80$ leads to local minima with a higher energy, $\mathcal{E}(\hat{\boldsymbol{x}}) = 2.64\ 10^4$ and $\mathcal{E}(\hat{\boldsymbol{x}}) = 2.72\ 10^4$, respectively—Fig. 8(b) and (c).

The ML solution—Fig. 5(c)—does not provide a useful initialization for direct local minimization. Instead, the GNC initial solution $\boldsymbol{x}^{(r_0)}$ corresponding to $T_\rho = 20$—Fig. 9(a)—is underregularized and it leads to a very acceptable solution, given in Fig. 9(b), where $\mathcal{E}(\hat{\boldsymbol{x}}) = 2.76\ 10^4$. Initialization, corresponding to a smaller threshold $T_\rho = 10$ provides a slightly different solution, shown in Fig. 9(c), and having a slightly higher energy $\mathcal{E}(\hat{\boldsymbol{x}}) = 2.79\ 10^4$. We calculated the same estimate using also a half-quadratic SA [21], but we obtained a shallow local minimum where $\mathcal{E}(\hat{\boldsymbol{x}}) = 10.26\ 10^4$.

B. Emission Tomography

The concentration of an isotope in a part of the body provides an image characterizing the metabolic functions and local blood flow [9], [23], [27]. In ECT, a radioactive drug is introduced in a region of the body and the emitted photons are recorded around it. Data are formed by the number of photons $y_j \geq 0$ reaching each detector, $j = 1, \cdots, N$. Data-fidelity $\mathcal{L}$, given in (2), is nonstrictly convex since $D^2\mathcal{L}(\boldsymbol{x}) = \Sigma_{j=1}^N y_j \boldsymbol{h}_j \boldsymbol{h}_j^T / (\boldsymbol{h}_j^T \boldsymbol{x}_j)^2$ is nonnegative definite but ill-conditioned; so $\underline{\mu} = 0$.

We treat the reconstruction of the 64×64 phantom in Fig. 10(a)–(b) with amplitudes in]0, 3.8], from the simulated ECT raw data in Fig. 10(c). Data correspond to $\rho = 8$, $\mu = 0.08$ in (2) and are collected on 64 arrays surrounding the object at equally spaced angles, each containing 64 detectors.

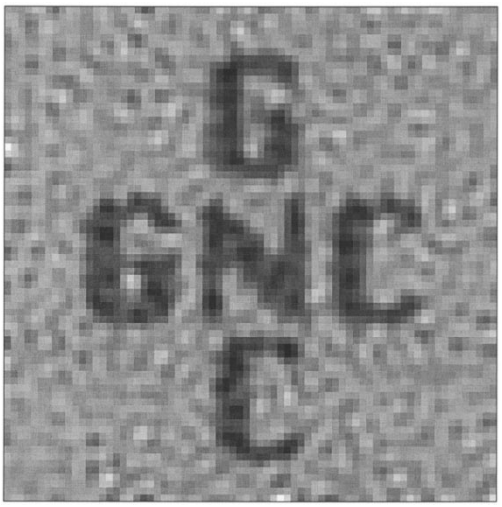
The MRF prior energies used in these reconstructions involve first and second order differences, where $\gamma^q = 1$ for $1 \leq q \leq 4$ and $\gamma^q = 0.01$ for $5 \leq q \leq 7$ in (16). In addition, a "soft" positivity constraint is imposed on the solution by appending to $\mathcal{E}_r$ a *convex* term Ψ with $\omega = 60$

$$\mathcal{E}_r(\boldsymbol{x}) + r\omega\Psi(\boldsymbol{x}), \Psi(\boldsymbol{x}) = \sum_{m \in \mathcal{S}} x_m^2 \mathbb{I}(x_m \leq 0). \tag{19}$$

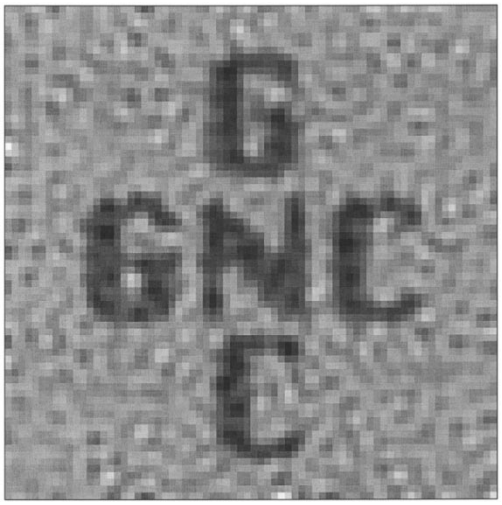
Weighting with r serves to improve conditioning for local optimization.

A reconstruction defined using a Huber PF (17) and parameters $(\alpha = 0.2, \beta = 0.5)$ is given in Fig. 11. Note that a reconstruction using a GG PF (18) with $\alpha = 1.1$ provided a closely similar solution. The reconstructed images allow a further interpretation by the user, but they are slightly smooth.

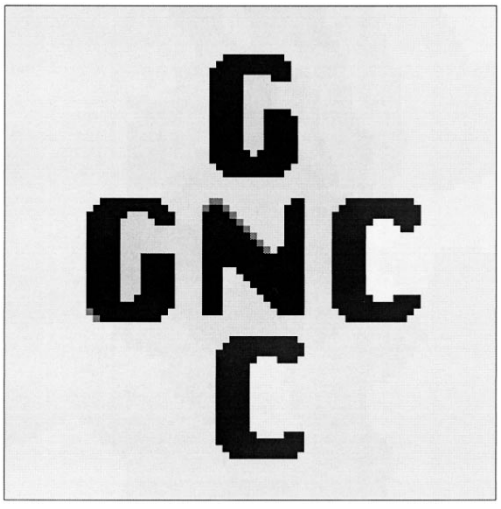
The last reconstruction is obtained using a concave PF with $(\alpha = 4, \beta = 1.5)$. The minimization is performed



(a)

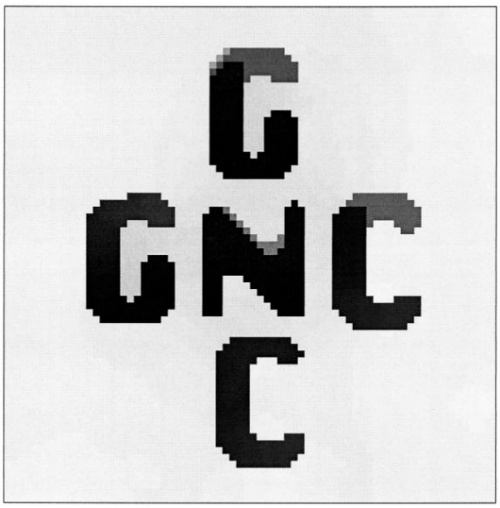


(b)

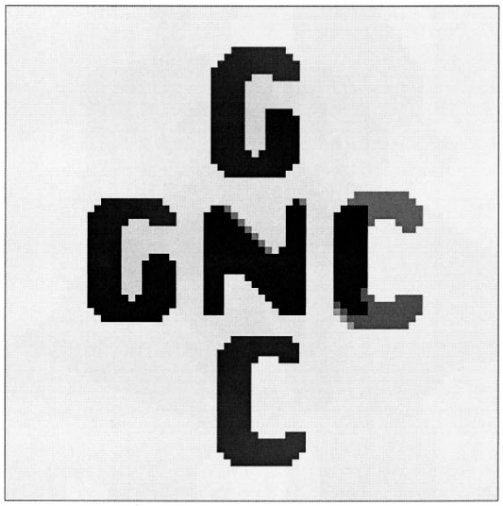


(c)

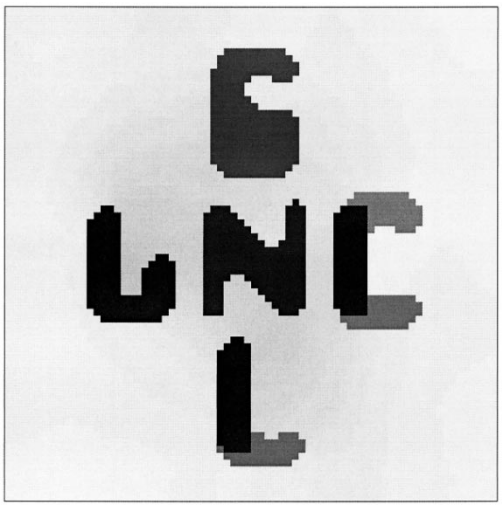
Fig. 7. MRF energy using a concave PF and GNC-minimization: $(\alpha = 12, \beta = 16)$ and γ = [1, 1, 1, 1, 0.045, 0.045, 0.045], *exponential* decrease of T_{r_k} with $T_\rho = 80$, $T_{r_K} = 0.01$, $K = 30$, $\tau = 1$, followed by $T_{r_{K+1}} = 0.003$. (a) The initial solution $\boldsymbol{x}^{(r_0)}$. (b) Intermediate solution $\boldsymbol{x}^{(\rho)}$. (c) The ultimate GNC solution $\hat{\boldsymbol{x}} = \boldsymbol{x}^{(r_K)}$; its energy is $\mathcal{E}(\hat{\boldsymbol{x}}) = 2.53\ 10^4$.



(a)

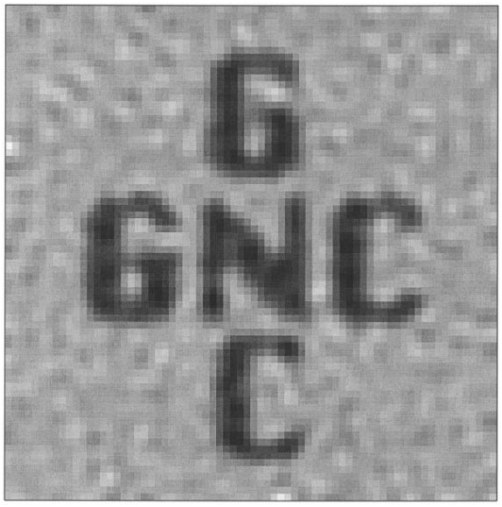


(b)

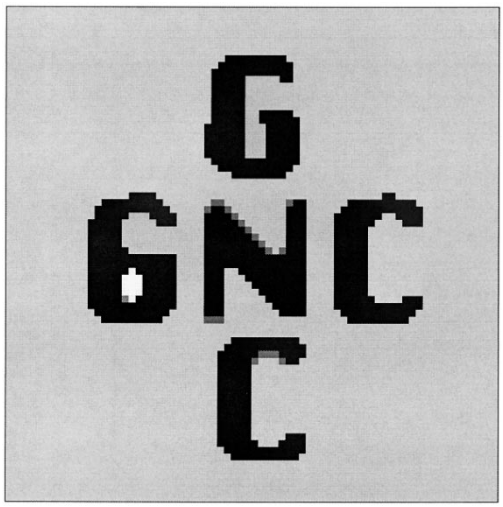


(c)

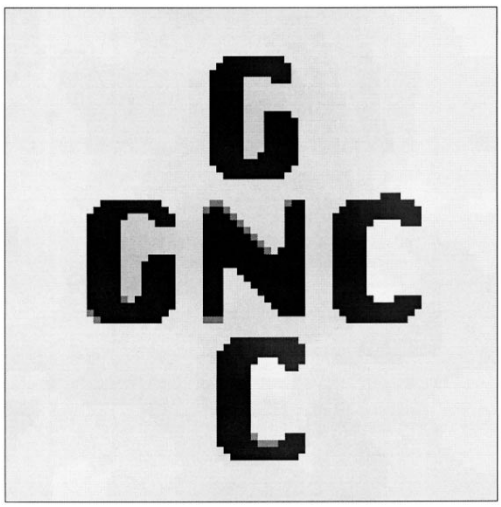
Fig. 8. Other relaxations (the same estimate as in [Fig. 7]). (a) GNC minimization starting from $T_\rho = 20$, with $T_{r_K} = 0.001$, $K = 30$, then $\mathcal{E}(\hat{\boldsymbol{x}}) = 2.73\ 10^4$. For (b) and (c), $T_\rho = 80$, $T_{r_K} = 0.001$, $K = 30$. (a) GNC minimization using a *linear* decrease of T_{r_k}, then $\mathcal{E}(\hat{\boldsymbol{x}}) = 2.64\ 10^4$. (b) GNC minimization using a *log*-decrease of T_{r_k}, in which case $\mathcal{E}(\hat{\boldsymbol{x}}) = 2.27\ 10^4$.



(a)

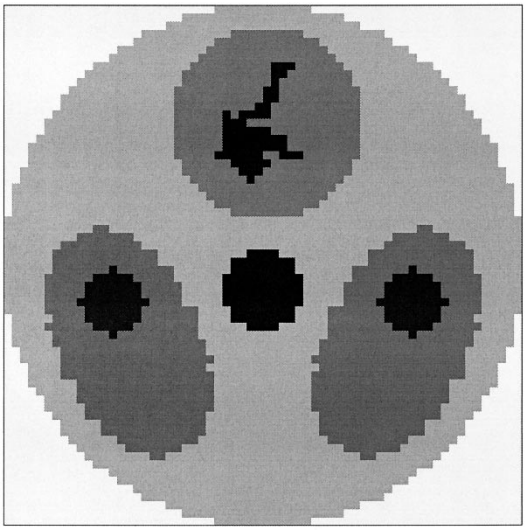


(b)

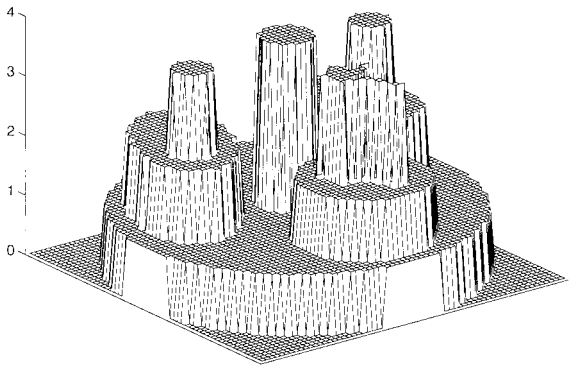


(c)

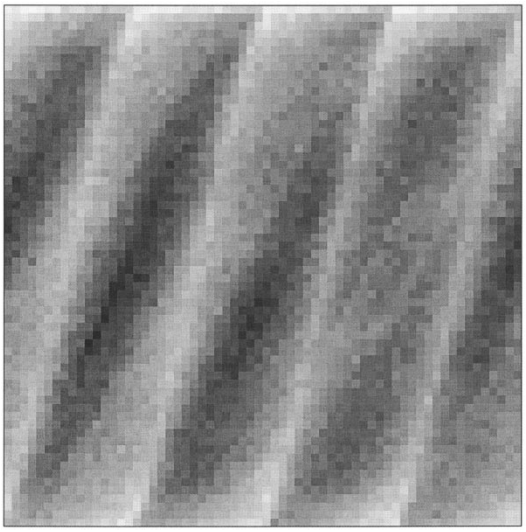
Fig. 9. Direct local minimization (the same estimate as in [Fig. 7]). (a) Initialization corresponding to $T_\rho = 20$. (b) Solution found by local minimization in its vicinity has energy $\mathcal{E}(\hat{\boldsymbol{x}}) = 2.76\ 10^4$. (c) Solution found using initalization with $T_\rho = 10$, then $\mathcal{E}(\hat{\boldsymbol{x}}) = 2.79\ 10^4$.



(a)



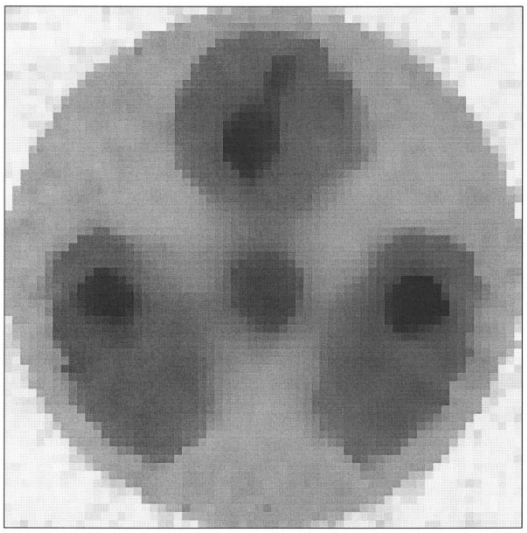
(b)



(c)

Fig. 10. Emission tomography (a)–(b) Original 72×72 phantom. (c) ECT simulated raw data with $\mu = 0.8$ and $\rho = 8$).

using GNC with auxiliary relaxation, where T_{r_k} decreases exponentially from $T_\rho = 40$ to $T_{r_K} = 0.01$ in $K = 30$ steps. Fig. 12(a) presents the initial GNC solution $\boldsymbol{x}^{(r_0)}$. The ultimate GNC solution—Fig. 12(b)–(c)—has a nice resolution and the contours are neat.



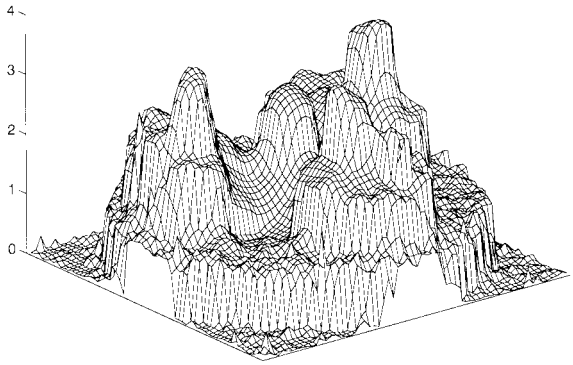
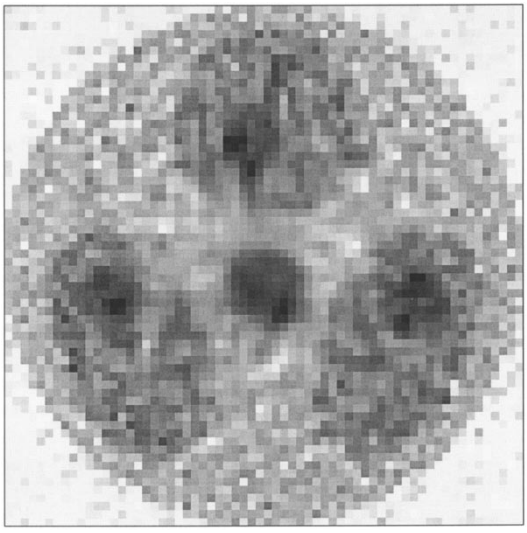


Fig. 11. Reconstruction using a second-order MRF energy with a Huber PF: γ = [1, 1, 1, 1, 0.01, 0.01, 0.01], $(\alpha = 0.2, \beta = 0.5)$ and a soft positivity constraint, $\omega = 60$ [cf. (17), (19)].

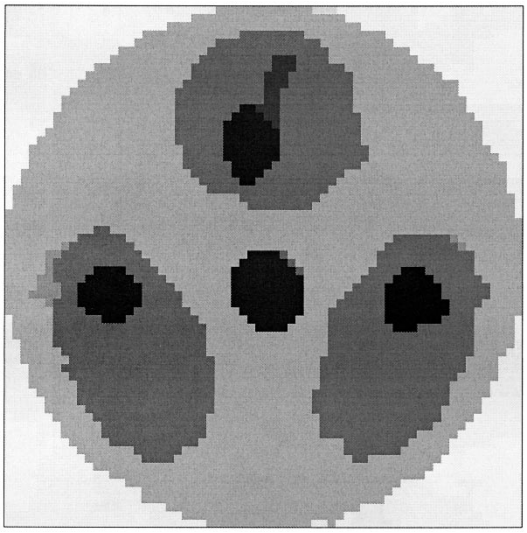
VIII. Concluding Remarks

In this paper, we proposed a family of efficient GNC-algorithms for Markovian MAP reconstruction of images and signals. Data are obtained at the output of a system giving rise to a convex data-fidelity term. Prior scene features are introduced in the reconstruction by the means of MRF energies, defined as the application of general nonconvex PF's to the differences between neighboring pixels. This allows to recover images (or signals) composed of homogeneous zones separated by neat edges.

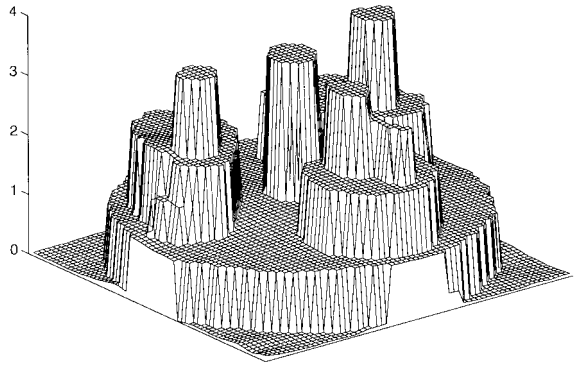
The relevant MAP energies are nonconvex and generally multimodal, and their optimization is a difficult task. Previously, the use of nonconvex PF's was limited to observation operators having either an extremely restricted support, or which are linear shift-invariant, and often to particular prior energies. In order to deal with the global optimization problem, we focused on the GNC algorithm which was initially proposed for the minimization of a MAP energy conceived for the denoising of images using truncated quadratic PF's. Based on a proper theoretical analysis, we developed an extension of GNC permitting to compute MAP estimates involving any convex data-fidelity term and any nonconvex and/or nonsmooth PF. Both theoretical and practical recommenda-



(a)



(b)



(c)

Fig. 12. Reconstruction using a second-order MRF energy with a concave PF: $\gamma = [1, 1, 1, 1, 0.01, 0.01, 0.01]$, $(\alpha = 4, \beta = 1.5)$ and $\omega = 60$. (a) The initial GNC solution corresponding to $T_\rho = 40$. (b)–(c) The ultimate solution $T_{r_K} = 0.01$ obtained after 30 GNC-steps using an exponential decrease for T_{r_k}.

tions for the construction of GNC algorithms were provided. The resultant algorithms are mathematically suboptimal, but definitely efficient. On the other hand, the GNC approach suggests how to define pertinent initializations for direct local minimization. The resultant method overcomes the limitations of classical ICM when faced with ill-posed problems.

Experimental results bear on the deconvolution of a blurred and noisy image and on the reconstruction of a phantom from simulated ECT raw data. The quality of the reconstructions obtained with MRF priors involving nonconvex PF's is due to the pertinence of the prior constraints.

Appendix A
Functions and Approximations

We present several PF's φ dependent on a parameter α. These PF's are illustrated in Fig. 2.

Truncated Quadratic PF: Such PF's are often used for the processing of images, where $\mathcal{A} = \mathcal{I}$ in the LG model (1) [5], [26], [31], [33], [42]. Extension to ill-posed linear inverse problems is proposed in [35].

$$\varphi(t) = \begin{cases} \alpha t^2, & \text{if } |t| < T, \\ 1, & \text{if } |t| \geq T, \end{cases} \qquad \begin{cases} T = 1/\sqrt{\alpha}, \\ \eta = \infty, \end{cases}$$

$$\varphi_r(t) = \begin{cases} \alpha t^2, & \text{if } |t| < a_r, \\ 1 - \dfrac{r(|t| - b_r)^2}{2(1-r)} & \text{if } a_r \leq |t| < b_r, \\ 1, & \text{if } b_r \leq |t|, \end{cases}$$

$$\begin{cases} a_r = \dfrac{T}{\sqrt{1 - 2\alpha + \dfrac{2\alpha}{r}}}, \\ b_r = T^2/a_r, \end{cases}$$

$$\begin{cases} T_r = (a_r + b_r)/2, \\ \eta_r = \dfrac{r}{1-r}, \end{cases}$$

$$\theta(t) = \begin{cases} 0, & \text{if } |t| < a_\rho, \\ \dfrac{1}{2} r(|t| - a_\rho)^2, & \text{if } a_\rho \leq |t| < b_\rho, \\ \dfrac{1}{2} r[(|t| - b_\rho)^2 - (a_\rho - b_\rho)^2], & \text{if } b_\rho \leq |t|. \end{cases}$$

Lorentzian PF: It has been used in [18] for SPECT image reconstruction, calculations being performed using ICM. Such PF's are also used for the denoising and the edge-enhancement of images in [36].

$$\varphi(t) = \frac{\alpha t^2}{1 + \alpha t^2}, \quad T = 1/\sqrt{\alpha}, \quad \eta = \alpha/2,$$

$$\varphi_r(t) = \frac{r\alpha t^2}{1 + r\alpha t^2}, \quad T_r = 1/\sqrt{r\alpha}, \quad \eta_r = r\alpha/2,$$

$$\theta(t) = \begin{cases} 0, & \text{if } |t| < u_\rho, \\ -\varphi_\rho(t) + |t|3/(8u_\rho) - 1/8, & \text{if } |t| \geq u_\rho, \end{cases}$$

$$u_\rho = \frac{1}{\sqrt{3\rho\alpha}}.$$

Gaussian PF: Such PF's has been used for the denoising and the segmentation of images, along with a MFA optimiza-

tion [41] and anisotropic diffusion [36].

$$\begin{aligned}
\varphi(t) &= 1-\exp(-\alpha t^2), \quad T=\sqrt{3/(2\alpha)}, \eta = 4\alpha e^{-3/2},\\
\varphi_r(t) &= 1-\exp(-r\alpha t^2), \quad T_r=\sqrt{3/(2r\alpha)},\\
\eta_r &= 4r\alpha e^{-3/2},\\
\theta(t) &= \begin{cases} 0, & \text{if } |t|<u_\rho,\\ e^{-\rho\alpha t^2}+e^{-(1/2)}(\sqrt{2\rho\alpha}|t|-2), & \text{if } |t|\geq u_\rho,\end{cases}\\
u_\rho &= \frac{1}{\sqrt{2\rho\alpha}}.
\end{aligned}$$

Concave PF: It has been applied for the restoration of blurred images, as well as for the reconstruction of SPECT images, where the solution is obtained using different forms of SA [20], [21], [45].

$$\begin{aligned}
\varphi(t) &= \frac{\alpha|t|}{1+\alpha|t|}, \quad T=0^+, \quad \eta=2\alpha^2,\\
\varphi_r(t) &= \begin{cases} \dfrac{a_r t^2}{1+b_r t^2}, & \text{if } |t|<\dfrac{1-r}{r},\\ \dfrac{\alpha|t|}{1+\alpha|t|}, & \text{if } |t|\geq\dfrac{1-r}{r},\end{cases}\\
&\quad \begin{cases} a_r=\dfrac{2r\alpha}{1-r},\\ b_4=\dfrac{r(r+2\alpha-2\alpha r)}{(1-r)^2},\end{cases}\\
&\quad \begin{cases} T_r=\dfrac{1}{b_r}<\dfrac{1-r}{r},\\ \eta_r=r\alpha/(1-r),\end{cases}\\
\theta(t) &= \begin{cases} 0, & \text{if } |t|<u_\rho\\ \dfrac{a_\rho}{4b_\rho}-\varphi_\rho(t)+(|t|-u_\rho)\varphi'_M, & \text{if } |t|\geq u_\rho\end{cases}\\
u_\rho &= \frac{1}{\sqrt{3b_\rho}}, \quad \varphi'_M=\frac{9a_\rho}{8\sqrt{3b_\rho}}.
\end{aligned}$$

Kronecker PF: It has been introduced for the segmentation and coding of images in [29], along with a GNC alogrithm involving a relaxation with Gaussians, $\varphi_r(t)=1-e^{-rt^2}$, $r\in]0,+\infty[$. This approach is applied later in [38]. Relaxation using Lorentzian functions improves the convergence (cf. Appendix B):

$$\begin{aligned}
\varphi(t) &= 1-\mathbb{I}(t=0), \quad T=0^{\pm}, \quad \eta=\infty,\\
\varphi_r(t) &= \frac{rt^2}{1-r+rt^2}, \quad T_r=\sqrt{\frac{1-r}{r}}, \quad \eta_r=\frac{1}{2}\frac{r}{1-r},\\
\theta(t) &= \begin{cases} 0, & \text{if } |t|<u_\rho,\\ -\varphi_\rho(t)+|t|\dfrac{3}{8u_\rho}-\dfrac{1}{8}, & \text{if } |t|\geq u_\rho,\end{cases}\\
u_\rho &= \frac{1}{\sqrt{3\rho}}.
\end{aligned}$$

Appendix B
Comparison of Relaxed Potentials

Kronecker PF can be relaxed using a regularization with Lorentzian or with Gaussian functions. The threshold of the Kronecker PF, equal to $T=0^+$, is less deformed by Lorentzians than by Gaussians:

$$\text{Lorentzian:} \quad T_r^{\mathrm{L}}=\frac{1}{\sqrt{2|\eta_r|}}\approx 0.7071\sqrt{\frac{1}{|\eta_r|}}$$

$$\text{Gaussian:} \quad T_r^{\mathrm{G}}=\sqrt{\frac{6}{|\eta_r|e^{3/2}}}\approx 1.1571\sqrt{\frac{1}{|\eta_r|}}$$

since $T_r^{\mathrm{L}}<T_r^{\mathrm{G}}$ for any η_r fixed. Experiments show that a relaxation using Lorentzians permits to obtain a better minimization of $\mathcal{E}$ than a relaxation using Gaussians.

The example in Fig. 3 presents the restoration of a gate-shaped 1-D signal $\boldsymbol{x}^o$ from blurred noise-free data $\boldsymbol{y}=\boldsymbol{a}*\boldsymbol{x}^o$, where $\boldsymbol{a}$ is a PSF. Energy $\mathcal{E}$ involves a Kronecker PF and it has two minima: a constant signal $\tilde{\boldsymbol{x}}=c\mathbf{1}$ and the original gate $\boldsymbol{x}^o$; here, $\mathbf{1}$ is composed of 1's and has the size of $\boldsymbol{x}$ and c is a scalar, $c=\arg\min_\zeta ||\boldsymbol{a}*(\zeta\mathbf{1}-\boldsymbol{x}^o)||$. The figures depict the section of the relaxed energies $\mathcal{E}_r$ along the line linking these minima, $e_r(t)=\mathcal{E}_r[t\tilde{\boldsymbol{x}}+(1-t)\boldsymbol{x}^o]$. The three sets of images correspond to three gates with decreasing widths. In Fig. 3(a), the global minimum is the original $\boldsymbol{x}^o$ and it is correctly retrieved using the two relaxations. The second gate, Fig. 3(b), is slightly narrower and it is still the global minimum of $\mathcal{E}$. In this case, a relaxation using Lorentzians lead to the global minimum $\boldsymbol{x}^o$ while a relaxation using Gaussians leads to the local minimum $\tilde{\boldsymbol{x}}$. The gate in Fig. 3(c) is still narrower and the global minimum of $\mathcal{E}$ corrsponds to $\tilde{\boldsymbol{x}}$. Although $\mathcal{E}(\tilde{\boldsymbol{x}})$ and $\mathcal{E}(\boldsymbol{x}^o)$ are close to each other, the two relaxations find the global minimum $\tilde{\boldsymbol{x}}$.

Appendix C
Differential and Hessian of the Prior Energy

In order to simplify notations, we write $\ddot{\Phi}_r$ for $D^2\Phi_r$. The elements of the differential of Φ_r^q are

$$\frac{\partial\Phi_r^q(\boldsymbol{x})}{\partial x_m}=\sum_{j\in C_0^q} d_j^q\dot{\varphi}_r^q(\boldsymbol{d}_{m+j}^q\boldsymbol{x})$$

Hessian $\ddot{\Phi}_r(\boldsymbol{x})$ reads $\ddot{\Phi}_r(\boldsymbol{x})=\Sigma_{q=1}^Q\gamma^q\ddot{\Phi}_r^q(\boldsymbol{x})$. The elements of the Hessian of Φ_r^q are shown at the bottom of the page. It is practical to consider the sums Σ_j above as infinite but having only a finite number of nonzero terms. Furthermore, recalling

$$\begin{aligned}
&[\ddot{\Phi}_r^q(\boldsymbol{x})]_{nn}=\textstyle\sum_j d_j^{q^2}\ddot{\varphi}_r^q(\boldsymbol{d}_{n+j}^q\boldsymbol{x}), && \text{for } n\in\mathcal{S}^{(q)},\\
&[\ddot{\Phi}_r^q(\boldsymbol{x})]_{nm}=\textstyle\sum_j d_j^q d_{j+n-m}^q\ddot{\varphi}_r^q(\boldsymbol{d}_{n+j}^q\boldsymbol{x}), && \text{if } n-m\in\bigcup_{j\in C_0^q} C_j^q\backslash\{0\},\\
&[\ddot{\Phi}_r^q(\boldsymbol{x})]_{nm}=0, && \text{if } n-m\notin\bigcup_{j\in C_0^q} C_j^q\backslash\{0\}.
\end{aligned}$$

that $\Sigma_m d_m^q = 0$ for any q, we can write

$$[\ddot{\Phi}_r^q(\boldsymbol{x})]_{nn} = - \sum_{m \in \mathcal{S}^{(q)} \setminus \{n\}} [\ddot{\Phi}_r^q(\boldsymbol{x})]_{nm}. \tag{20}$$

Appendix D
Proof of Proposition 1

We now develop the numerator of (9), $\boldsymbol{v}^T[\ddot{\Phi}_r(\boldsymbol{x})]\boldsymbol{v} = \Sigma_{q=1}^Q \boldsymbol{v}^T[\ddot{\Phi}_r^q(\boldsymbol{x})]\boldsymbol{v}$, where

$$\begin{aligned}
&\boldsymbol{v}^T[\ddot{\Phi}_r^q(\boldsymbol{x})]\boldsymbol{v} \\
&\quad = \sum_{k \in \mathcal{S}^{(q)}} \sum_{m \in \mathcal{S}} [\ddot{\Phi}_r^q(\boldsymbol{x})]_{km} v_k v_m \\
&\quad = \sum_k \left\{ \sum_{m \neq k} [\ddot{\Phi}_r^q(\boldsymbol{x})]_{km} v_k v_m + [\ddot{\Phi}_r^q(\boldsymbol{x})]_{kk} v_k^2 \right\}.
\end{aligned}$$

Thanks to (20), it is obtained

$$\begin{aligned}
&\boldsymbol{v}^T[\ddot{\Phi}_r^q(\boldsymbol{x})]\boldsymbol{v} \\
&\quad = \sum_k \sum_{m \neq k} [\ddot{\Phi}_r^q(\boldsymbol{x})]_{km} (v_k v_m - v_k^2) \\
&\quad = \sum_k \sum_{m \neq k} (v_k v_m - v_k^2) \sum_j d_j^q d_{j+k-m}^q \ddot{\varphi}_r^q(\boldsymbol{d}_{k+j}^q \boldsymbol{x}).
\end{aligned}$$

Let us set $n = k + j$ and $l = k + j - m$. Then $m \neq k$ leads to $l \neq j$, so

$$\boldsymbol{v}^T[\ddot{\Phi}_r^q(\boldsymbol{x})]\boldsymbol{v} = \sum_n \ddot{\varphi}_r^q(\boldsymbol{d}_n^q \boldsymbol{x}) \sum_j \sum_{l \neq j} d_j^q d_l^q (v_{n-j} v_{n-l} - v_{n-j}^2).$$

Since $\Sigma_{l \neq j} d_l^q = -d_j^q$ for any $j \in C_0^q$ (cf., Section II-A), $\Sigma_{l \neq j} v_{n-j}^2 d_j^q d_l^q = v_{n-j}^2 d_j^q \Sigma_{l \neq j} d_l^q = -v_{n-j}^2 (d_j^q)^2$. Then

$$\begin{aligned}
\boldsymbol{v}^T[\ddot{\Phi}_r^q(\boldsymbol{x})]\boldsymbol{v} &= \sum_n \ddot{\varphi}_r^q(\boldsymbol{d}_n^q \boldsymbol{x}) \left(\sum_j d_j^q v_{n-j} \right)^2 \\
&= \sum_n (\boldsymbol{d}_n \boldsymbol{v})^2 \ddot{\varphi}_r^q(\boldsymbol{d}_n^q \boldsymbol{x}).
\end{aligned}$$

Then we have

$$\boldsymbol{v}^T[\ddot{\Phi}_r(\boldsymbol{x})]\boldsymbol{v} = \sum_{q=1}^Q \gamma^q \sum_n (\boldsymbol{d}_n \boldsymbol{v})^2 \ddot{\varphi}_r^q(\boldsymbol{d}_n^q \boldsymbol{x}).$$

Using that $\eta_r^q = \min_t \ddot{\varphi}_r^q(t) < 0$, two inequalities can be extracted from the previous relation:

$$\begin{aligned}
&\boldsymbol{v}^T[\ddot{\Phi}_r(\boldsymbol{x})]\boldsymbol{v} \geq \sum_{q=1}^Q \eta_r^q \gamma^q \sum_n (\boldsymbol{d}_n \boldsymbol{v})^2 \geq ||\boldsymbol{v}||^2 \sum_{q=1}^Q \eta_r^q \overline{\nu}^q, \\
&\boldsymbol{v}^T[\ddot{\Phi}_r(\boldsymbol{x})]\boldsymbol{v} \geq \underline{\eta_r} \sum_{q=1}^Q \gamma^q \sum_n (\boldsymbol{d}_n \boldsymbol{v})^2 \geq ||\boldsymbol{v}||^2 \underline{\eta_r} \underline{\nu}
\end{aligned}$$

according to the definition of $\overline{\nu}^q$ and $\overline{\nu}$. Hence, the proposition.

Appendix E
Proof of Proposition 2

Let us consider the functions f_1, f_2, and f:

$$\begin{aligned}
f(r) &= \max_{i \in \{1,2\}} f_i(r) \quad \text{where} \\
f_1(r) &= \underline{\mu} + \beta \underline{\eta_r} \overline{\nu} \quad \text{and} \\
f_2(r) &= \underline{\mu} + \beta \sum_{q=1}^Q \eta_r^q \overline{\nu}^q.
\end{aligned}$$

Since $\lim_{r \to 1} \mathcal{E}_r(\boldsymbol{x})$ is strictly nonconvex, $\lim_{r \to 1} f_i(r) < 0$ for $i = 1, 2$. On the other hand, (r3) ensures that both terms, $\beta \underline{\eta_r} \overline{\nu} < 0$ and $\beta \Sigma_{q=1}^Q \eta_r^q \overline{\nu}^q < 0$, increase toward zero strictly monotonously along with $r \to 0$. So, both functions f_1 and f_2 increase strictly monotonously when $r \to 0$; then f is monotone incresing with $r \to 0$, as well. As $\underline{\mu}$ is strictly positive, f_1 and f_2 reach a positive value for r close to zero; the same is true for f as well. In conclusion, there exists a unique $r_0 \in]0, 1[$ which satisfies $f(r_0) = 0$.

Furthermore, the Hessian $D^2 \mathcal{E}_r(\boldsymbol{x})$, given in (12), is positive definite, since $f(r) > 0$ for $r < r_0$.

References

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

[2] ———, "Digital image processing: Toward Bayesian image analysis," *J. Appl. Stat.*, vol. 16, pp. 395–407, 1989.

[3] G. Bilbro, W. Snyder, and R. Mann, "Mean-field approximation minimizes relative entropy," *J. Opt. Soc. Amer.*, vol. 8, pp. 290–294, Feb. 1991.

[4] G. Bilbro, W. Snyder, S. Garnier, and J. Gault, "Mean-field annealing: A formalism for constructing GNC-like algorithms," *IEEE Trans. Neural Networks*, vol. 3, pp. 131–138, Jan. 1992.

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

[6] A. Blake, "Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 2–12, Jan. 1989.

[7] C. Bouman and B. Liu, "Segmentation of textured images using a multiple resolution approach," in *Proc. IEEE ICASSP*, Apr. 1988, pp. 1124–1127.

[8] ———, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Processing*, vol. 2, pp. 296–310, July 1993.

[9] C. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. Image Processing*, vol. 5, pp. 480–492, Mar. 1996.

[10] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing*, vol. 5, Dec. 1996.

[11] T. S. Chiang, C. R. Hwang, and S. J. Sheu, "Diffusion for Global Optimization in $\mathbb{R}^n$," *SIAM J. Contr. Optim.*, vol. 25, p. 737, May 1987.

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

[13] M. Duflo, *Algorithmes Stochastiques, Collection Mathématiques et Applications 23*. Berlin, Germany: Springer-Verlag, 1996.

[14] 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, May 1991.

[15] S. Gelfand and S. Mitter, "Recursive stochastic algorithms for global optimization in $\mathbb{R}^M$," *SIAM J. Contr. Optim.*, vol. 29, pp. 999–1018, Sept. 1991.

[16] ———, "Metropolis-type annealing for global optimization in $\mathbb{R}^M$," *SIAM J. Contr. Optim.*, vol. 31, pp. 111–131, Jan. 1993.

[17] 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, Nov. 1984.

[18] S. Geman and D. McClure, "Statistical methods for tomographic image reconstruction," *Proc. 46th Session of the ISI, Bull. ISI*, vol. 52, pp. 22–26, 1997.
[19] D. Geman, *Random Fields and Inverse Problems in Imaging: Lecture Notes in Mathematics*. Berlin, Germany: Springer-Verlag, 1990.
[20] D. Geman and G. Reynolds, "Constrained restoration and recovery of discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 367–383, Mar. 1992.
[21] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Processing*, vol. 4, pp. 932–946, July 1995.
[22] D. M. Grieg, B. T. Porteous, and A. H. Seheult, "Exact maximum *a posteriori* estimation for binary images," *J. R. Stat. Soc. B*, vol. 51, pp. 271–279, 1989.
[23] G. Herman, *Image Reconstruction from Projections. The Fundamentals of Computerized Tomography*. New York: Academic, 1980.
[24] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
[25] B. Hunt, "The application of constrained least squares estimation to image restoration by digital computer," *IEEE Trans. Commun.*, vol. COMM-22, pp. 805–812, 1973.
[26] F. Jeng and J. Woods, "Simulated annealing in compound Gaussian random fields," *IEEE Trans. Inform. Theory*, vol. 36, pp. 94–107, Jan. 1990.
[27] A. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. New York: IEEE Press, 1987.
[28] L. Le Cam, "Maximum likelihood: An introduction," *Int. Stat. Rev.*, vol. 58, pp. 153–171, 1990.
[29] Y. Leclerc, "Constructing simple stable description for image partitioning," *Int. J. Comput. Vis.*, vol. 3, pp. 73–102, 1989.
[30] S. Z. Li, "On discontinuity-adaptive smoothness priors in computer vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 576–586, June 1995.
[31] J. Marroquin, "Deterministic interactive particle models for image processing and computer graphics," *Comput. Vis., Graph., Image Process.*, vol. 55, pp. 408–417, 1993.
[32] R. Molina, A. Katsaggelos, J. Mateos, and A. Hermoso, *Restoration of Severely Blurred High Range Images Using Stochastic and Deterministic Relaxation Algorithms in Compound Gauss Markov Random Fields*, in M. Pelilo, E. R. Hancock, Eds, *Proc. EMMCVPR'97*.
[33] D. Mumford and J. Shah, "Boundary detection by minimizing functionals," in *Proc. IEEE ICASSP*, 1985, pp. 22–26.
[34] M. Nikolova, "Non convex regularization and the recovery of edges," in *Proc. IEEE Workshop on Nonlinear Signal and Image Processing*, June 1995, pp. 1042–1045.
[35] M. Nikolova, J. Idier, and A. Mohammad-Djafari, "Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF," *IEEE Trans. Image Processing*, vol. 8, pp. 571–585, Apr. 1999.
[36] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 629–639, July 1990.
[37] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C, the Art of Scientific Computing, 2nd ed.*. Cambridge, U.K.: Cambridge Univ. Press, 1992.
[38] N. Saito, "Superresolution of noisy band-limited data by data adaptive regularization and its application to seismic trace inversion," in *Proc. IEEE ICASSP*, Albuquerque, NM, Apr. 1990, pp. 1237–1240.
[39] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Processing*, vol. 41, pp. 534–548, Feb. 1993.
[40] T. Simchony, R. Chellappa, and Z. Lichtenchtein, "Graduated nonconvexity algorithm for image estimation using compound Gauss Markov field models," in *Proc. IEEE ICASSP*, 1989, pp. 1417–1420.
[41] W. Snyder *et al.*, "Image relaxation: Restoration and feature extraction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 620–624, June 1995.
[42] D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 413–424, July 1986.
[43] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC: Winston, 1977.
[44] E. Wasserstrom, "Numerical solutions by the continuation method," *SIAM Rev.*, vol. 15, pp. 89–119, Jan. 1973.
[45] C. Yang, "Efficient stochastic algorithms on locally bounded image space," *Comput. Vis., Graph., Image Process.: Graph. Models Image Process.*, vol. 55, pp. 494–506, Nov. 1993.

Mila Nikolova received the Ph.D. degree in signal processing from the Université Paris-Sud, Paris, France, in 1995.

She is an Assistant Professor at the Université René Descartes—Paris V. Her research interests are in inverse problems and reconstruction of images and signals.