J Math Imaging Vis 26: 277–291, 2006

DOI: 10.1007/s10851-006-0644-3

Image Restoration with Discrete Constrained Total Variation Part II: Levelable Functions, Convex Priors and Non-Convex Cases

JÉRÔME DARBON
EPITA Research and Development Laboratory (LRDE), 14-16 rue Voltaire F-94276 Le Kremlin-Bicêtre, France; École Nationale Supérieure des Télécommunications (ENST), 46 rue Barrault, F-75013 Paris, France
jerome.darbon@{lrde.epita.fr, enst.fr}

MARC SIGELLE
École Nationale Supérieure des Télécommunications (ENST) / LTCI CNRS UMR 5141, 46 rue Barrault, F-75013 Paris, France
marc.sigelle@enst.fr

Published online: 30 November 2006

Abstract. In Part II of this paper we extend the results obtained in Part I for total variation minimization in image restoration towards the following directions: first we investigate the decomposability property of energies on levels, which leads us to introduce the concept of levelable regularization functions (which TV is the paradigm of). We show that convex levelable posterior energies can be minimized exactly using the level-independant cut optimization scheme seen in Part I. Next we extend this graph cut scheme to the case of non-convex levelable energies. We present convincing restoration results for images corrupted with impulsive noise. We also provide a minimum-cost based algorithm which computes a global minimizer for Markov Random Field with convex priors. Last we show that non-levelable models with convex local conditional posterior energies such as the class of generalized Gaussian models can be exactly minimized with a generalized coupled Simulated Annealing.

Keywords: total variation, level sets, convexity, Markov Random fields, graph cuts, levelable functions

Abbreviations: KAP, Kluwer Academic Publishers; compuscript, Electronically submitted article

1. Introduction

Total variation (TV) is widely used as regularization term for image restoration purposes because of its edge preserving behavior [2, 19–22]. Data fidelity terms are often convex functions because it leads to convex optimization problems. We have dealt with such models in Part I of this paper.

In Part II of this paper, we extend the theoretical framework detailed in Part I for computing an exact and fast minimizer of the following problem

$$E_v(u) = \int_\Omega f(u(x),\ v(x))\, \mathrm{d}x + \beta \int_\Omega g(|\nabla u|)\, \mathrm{d}x. \quad (1)$$

Data fidelity term f is often a convex function and typically takes the following L^p form: $f\,(u(x), v(x))) = |u(x) - v(x)|^p$, $p \geq 1$. Laplace and Gaussian additive noise correspond to the particular cases $p = 1$ and $p = 2$, respectively.

Concerning the regularization term, the L^q form: $g(x) = |x|^q,\ q \geq 1$ is of great importance (with TV corresponding to $q = 1$), since it is widely used for image restoration [18]. This total variation case as well as the L^1+ TV and L^2+ TV cases were thoroughly investigated in Part I. Our approach relies on reformulating this problem within a *discrete* framework into independent binary Markov Random Fields (MRFs) attached to each level set of an image. Exact minimization is performed thanks to a minimum cost cut algorithm [12].

In this paper we extend these results towards the following directions:

First we generalize the decomposability property of energies on the level sets, which leads us to introduce the concept of *levelable* functions. We prove that *convex* levelable posterior energies can be minimized exactly using the level-independent graph cut optimization scheme seen in Part I. During the revision of this paper, we became aware of the work of Zalesky on submodular energy functions described in [23]. Zalesky studies the same class of functions as us. He gives conditions on these energies so that they can be exactly optimized via a modified submodular function minimization algorithm [15]. However, no numerical results are presented.

Then we extend this optimization scheme to *non-convex* levelable energies but with convex priors. In this case we provide a graph construction such that its minimum cost cut determines a global minimizer. No conditions are set on data fidelity terms. Ishikawa [14] solves exactly the same problem using a minimum cost cut based approach. The main difference between the two methods consists in the different graph structures obtained. Experiments show both convincing restoration results for images corrupted with impulse noise and a quite faster minimum cost cut-based algorithm applied on our graph than on Ishikawa's graph. Approaches which compute only approximate solutions are found in [3, 4, 8, 10].

Last we show that a large subset of non-levelable convex posterior energies can be also minimized with a stochastic coupled Simulated Annealing algorithm. This applies for instance to the widely used generalized Gaussian [5] $L^p + L^q$ models (with $p, q \geq 1$).

The rest of this paper is organized as follows. In Section 2, we recall the main results obtained in Part I of this paper which will be useful afterwards. Then in Section 3, we introduce and develop the main properties of levelable functions In Section 4 we show the graph-cut property of non-convex levelable energies thanks to a nice constrained scheme and we present some results. Last in Section 5, we extend all previous schemes to fairly general cases. For sake of clarity, the main demonstrations of theorems have been placed in related Appendices at the end of this Part II.

2. Recalls

In this section let us recall the main results obtained in Part I that will be used in the sequel of this work. We assume that u is defined on a finite discrete lattice S and takes values in the discrete integer set $\mathcal{L} = [0, L-1]$. We denote by u_s the value of the image u at the site $s \in S$. An image is decomposed into its level sets using the decomposition principle [13]. It corresponds to considering all thresholding images u^λ where $u_s^\lambda = \mathbb{1}_{u_s \leq \lambda}$. Note the original image can be reconstructed from its level sets using $u_s = \min\{\lambda, u_s^\lambda = 1\}$ as shown in [13]. Recall that we note by $s \sim t$ the neighboring relationship between sites s and t, by (s, t) the related clique of order two and by N_s the local neighborhood of site s. For simplicity purpose we shall in the sequel write sums on cliques of order one and two by $\sum_s$ and $\sum_{(s,t)}$ respectively.

Proposition 1. *The discrete version of energy* (1) *rewrites as follows*

$$E_v(u) = \sum_{\lambda=0}^{L-2} E^\lambda(u^\lambda) + C, \quad \textit{where} \tag{2}$$

$$E^\lambda(u^\lambda) = \beta \left[\sum_{(s,t)} w_{st}\big((1 - 2u_t^\lambda)u_s^\lambda + u_t^\lambda\big) \right] + \sum_s (g_s(\lambda+1) - g_s(\lambda))\big(1 - u_s^\lambda\big) \tag{3}$$

$$g_s(x) = f(x,\ v_s) \quad \forall s \in S \quad \textit{and} \quad C = \sum_s g_s(0).$$

Note that each $E_v^\lambda(\cdot)$ is a binary MRF with an Ising prior model. We endow the space of binary configurations by the following order: $a \preceq b$ iff $a_s \leq b_s\ \forall s \in S$. In order to minimize $E_v(\cdot)$ one would like to minimize all $E_v^\lambda(\cdot)$ independently. Thus we get a family $\{\hat{u}^\lambda\}$ which are respectively minimizers of $E_v^\lambda(\cdot)$. Suppose we do so, then clearly the summation will be minimized and thus we have a minimizer of $E_v(\cdot)$ provided this family is monotone, i.e.,

$$\hat{u}^\lambda \preceq \hat{u}^\mu \Leftrightarrow \hat{u}_s^\lambda \leq \hat{u}_s^\mu \quad \forall \lambda \leq \mu, \quad \forall s \in S. \tag{4}$$

If this property holds then the optimal solution is given by the reconstruction formula from level sets [13]: $\hat{u}_s = \min\{\lambda, \hat{u}_s^\lambda = 1\}\ \forall s$. Else the family $\{u^\lambda\}$ does not define a function, and thus our optimization scheme is no more valid.

The following Lemma is absolutely crucial for our point.

Lemma 1. *If the local conditional posterior energy at each site s can be written as*

$$E_v(u_s \mid N_s) = \sum_{\lambda=0}^{L-2} \big(\Delta\phi_s(\lambda)\, u_s^\lambda + \chi_s(\lambda)\big), \tag{5}$$

where $\Delta\phi_s(\lambda)$ is a non-increasing *function of λ and $\chi_s(\lambda)$ is a function which does not depend on u_s^λ, then one can exhibit a "coupled" stochastic algorithm minimizing each total posterior energy $E_v^\lambda(u^\lambda)$ while preserving the monotone condition: $\forall s, u_s^\lambda$ is non-decreasing with λ.*

This Lemma states that given a binary solution $a^\star$ to the problem $E_v^\lambda(\cdot)$, there exists at least one solution $\hat{b}$ to the problem $E_v^\mu(\cdot)$ such that $a^\star \preceq \hat{b}\ \forall\lambda \leq \mu$.

We also recall that a one-dimensional discrete function f defined on $]A, B[$ is *convex* on $]A, B[$ iff $2f(x) \leq f(x-1) + f(x+1)\ \forall x \in]A, B[$, or equivalently, $f(x+1) - f(x)$ is a *non-decreasing* function on $[A, B[$.

The following important equivalence was then established:

Lemma 2. *The requirements stated by this Lemma are equivalent to these: all conditional energies $E_v(u_s|N_s)$ are* convex *functions of grey level $u_s \in]0, L-1[$, for any neighborhood configuration and local observed data.*

It is indeed easy to prove that

$$\begin{aligned}\Delta\phi_s(\lambda) &= E_v^\lambda\big(u_s^\lambda = 1|N_s^\lambda\big) - E_v^\lambda\big(u_s^\lambda = 0|N_s^\lambda\big)\\ &= -(E_v(\lambda+1|N_s) - E_v(\lambda|N_s)).\end{aligned} \quad (6)$$

As a consequence Lemma 1 applies for both the models convex data fidelty with TV as prior.

3. Levelable Regularization Energies

In this section we propose an extension of the decomposition property of posterior energies on levels previously seen for the TV case in view of minimization purposes. To this aim we first introduce the concept of *levelable* functions of several variables. Then we show their decomposition properties on binary MRFs.

3.1. A Deductive Construction of Decomposable Energies

We first introduce our concept of *levelable* functions.

Definition 1. A function ϕ of one or several variables $x, y \ldots \in \mathcal{L}$ is *levelable* if it can be decomposed as a sum on levels, of functions of its variables level-set indicatrices at current level:

$$\phi(x, y\ldots) = \sum_{\lambda=0}^{L-1} \psi(\lambda,\ \mathbb{1}_{\lambda<x}, \mathbb{1}_{\lambda<y}\ldots).$$

We now give the main properties of levelable functions.

Proposition 2. *Every function of a single variable is levelable.*

Proof: it relies on a straightforward "discrete integration theory" result:

$$\begin{aligned}\forall k \in \mathcal{L}\ \ \phi(k) &= \sum_{\lambda=0}^{k-1}(\phi(\lambda+1) - \phi(\lambda)) + \phi(0)\\ &= \sum_{\lambda=0}^{L-2}(\phi(\lambda+1) - \phi(\lambda))\ \mathbb{1}_{\lambda<k} + \phi(0).\end{aligned} \quad (7)$$

The summation runs indeed up to $\lambda = L-2$ and yields coherent results for both $k = 0$ and $k = L-1$, since $\forall k \in \mathcal{L},\ \mathbb{1}_{\lambda<k} = 0$ when $\lambda = L-1$. □

The following proposition, whose proof is given in Appendix A, connects a levelable total Gibbs energy with each of its cliques components.

Proposition 3. *The following two assumptions are equivalent:*

A 1. *The total Gibbs energy function is levelable.*

A 2. *Each clique energy is a levelable function.*

Since we shall be concerned by cliques of order two at most in the sequel, we give the closed form of clique of order two *symmetric* levelable energies.

Proposition 4. *A* symmetric *function of two variables $U(x, y)$ is levelable* iff *it writes in both equivalent ways:*

$$\begin{aligned}&U(x, y)\\ &\quad = S(\max(x, y)) - S(\min(x, y)) + D(x) + D(y) &(8)\\ &\quad = F(\max(x, y)) - G(\min(x, y)), &(9)\end{aligned}$$

where $S, D, F = S+D$ and $G = S-D$ are functions mapping $\mathcal{L} \mapsto I\!R$.

See proof in Appendix B.

We now characterize the behavior of cliques of order two potentials under mild conditions (from an image processing point of view).

Proposition 5. *If a symmetric function of two variables $U(x, y)$ satisfies both assumptions:*

A 3. *$U(x, y)$ is a levelable function of x, y.*

A 4. *$\forall y \in \mathcal{L}$, $U(x, y)$ attains a minimum at $x = y$.*

then the functions F, G and $S = (F + G)/2$ defined in Proposition 4 should be increasing *on $\mathcal{L}^* = [0, L-2]$. The related energy writes in this case:*

$$\begin{aligned} U(x, y) &= |S(x) - S(y)| + D(x) + D(y) \\ &= \sum_{\lambda=0}^{L-2} R(\lambda) \mid \mathbb{1}_{\lambda<x} - \mathbb{1}_{\lambda<y} \mid + D(x) + D(y), \end{aligned} \tag{10}$$

where $R(\lambda) = S(\lambda + 1) - S(\lambda)$ is a positive *function on $\mathcal{L}^*$.*

See proof in Appendix C. In other words, $\forall y \in \mathcal{L}$, one has from (9):

$$U(x, y) = \begin{cases} F(y) - G(x) \text{ is a decreasing function of } x & \text{for } x \leq y \\ F(x) - G(y) \text{ is an increasing function of } x & \text{for } x \geq y \end{cases} \tag{11}$$

$U(\cdot, y)$ is thus a *quasi-convex* function [6, 11] attaining its minimum in y. A function f is said quasi-convex iff its lower level-sets are convex sets. In 1D this is equivalent to the monotony property by intervals (11).

Examples of such quasi-convex functions are shown in Fig. 1.

As a further consequence the set of *convex* levelable energy functions satisfying $D(x) = 0$ i.e., such that $\min_x U(x, y) = 0 \; \forall y \in \mathcal{L}$, reduces to: $S(x) = \beta x$, since in this case by (11) $S(x) = F(x) = G(x)$ should be both convex and concave on $]0, L-1[$. We find again the importance of the TV family.

Let us also stress again that in general no condition is imposed on function $D(x)$ excepted that $S \pm D$ should be increasing functions on $\mathcal{L}$.

3.2. *On the Structure of Levelable MRF Posterior Energies*

Let us assume for sake of generality levelable pairwise regularization energies satisfying Propositions 4 and 5 in some Γ-connectivity.

We shall note γ_i the number of neighbors of type $i \in [1, I]$, $\tau(s, t) = i$ the type of clique (s, t) and $U_i(u_s, u_t)$ the associated regularization energy. This is an extension of the case $U_i(u_s, u_t) = w_{st} \, |u_s - u_t|$ previously seen in Part I (where w_{st} should indeed be noted w_i). Each single term $D_i(u_s)$ appearing in all energy terms $U_i(u_s, u_t)$ such that $t \sim s$, $\tau(s, t) = i$ will thus be accounted γ_i times in the total regularization energy. We therefore define $D(u_s) = \sum_{i=1}^{I} \gamma_i \, D_i(u_s)$, so that the following proposition holds:

Proposition 6. *The general form of a total posterior levelable energy satisfying the prerequisites of Proposition 5 is:*

$$\begin{aligned} E_v(u) &= \sum_{i=1}^{I} \Bigg[\sum_{\substack{(s,t) \\ \tau(s,t)=i}} |S_i(u_s) - S_i(u_t)| + D_i(u_s). \\ &\qquad + D_i(u_t) \Bigg] + \sum_s g_s(u_s) && (12) \\ &= \sum_{\lambda=0}^{L-2} \Bigg\{ \sum_{i=1}^{I} \sum_{\substack{(s,t) \\ \tau(s,t)=i}} R_i(\lambda) \left| u_s^\lambda - u_t^\lambda \right| \\ &\qquad + \sum_s \delta(\lambda, v_s) \left(1 - u_s^\lambda\right) \Bigg\} + \tilde{C}, && (13) \end{aligned}$$

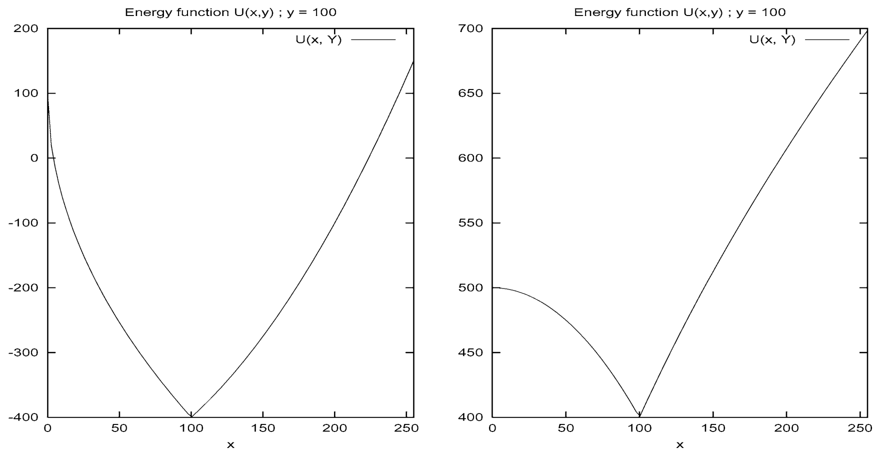
where

- *constant $\tilde{C} = \sum_s (g_s(0) + D(0))$,*
- *regularization coefficient term at level λ: $R_i(\lambda) = S_i(\lambda + 1) - S_i(\lambda)$ is a* positive *function on $\mathcal{L}$ $\forall i \in [1, I]$,*
- *and* effective *attachment to data coefficient term $\delta(\lambda, v_s)$ is: $\delta(\lambda, v_s) = D(\lambda + 1) - D(\lambda) + g_s(\lambda + 1) - g_s(\lambda)$.*

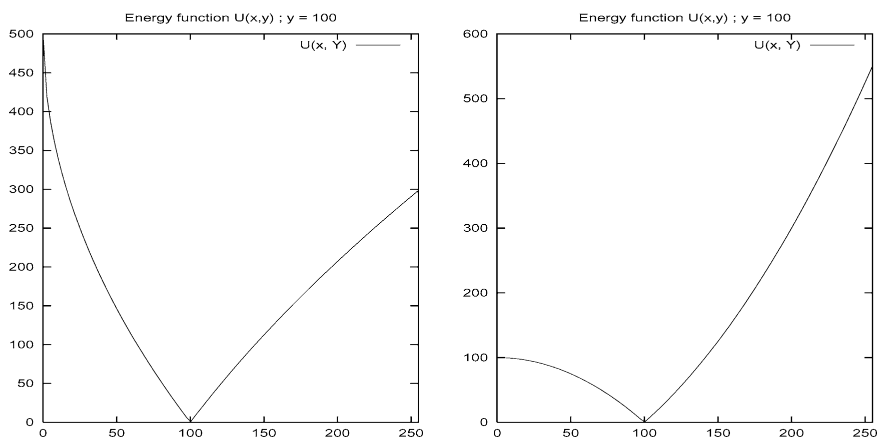
Since $|\mathbb{1}_{\lambda<u_s} - \mathbb{1}_{\lambda<u_t}| = |u_s^\lambda - u_t^\lambda| \;\; \forall u_s, u_t \in \mathcal{L}$, the proof results straightforwardly from Propositions 1 and 5.

Note again that we set no conditions on functions $D(u_s)$ and $g_s(u_s)$ for now. Now from (13), all local conditional posterior energies are levelable:

$$E_v(u_s | N_s) = \sum_{\lambda=0}^{L-2} E_v^\lambda\left(u_s^\lambda \middle| N_s^\lambda\right), \quad \text{with} \tag{14}$$



a) left: $F(x) = x^2/100 \quad G(x) = 50\sqrt{x}$
- right: $F(x) = 50\sqrt{x} \quad G(x) = x^2/100$



b) left: $F(x) = G(x) = 50\sqrt{x}$ - right: $F(x) = G(x) = x^2/100$

Figure 1. $U(x, y) = F(\max(x, y)) - G(\min(x, y))$ as a function of x $(y = 100)$.

$$E_v^\lambda\big(u_s^\lambda \big| N_s^\lambda\big) = \sum_{i=1}^{I} \sum_{\substack{(s,t)\\ \tau(s,t)=i}} R_i(\lambda)\big|u_s^\lambda - u_t^\lambda\big| + \sum_s \delta(\lambda, v_s)\big(1 - u_s^\lambda\big). \tag{15}$$

From all what precedes we obtain the following Proposition:

Proposition 7. *If the total posterior energy $E_v(u)$ is* levelable *and such that* all *local conditional posterior energies $E_v(\cdot \mid N_s)$ are* convex *functions (even if their regularization or observation part may not), then the same level-by-level minimization scheme of Part I applies for global optimization.*

In [23], Zalesky studies the general form of levelable function while we have focused on levelable energies dedicated to image processing. In next section we treat the minimization of *non-convex* levelable energies.

4. Minimization Algorithms for Levelable Energies

The previous level-independent minimization scheme can no more hold for non-convex levelable energies (non-convexity could arise from the attachment to data as well as from the regularization energies). We are thus led to devise a graph construction such that its minimum cost cut provides a global minimizer. We would to emphasize that Zalesky proposes in [23] a minimization algorithm which relies on submodular function minimization [15], while we propose an algorithm which relies on minimum cuts. In the specific total variation case we compare our graph structure to the one proposed by Ishikawa in [14].

4.1. Reformulation of the Total Levelable Posterior Energy

First we drop out constant $\tilde{C}$ in Eq. (13) and write the total posterior energy as a functional of the $(L-1)$ binary images u^λ:

$$E_v(\{u^\lambda\}) = \sum_{\lambda=0}^{L-2} \left\{ \sum_{i=1}^{I} \sum_{\substack{(s,t)\\ \tau(s,t)=i}} R_i(\lambda)\left|u_s^\lambda - u_t^\lambda\right| + \sum_s \delta(\lambda, v_s)\left(1 - u_s^\lambda\right) \right\}, \qquad (16)$$

where $R_i(\cdot)$ is a *positive* non-decreasing function on $\mathcal{L}$. As the convexity assumption no more holds, the monotony condition (4): $\forall s\, \forall \lambda \le \mu\ u_s^\lambda \le u_s^\mu$ can thus no more be guaranteed. Since this condition is equivalent to

$$\forall s\, \forall \lambda \in \mathcal{L}^*\ u_s^\lambda \le u_s^{\lambda+1}, \qquad (17)$$

the minimization of (16) is equivalent to the following constrained problem:

$$\min_{\{u^\lambda\}} \ E_v(\{u^\lambda\}) \text{ subject to constraints}$$
$$\forall s \forall \lambda \in \mathcal{L}^*\ H\left(u_s^\lambda - u_s^{\lambda+1}\right) \le 0, \qquad (18)$$

where $H : I\!R \mapsto I\!R$ is the Heaviside function defined as

$$H(x) = \begin{cases} 0 & \text{if } x \le 0, \\ 1 & \text{else.} \end{cases} \qquad (19)$$

We define for this purpose the new unconstrained energy $E_{\alpha,v}(\{\cdot^\lambda\})$ as:

$$E_{\alpha,v}(\{u^\lambda\}) = E_v(\{u^\lambda\}) + \alpha \sum_{\lambda=0}^{L-2} \sum_s H\left(u_s^\lambda - u_s^{\lambda+1}\right) \quad \text{with} \quad \alpha > 0. \qquad (20)$$

4.2. Graph-Cut Representation and Minimization

In [16], Kolmogorov *et al.* gives sufficient and necessary conditions for a fonction to be minimized via graph cuts. Such functions are also called *submodular* [16]. We can now formulate the following proposition:

Proposition 8. *The unconstrained functional $E_{\alpha,v}(\{u^\lambda\})$ defined by* (20) *is submodular, i.e., graph-representable in the sense of Kolmogorov et al.* [16].

Since each function of a single binary variable is graph-representable it is enough to prove the regularity property (defined in [16]):

$$A(0,0) + A(1,1) \le A(0,1) + A(1,0)$$

for each function $A(\cdot,\cdot)$ of two binary variables appearing in (20).

See proof in Appendix D.

We now show in which conditions the minimization of (20) can lead to a minimizer of (13). We first notice the following Proposition:

Proposition 9. *Suppose that $\{\hat{u}^\lambda\}$ is a minimizer of $E_{\alpha,v}(\{\cdot^\lambda\})$ such that $\forall s\, \forall \lambda\ \ \hat{u}_s^\lambda \le \hat{u}_s^{\lambda+1}$. Then the image $\hat{u} \in \mathcal{L}^S$ defined as:*

$$\forall s\ \ \hat{u_s} = \min\left\{\lambda, u_s^\lambda = 1\right\},$$

is a minimizer of $E_v(\cdot)$.

See proof in Appendix E.

We now investigate the cases where the hypotheses of Proposition 9 always hold. To this aim we exhibit a finite positive value for α such that the monotony constraint (17) holds.

Proposition 10. *Let $\alpha > \Gamma \sup_{\lambda,i} R_i(\lambda) - \inf_{\lambda \ge 1, v_s} \delta(\lambda, v_s)$, where Γ is the total connectivity. Then a minimizer $\{\hat{u}^\lambda\}$ of $E_{\alpha,v}(\{\cdot^\lambda\})$ verifies the inequality*

$$\forall s\, \forall \lambda\ u_s^\lambda \le u_s^{\lambda+1}.$$

See proof in Appendix F.

The value of α could in fact be adjusted locally for each site s. Experiments show indeed that α needs not to be so high.

In other words, we have shown that a levelable function with every potential satisfying conditions of Proposition 5 is exactly minimizable via a minimum cost cut. In terms of submodularity, it means that such a function is a sum of submodular functions over all level sets.

4.3. Notes on the Graph Construction

We now consider the case of the total variation minimization. We briefly compare the approach proposed by Ishikawa [14] which also solves exactly this problem and ours. The Ishikawa's approach also consists in building a graph on which a minimum cost cut is computed. The structure of this graph is composed of layers. Each layer corresponds to a grey level. More precisely, a node is created for each value that a pixel can take. The node n_s^l which corresponds to the pixel at site s which takes the value l is connected to spatial neighbors at same level, i.e., to nodes n_t^l with $t \sim s$. It is also connected to nodes n_s^m with $|m - l| = 1$. The nodes located at the first (respectively last) layers are connected to the source (sink respectively). The structure of this graph structurally imposes the monotone property to hold. The Ishikawa's graph is depicted in Fig. 2(a).

Contrary to the graph of Ishikiwa, our graph is not made of layers. However, it contains the same number of nodes and arcs. A node is created for each binary variable u_s^λ. Each of these nodes is connected to the source, the sink, the nodes u_t^λ with $s \sim t$, and the nodes u_s^μ with $\mu = \lambda \pm 1$. Compared to the graph of Ishikiwa, the main difference is that all nodes are connected to the source and to the sink. The minimal path from the source to the sink is equal to 2 for our graph and to L for the Ishikiwa's one. The graph we build is depicted in Fig. 2(b)). A main class of minimum-cut/maximum flow algorithms are based on the augmenting path approach [1]. Such algorithms work iteratively and look for a path from the source to the sink where the flow can be increased. Since our graph is more compact (in the sense that each node is connected to both the source and the sink) it is reasonable to conjecture that our graph is better suited for augmented path based algorithms. These considerations are confirmed by the numerical experiments presented in the next section.

4.4. Experiments and Discussion

Our implementation relies on the graph construction proposed by Kolmogorov *et al.* in [16] and we use the graph cut algorithm described by Boykov *et al.* in [7]. For all experiments, we have checked that the minimizers computed by the Ishikawa's approach [14] and our algorithm have the same energy. We show here our associated restoration results when the image is corrupted by impulsive noise of parameter p. The attachment to data term is thus:

$$g_s(u_s) = f(u_s, v_s) = \begin{cases} -\ln\left((1-p) + \dfrac{p}{L}\right) & \text{if } u_s = v_s, \\ -\ln \dfrac{p}{L} & \text{else.} \end{cases}$$

This implies that the total energy $E_v(\cdot)$ is highly *nonconvex*. We used a 3 GHz Pentium4 with 1024 Kbytes

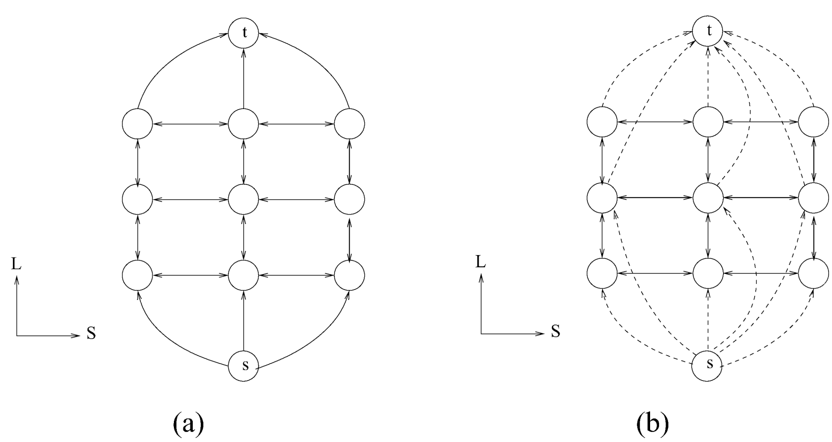


Figure 2. The graph constructed with the Ishikawa approach is depicted in (a) while our graph is presented in (b). We consider a one dimensional signal with 3 sites ($|S| = 3$) with 3 available labels ($|L| = 3$). The source and the sink are respectively denoted by s and t. Note that all nodes are connected to the source s and to the sink t for our graph although all arcs are not depicted in b).

Table 1. CPU times (in seconds) for total variation minimization with an impulsive noise of parameter p and for 256 × 256 size images. The gain factor obtained by our algorithm (Levelable time wrt. Ishikawa time) is also shown.

Image	Impulsive p	Levelable	Ishikawa	Gain
Lena	0.20	114.78	425.14	3.70
Lena	0.40	159.14	633.09	3.98
Lena	0.70	252.67	1203.22	4.76
Girl	0.20	114.09	469.44	4.11
Girl	0.40	171.68	648.72	3.78
Girl	0.70	272.72	1553.66	5.70

cache memory for our tests. Each experiment has been repeated 30 times, using the 4-connectivity approximation of the total variation. Average CPU times for *lena* and *girl* images are given in Table 1 for both our algorithm and Ishikawa's (image size: 256 × 256). Time results show that our graph construction yields a gain factor about 4 compared to Ishikawa's method.

Associated denoised images for the image *girl* are presented in Fig. 3 for noise parameter p with respective values 20%, 40% and 70%. In each of these cases the regularization parameter β was chosen in order to yield the best visual result. We also observe on Figs. (b) and (d) that minimizers have some pixels which are not regularized enough. We verify experimentally that these pixels do not change during the algorithm. This is to compare with observations of Nikolova in [17, 18] for the detection of outliers. In Fig. 4. we compare to the results from the (sub-optimal) direct random descent. Its principle is the following: pick randomly a site, and a new gray level value. If this new value for this pixel makes decrease the energy then this transition is accepted, otherwise the pixel keeps its value. We iterate this process until no moves are allowed. To conclude at this point, our graph cut method yields very good results visually speaking, but comparison with other methods such as the morphological approach of Coupier *et al.* [9] remains to be done.

5. Extension to Convex Non-Levelable Regularization Energies

In this section we proceed even further by extending the previous results to non-levelable regularization energies provided they have a *convex* form, while no conditions are imposed on data fidelity terms. As an example, this section shows how to find a global minimizer of energies which has the following form [18]:

$$E_v(u)=\sum_{s\in S} j_s|u_s-v_s|^p+\beta\sum_{(r,s)} b_{rs}|u_s-u_t|^q \quad p,q\geq 1,$$

where $j_s \geq 0$ is a local "outlier" detector and $b_{rs} \geq 0$ is a local boundary detector can be exactly minimized. First of all, we present an algorithm based on graph-cuts which computes an exact minimizer. Note that in [14], Ishikawa solves exactly the same problem with the use of graph-cuts. However, the graphs built are quite different from each others. We also propose an alternative condition for devising stochastic optimization.

5.1. *Reformulation of the Total Posterior Energy*

Our approach for exact minimization is similar to that previously encountered for the levelable case. Recall that we assume that priors are *convex* functions. Moreover we shall assume in the sequel that each regularization potential (noted $U_{st}(u_s, u_t)$ for sake of generality) is also *symmetric* and depends on the *difference* of its two grey level variables as in Eq. (1), i.e.,

$$U_{st}(u_s,u_t) = V_{st}(u_s-u_t) = V_{st}(u_t-u_s).$$

These assumptions imply the convexity of each $V_{st}(\cdot)$ on $]-L+1, L-1[$.

Now, proceeding similarly to paragraph 4.1 we first reformulate the total posterior energy as a functional of the $\mathcal{L}-1$ binary images u^λ.

Proposition 11. *The total posterior energy* (1) *can be rewritten as:*

$$\begin{aligned} E_v(\{u^\lambda\}) = \sum_{(s,t)} \Bigg\{ &\sum_{\mu=0}^{L-2}\sum_{\lambda=0}^{L-2} U_{st}^{\lambda,\mu}\big(u_s^\lambda, u_t^\mu\big) \\ &+ \sum_{\lambda=0}^{L-2}\big[D_{st}^\lambda\big(u_s^\lambda\big) + D_{st}^\lambda\big(u_t^\lambda\big)\big]\Bigg\} \\ &+ \sum_s\sum_{\lambda=0}^{L-2}\Delta_s^\lambda\big(u_s^\lambda|v\big) + C, \quad \textit{where} \end{aligned} \tag{21}$$

$$\begin{aligned} U_{st}^{\lambda,\mu}\big(u_s^\lambda, u_t^\mu\big) &= G_{st}(\lambda,\mu)\big(1-u_s^\lambda\big)\big(1-u_t^\mu\big), \\ \textit{with}\quad G_{st}(\lambda,\mu) &= 2V_{st}(\lambda-\mu) - V_{st}(\lambda-\mu-1) \\ &\quad - V_{st}(\lambda-\mu+1) \leq 0, \\ D_{st}^\lambda\big(u_s^\lambda\big) &= (V_{st}(\lambda+1)-V_{st}(\lambda))\big(1-u_s^\lambda\big), \\ \Delta_s^\lambda\big(u_s^\lambda|v\big) &= (g_s(\lambda+1)-g_s(\lambda))\big(1-u_s^\lambda\big), \\ \textit{and}\quad C &= \sum_s g_s(0). \end{aligned}$$

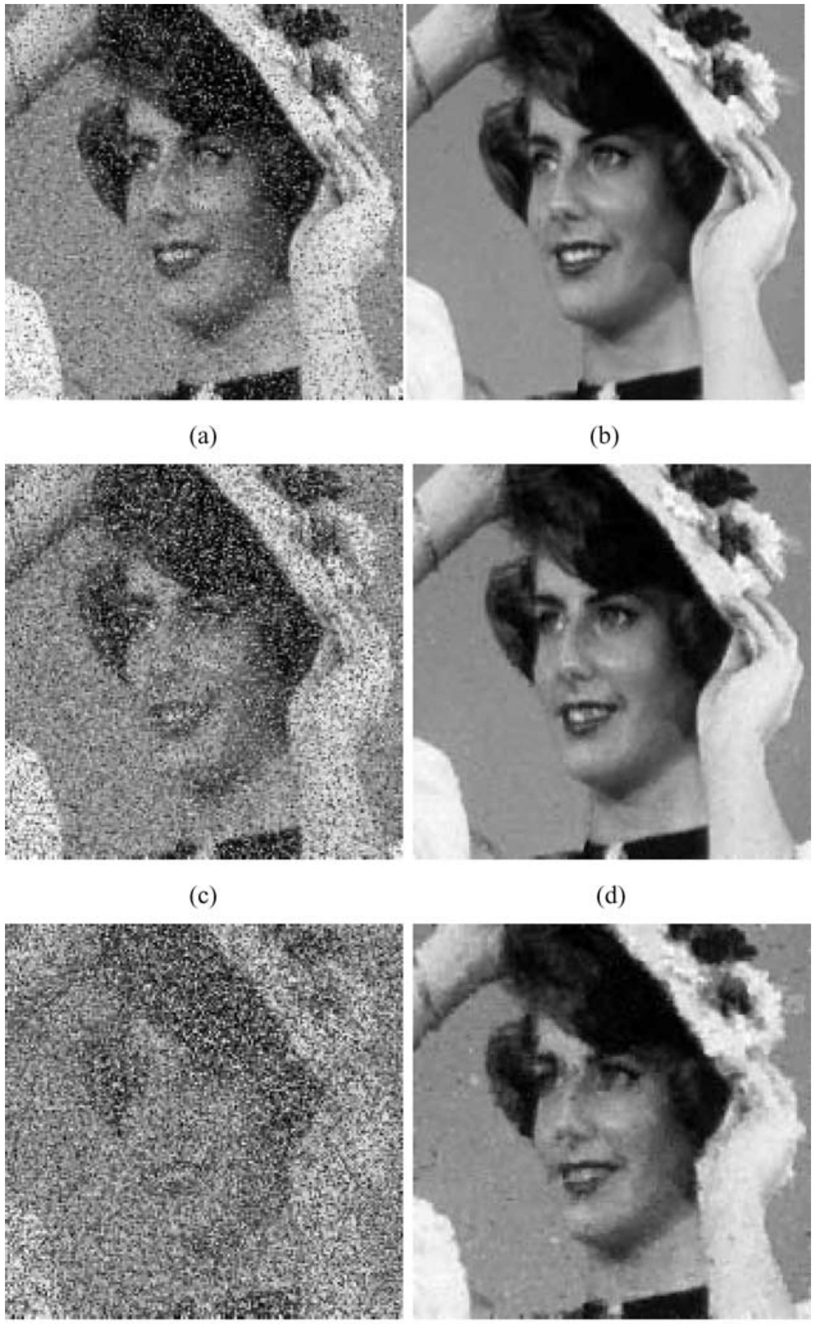

(a) (b) (c) (d) (e) (f)

Figure 3. Image *girl* corrupted with impulse noise. Noise parameter value: (a): 20%, (c): 40%, (e): 70%. Associated respective restoration results: (b), (d), (f).

Figure 4. Results of the restoration of the image *girl* corrupted with impulse noise using a direct random descent algorithm. Figures (a), (b) and (c) correspond to minimizer of images corrupted with an impulse noise of parameter 20%, 40% and 70%, respectively.

The principle relies on decomposing each regularization potential as two sums over grey level variables instead of one. See Proof in Appendix G.

5.2. Graph-Cut Representation and Minimization

The proposed approach is similar to the one presented for levelable energies. We define the constrained total posterior energy as: $E_{\alpha,v}(\cdot)$

$$E_{\alpha,v}(\{u^\lambda\}) = E_v(\{u^\lambda\}) + \sum_s \sum_{\lambda=0}^{L-2} \alpha H\left(u_s^\lambda - u_s^{\lambda+1}\right) \quad \text{with } \alpha > 0, \qquad (22)$$

where H is the Heaviside function defined by Eq. (19). The purpose of this last term is, as before, to enforce the monotone property Eq. (4) to hold i.e.,

$$\forall \lambda \; u_s^\lambda \leq u_s^{\lambda+1}.$$

The fact that regularization terms $\{V_{st}\}$ are *convex* functions is of capital importance for the minimization algorithm that we present now. Notice that results hold even for non-symmetric convex regularization energies.

Proposition 12. *Let α be a non-negative value. The energy $E_{\alpha,v}(\cdot)$ in Eq.* (22) *is submodular, i.e., it is graph-representable in the Kolmogorov et al.* [16] *sense.*

The main argument relies on the fact that each function of two binary variables in Eq. (21) is graph-representable since all terms $G_{st}(\lambda, \mu) \leq 0$ by the *convexity* hypothesis. See Proof in Appendix H.

We now build a graph associated to $E_{\alpha,v}(\{\cdot^\lambda\})$ such that its minimum cost cut yields a global minimizer. We show that there exists a finite value for α such that we can construct a minimizer for $E_v(\cdot)$ from a minimizer for $E_{\alpha,v}(\{\cdot^\lambda\})$.

Proposition 13. *Let α be a positive value such that*

$$\alpha > \sup_{(s\sim t),\lambda,\mu} \left|U_{st}^{\lambda,\mu}\left(u_s^\lambda, u_t^\mu\right)\right| + \Gamma \sup_{(s\sim t),\lambda} \left|D_{st}^\lambda\left(u_s^\lambda\right)\right| + \sup_{s,\lambda} \left|\Delta_s^\lambda\left(u_s^\lambda|v\right)\right|,$$

where Γ is the the total connectivity. Then a minimizer $\{\hat{u}^\lambda\}$ of $E_{\alpha,v}(\{\cdot^\lambda\})$ satisfies the inequality

$$\forall s\, \forall \lambda\ u_s^\lambda \leq u_s^{\lambda+1}.$$

Proof: The proof is similar to the one given for Propositions 9 and 10. □

This last proposition shows that a Markov Random Field with convex regularization terms can be globally optimized using minimum cost cuts. The same result is proposed by Ishikawa in [14].

In the next paragraph we show that the coupled Simulated Annealing investigated in Part I applies as well to this convex non-levelable case.

5.3. Stochastic Minimization

Since we assume that the levelable hypothesis no more holds, we have now to keep the general dependence wrt. neighborhood grey variables $\{u_t\}_{t\sim s}$ in each local posterior conditional energy $E_v(\cdot|\, N_s)$. However we claim the following as well, which is essentially the same as Lemma 2:

Proposition 14. *A necessary and sufficient condition that the monotony conditions hold is that all local conditional energies $E_v(\cdot \mid N_s)$ are* convex *functions for any neighborhood configuration N_s and local observed data v_s.*

See proof in Appendix I.

We can thus devise the same coupled Gibbs sampler and Simulated Annealing that was devised in Part I for the exact minimization of total energy. This concludes the stochastic approach to this work and to the work of [14].

6. Conclusion

In this paper we have presented an algorithm which computes an exact solution for the minimization of the total variation under a convex constraint. The method relies on the decomposition of the problem into binary ones using the level sets of an image. Moreover, this algorithm is quite fast. In the second Part of this paper we have extended these results to the following trends: first we have developed the concept of *levelable* functions which generalizes the decomposability property of total variation on level sets. Then we have shown that convex levelable posterior energies can be exactly minimized as in Part I. Then we have extended our graph-cut optimization scheme to the case of non-convex levelable energies with impulsive noise restoration as an example. Further comparison to existing methods remains to be done in this case. Last we proved more generally that all convex models of the form

$$E_v(u) = \sum_{s\in S} j_s|u_s - v_s|^p + \beta \sum_{(r,s)} b_{rs}|u_s - u_t|^q \qquad p, q \geq 1,$$

where $j_s \geq 0$ is a local "outlier" detector and $b_{rs} \geq 0$ is a local boundary detector can be exactly minimized. We have also proposed an exact optimization algorithm based on minimum cost cut for MRFs with convex priors.

A fast algorithm based on graph cuts for exact optimization of MRFs with both convex data fidelity terms and priors will be presented in a forthcoming paper. We also work on using this formulation for vectorial images. We are currently investigating a class of models enjoying the perfect sampling property. This might in turn pave the way for fast hyperparameter estimation, a problem still of importance for the variational community.

Appendix

A. Proof of Proposition 3

Case b) ⇒ a). The set of levelable functions is clearly a vector space.

Case a) ⇒ b). We shall be concerned by cliques of order two at most in the sequel, however the proof is easily established by induction to any finite order clique energies. Assume the total energy decomposes as

$$\phi(u_s,\ u_t,\ u_r \ldots) = \sum_{(s,t)} U(u_s,\ u_t) + \sum_s T(u_s) = \sum_{\lambda=0}^{L-2} \psi(\lambda,\ \mathbb{1}_{\lambda<u_s},\ \mathbb{1}_{\lambda<u_t},\ \mathbb{1}_{\lambda<u_r}, \ldots)$$

Assign $u_r = 0\ \forall r \notin \{s,\ t\}$ in previous formula. One gets

$$\begin{aligned} U(u_s,\ u_t) + T(u_s) + T(u_t) + C \\ = \sum_{\lambda=0}^{L-2} \underbrace{\psi(\lambda,\ \mathbb{1}_{\lambda<u_s},\ \mathbb{1}_{\lambda<u_t},\ 0,\ 0\ldots)}_{\downarrow} \\ = \sum_{\lambda=0}^{L-2} \psi_{st}(\lambda,\ \mathbb{1}_{\lambda<u_s},\ \mathbb{1}_{\lambda<u_t}). \end{aligned}$$

Since by Proposition 2 single variable functions $T(u_s)$ and $T(u_t)$ are levelable, the proof is established for clique of order two energy $U(u_s,\ u_t)$. □

B. Proof of Proposition 4

- Sufficient - Let us assume that (8) holds. First note that from Proposition 2, $D(x)+D(y)$ is a symmetric levelable function of both variables x and y. On the other hand the following formula can be written:

$$\begin{aligned} S(\max(x, y)) - S(\min(x, y)) \\ = S(x) + S(y) - 2\ S(\min(x, y)) \end{aligned}$$

  Since $\mathbb{1}_{\lambda<\min(x,y)} = \mathbb{1}_{\lambda<x} \cdot \mathbb{1}_{\lambda<y}$ one gets from (7):

$$\begin{aligned} & S(\max(x, y)) - S(\min(x, y)) \\ &= \sum_{\lambda=0}^{L-2} (S(\lambda+1) - S(\lambda))(\mathbb{1}_{\lambda<x} + \mathbb{1}_{\lambda<y} - 2\ \mathbb{1}_{\lambda<x} \cdot \mathbb{1}_{\lambda<y}) \\ &= \sum_{\lambda=0}^{L-2} (S(\lambda+1) - S(\lambda))\ |\ \mathbb{1}_{\lambda<x} - \mathbb{1}_{\lambda<y}\ |. \end{aligned}$$

- Necessary - Previous argument leads us to expand $U(x, y)$ on the basis of $\{\mathbb{1}_{\lambda<x},\ \mathbb{1}_{\lambda<y}\}_\lambda$ and $\{|\mathbb{1}_{\lambda<x} - \mathbb{1}_{\lambda<y}|\}_\lambda$ rather than $\{\mathbb{1}_{\lambda<x},\ \mathbb{1}_{\lambda<y}\}_\lambda$ and $\{\mathbb{1}_{\lambda<x} \times \mathbb{1}_{\lambda<y}\}_\lambda$:

$$\begin{aligned} & U(x, y) \\ &= \sum_{\lambda=0}^{L-2} R(\lambda)|\mathbb{1}_{\lambda<x} - \mathbb{1}_{\lambda<y}| + \sum_{\lambda=0}^{L-2} \theta(\lambda)[\mathbb{1}_{\lambda<x} + \mathbb{1}_{\lambda<y}] \\ &= \sum_{\lambda=0}^{L-2} (S(\lambda+1) - S(\lambda))\ |\ \mathbb{1}_{\lambda<x} - \mathbb{1}_{\lambda<y}| \\ &\quad + \sum_{\lambda=0}^{L-2} (D(\lambda+1) - D(\lambda))[\ \mathbb{1}_{\lambda<x} +\ \mathbb{1}_{\lambda<y}], \end{aligned}$$

  where $S(\lambda)$ are $D(\lambda)$ are respective "primitives" of $R(\lambda)$ and $\theta(\lambda)$, e.g.

$$\begin{cases} S(0) = D(0) = 0, \\ S(\lambda) = \sum_{\mu<\lambda} R(\mu),\ \ D(\lambda) = \sum_{\mu<\lambda} \theta(\mu)\ \ \forall\lambda \geq 1. \end{cases}$$

  Then (7) implies, up to a constant:

$$\begin{aligned} & U(x, y) \\ &= S(\max(x, y)) - S(\min(x, y)) + D(x) + D(y) \\ &= S(\max(x, y)) - S(\min(x, y)) + D(\max(x, y)) \\ &\quad + D(\min(x, y)). \end{aligned}$$ □

C. Proof of Proposition 5

The minimum energetic condition A 4. writes from Eq. (9) as:

$$\begin{aligned} (a)\ \forall x \geq y\ \ & U(x, y) = F(x) - G(y) \geq \\ & U(y,\ y) = F(y) - G(y) \Rightarrow F(x) \geq F(y) \\ (b)\ \forall x \leq y\ \ & U(x, y) = F(y) - G(x) \geq \\ & U(y,\ y) = F(y) - G(y) \Rightarrow G(x) \leq G(y) \end{aligned}$$

These inequalities must hold $\forall y \in \mathcal{L}$. Thus F, G and $S = (F + G)/2$ should be *increasing* functions on $\mathcal{L}$. This in turn implies that $R(\lambda) = S(\lambda+1) - S(\lambda)$ should be a *positive* function on $\mathcal{L}^*$. The level summation formula (10) results immediately from the previous proof of Proposition 4. □

D. Proof of Proposition 8

It is sufficient for this purpose to prove that each term

$$R_{st}^{i,\lambda}\left(u_s^\lambda, u_t^\lambda\right) = R_i(\lambda)\left|u_s^\lambda - u_t^\lambda\right|,$$

and

$$H_{st}^\lambda\left(u_s^\lambda, u_t^\lambda\right) = \alpha H\left(u_s^\lambda - u_s^{\lambda+1}\right),$$

should verify the regularity property i.e.,

$$R_{st}^{i,\lambda}(0, 0) + R_{st}^{i,\lambda}(1, 1) \leq R_{st}^{i,\lambda}(0, 1) + R_{st}^{i,\lambda}(1, 0), \tag{23}$$

and

$$H_{st}^\lambda(0, 0) + H_{st}^\lambda(1, 1) \leq H_{st}^\lambda(0, 1) + H_{st}^\lambda(1, 0). \tag{24}$$

This is obviously the case: first inequality (23) holds (R_i is a positive function), then inequality (24) also holds since $\alpha > 0$ and from the definition of H. The total energy $E_{\alpha,v}(\{u^\lambda\})$ in (20) is thus a sum of graph-representable terms and is thus itself graph-representable [16]. □

E. Proof of Proposition 9

Notice that minimizer $\{\hat{u}^\lambda\}$ verifies the monotony property. We thus define the grey level value $\hat{u}_s$ from the usual reconstruction formula $\forall s\ \hat{u_s} = \min\{\lambda, u_s^\lambda = 1\}$ [13]. Is is clear that with these conditions one has

$$E_v(\hat{u}) = E_{\alpha,v}(\{\hat{u}^\lambda\}),$$

since $\forall s\ \alpha\ H(\hat{u}_s^\lambda - \hat{u}_s^{\lambda+1}) = 0$, and since the other components of energies (16) and (20) have identical values. This concludes the proof.

F. Proof of Proposition 10

Let us assume that minimizer $\{\hat{u}^\lambda\}$ of $E_{\alpha,v}(\{\cdot^\lambda\})$ is such that

$$\begin{aligned}&\exists s,\ \lambda \text{ s.t. } H_\alpha\big(\hat{u}_s^\mu - \hat{u}_s^{\mu+1}\big) = 0 \quad \forall \mu < \lambda\\ &\quad\text{and}\quad H_\alpha\big(\hat{u}_s^\lambda - \hat{u}_s^{\lambda+1}\big) > 0 \Leftrightarrow \hat{u}_s^\lambda = 1 \text{ and } \hat{u}_s^{\lambda+1} = 0.\end{aligned}$$

i.e., the monotony condition is satisfied for s up to $\mu \le \lambda$. We want to discard this solution in order to preserve monotony, and thus

$$\begin{aligned}&\Delta\phi_s(\lambda+1)\\ &\quad = E_v^\lambda\big(u_s^{\lambda+1}{=}1 \mid N_s^{\lambda+1}\big) - E_v^\lambda\big(u_s^{\lambda+1}{=}0 \mid N_s^{\lambda+1}\big) < 0.\end{aligned}$$

From (16) this means

$$\begin{aligned}&\Delta\phi_s(\lambda+1)\\ &\quad = \sum_{i=1}^{I} \sum_{\substack{t\sim s\\ \tau(s,t)=i}} R_i(\lambda+1)\,\big(2u_t^{\lambda+1}-1\big) - \delta(\lambda, v_s) - \alpha < 0.\end{aligned}$$

In the conditions of the proposition one has effectively

$$\begin{aligned}\alpha &> \Gamma \sup_{\lambda,i} R_i(\lambda) - \inf_{\lambda\ge 1, v_s} \delta(\lambda, v_s)\\ &> \sum_{i=1}^{I} \sum_{\substack{t\sim s\\ \tau(s,t)=i}} R_i(\lambda+1)\big(1 - 2u_t^{\lambda+1}\big) - \delta(\lambda, v_s)\end{aligned}$$

Thus $\{\hat{u}^\lambda\}$ cannot be a minimizer of $E_{\alpha,v}(\{\cdot^\lambda\})$. This concludes the proof. □

G. Proof of Proposition 11

We only deal with the regularization terms since the decomposition has already been performed for data terms in Section 3. We decompose $g(\cdot,\cdot)$ in cascade over the two variables using (7), so that we have $\forall k, l \in [0, L-1]^2$:

$$\begin{aligned}U(k,l) &= \sum_{\lambda=0}^{L-2} (U(\lambda+1, l) - U(\lambda, l))\, \mathbb{1}_{\lambda<k} + U(0,l)\\ &= \sum_{\mu=0}^{L-2}\sum_{\lambda=0}^{L-2} \{U(\lambda+1, \mu+1) - U(\lambda, \mu+1)\\ &\quad - U(\lambda+1,\mu) + U(\lambda,\mu)\}\ \mathbb{1}_{\lambda<k}\mathbb{1}_{\mu<l}\\ &\quad + \sum_{\lambda=0}^{L-2} (U(\lambda+1, 0) - U(\lambda, 0))\, \mathbb{1}_{\lambda<k}\\ &\quad + \sum_{\mu=0}^{L-2} (U(0, \mu+1) - U(0, \mu))\, \mathbb{1}_{\mu<l}\\ &\quad + U(0,0).\end{aligned} \tag{25}$$

This decomposition holds whatever function $U(\cdot,\cdot)$. We now inject the assumption that $U(x,y) = V(x-y)$ $\forall x, y \in \mathcal{L}$ into Eq. (25), and we get:

$$\begin{aligned}U(k,l) = &\sum_{\mu=0}^{L-2}\sum_{\lambda=0}^{L-2} \{2V(\lambda-\mu) - V(\lambda-\mu-1)\\ &- V(\lambda-\mu+1)\}\ \mathbb{1}_{\lambda<k}\mathbb{1}_{\mu<l}\\ &+ \sum_{\lambda=0}^{L-2} (V(\lambda+1) - V(\lambda))\,[\mathbb{1}_{\lambda<k} + \mathbb{1}_{\lambda<l}] + V(0).\end{aligned} \tag{26}$$

Since we assume that $V(\cdot)$ is also a *convex* function, the following holds:

$$\begin{aligned}&\forall \lambda, \mu\ \ G(\lambda,\mu)\\ &\quad = 2V(\lambda-\mu) - V(\lambda-\mu-1) - V(\lambda-\mu+1) \le 0.\end{aligned} \tag{27}$$

The work is almost done using Eq. (26) for regularization terms. We inject this equality and the one for the data terms (given by Proposition 2) into Eq. (1). Besides, note that $\mathbb{1}_{\lambda<u_s} = (1 - u_s^\lambda)$. So we get the expected result. This concludes the proof. □

H. Proof of Proposition 12

Each term of Eq. (22) is submodular, i.e., graph-representable in the Kolmogorov et al. [16]:

- H : the justification is given in Section 4.
- D_{st}^{λ} and Δ_s^{λ} in Eq. (21) are functions of a single binary variable. Thus they are graph-representable.
- each term $U_{st}^{\lambda,\mu}(u_s^{\lambda}, u_t^{\mu}) = G_{st}(\lambda, \mu)\,(1-u_s^{\lambda})(1-u_t^{\mu})$ involving two binary variables in Eq. (21) is graph-representable because it satisfies the regularity property (see Appendix D Eqs. (23) and (24)):

$$U_{st}^{\lambda,\mu}(0,0) + U_{st}^{\lambda,\mu}(1,1) \leq U_{st}^{\lambda,\mu}(1,0) + U_{st}^{\lambda,\mu}(0,1),$$

since $G_{st}(\cdot,\cdot)$ is non-positive by the *convexity* assumption.

In [16], Kolmogorov *et al.* show that the sum of graph-representable functions is graph-representable, which is the case for the energy $E_\alpha(\{\cdot^{\lambda}\}|v)$. This concludes the proof. □

I. Proof of Proposition 14

Each local conditional energy may be decomposed similarly to Part I as:

$$\begin{aligned} &E_v(u_s \mid N_s) \\ &= \sum_{\lambda=0}^{L-2} (E_v(\lambda+1 \mid N_s) - E_v(\lambda \mid N_s))\left(1 - u_s^{\lambda}\right) \\ &\quad + E_v(0 \mid N_s) \\ &= \sum_{\lambda=0}^{L-2} (\Delta\phi_s(\lambda)\, u_s^{\lambda} + \chi_s(\lambda)) \\ &\text{where } \Delta\phi_s(\lambda) = -(E_v(\lambda+1 \mid N_s) - E_v(\lambda \mid N_s)). \end{aligned} \tag{28}$$

The local posterior Gibbs distribution at a given level writes thus:

$$P\left(u_s^{\lambda} = 1 \mid N_s,\ v_s\right) = \frac{1}{1 + \exp\{\Delta\phi_s(\lambda)\}}.$$

Validating the coupled Gibbs sampling scheme of Part I requires that

$$\forall N_s,\ v_s,\ P\left(u_s^{\lambda} = 1 | N_s, v_s\right) \text{ is non-decreasing with } \lambda,$$

and thus from (28):

$$\begin{aligned} &\forall N_s,\ v_s, -\Delta\phi_s(\lambda) \\ &= E_v(\lambda+1|N_s) - E_v(\lambda|N_s) \text{ is non-decreasing with } \lambda. \end{aligned}$$

Thus $\forall N_s,\ v_s,\ E_v(\cdot|\ N_s)$ should be a convex function on $]0, L-1[$. □

Acknowledgments

The authors would like to thank deeply Vincent Arsigny (INRIA Sophia EPIDAURE Project), Jean-Francois Aujol (CMLA ENS Cachan) and Antonin Chambolle (CMAP Ecole Polytechnique) for enlighting discussions.

References

1. R. Ahuja, T. Magnanti, and J. Orlin, *Network Flows: Theory, Algorithms and Applications*. Prentice Hall, 1993.
2. S. Alliney, "An algorithm for the minimization of mixed l^1 and l^2 norms with application to bayesian estimation," *IEEE Transactions on Signal Processing*, Vol. 42, No. 3, pp. 618–627, 1994.
3. J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistics Society*, Vol. 48, pp. 259–302, 1986.
4. A. Blake and A. Zisserman, *Visual Reconstruction*. MIT Press, 1987.
5. C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Transactions on Transactions on Signal Processing*, Vol. 2, No. 3, pp. 296–310, 1993.
6. S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
7. Y. Boykov and V. Kolmogorov, "An experimental comparison of Min-Cut/Max-Flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 9, 1124–1137, 2004.
8. Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 11, pp. 1222–1239, 2001.
9. D. Coupier, A. Desolneux, and B. Ycart, "Image denoising by statistical area thresholding," *Journal of Mathematical Imaging and Vision*, Vol. 22, No. 2–3, pp. 183–197, 2005.
10. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images," *IEEE Pattern Analysis and Machine Intelligence*, Vol. 6, No. 6, pp. 721–741, 1984.
11. G. Giorgi, A. Guerraggio, and J. Thierfelder, *Mathematics of Optimization: Smooth and Nonsmooth Case*. Elsevier Science, 2004.
12. D. Greig, B. Porteous, and A. Seheult, "Exact maximum a posteriori estimation for binary images," *Journal of the Royal Statistics Society*, Vol. 51, No. 2, pp. 271–279, 1989.
13. F. Guichard and J. Morel, "Mathematical morphology "almost everywhere,"" in *Proceedings of ISMM*, 2002, pp. 293–303.
14. H. Ishikawa, "Exact optimization for Markov random fields with convex priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 10, pp. 1333–1336, 2003.
15. S. Iwata, L. Fleischer, and S. Fujishige, "A combinatorial strongly polynomial algorithm for minimizing submodular functions," *Journal of the ACM*, Vol. 48, No. 4, pp. 761–777, 2001.

16. V. Kolmogorov and R. Zabih, "What energy can be minimized via graph cuts?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 2, pp. 147–159, 2004.
17. M. Nikolova, "Local strong homogeneity of a regularized extimator," *SIAM Journal on Applied Mathematics*, Vol. 61, pp. 633–658, 2000.
18. M. Nikolova, "A variational approach to remove outliers and impulse noise," *Journal of Mathematical Imaging and Vision*, Vol. 20, pp. 99–120, 2004.
19. S. Osher, A. Solé, and L. Vese, "Image decomposition and restoration using total variation minimization and the $\mathbf{H}^{-1}norm$," *J. Mult. Model. and Simul.*, Vol. 1, No. 3, 2003.
20. I. Pollak, A. Willsky, and Y. Huang, "Nonlinear evolution equations as fast and exact solvers of estimation problems," *IEEE Transactions on Signal Processing*, Vol. 53, No. 2, pp. 484–498, 2005.
21. L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D.*, Vol. 60, pp. 259–268, 1992.
22. K. Sauer and C. Bouman, "Bayesian estimation of transmission tomograms using segmentation based optimization," *IEEE Transactions on Nuclear Science*, Vol. 39, No. 4, pp. 1144–1152, 1992.
23. B. Zalesky, "Efficient determination of Gibbs estimators with submodular energy functions," `http://www.citebase.org/cgi-bin/citations?id=oai:arXiv.org:math/0304041`, 2003.

Jérôme Darbon was born in Chenôve, France in 1978. From 1998 to 2001, he studied computer science at Ecole Pour l'Informatique et les Techniques Avancées (EPITA), France. He received the M.Sc. degree in applied mathematics from E.N.S. de Cachan, France, in 2001. In 2005, he received the Ph.D. degree from Ecole Nationale des Télécommunications (ENST), Paris, France. He is currently a postdoc at the Department of Mathematics, University of California, Los Angeles, hosted by Prof. T.F. Chan. His main research interests include fast algorithms for exact energy minimization and mathematical morphology.

Marc Sigelle was born in Paris on 18th March 1954. He received an engeneer diploma from Ecole Polytechnique Paris 1975 and from Ecole Nationale Supérieure des Télécommunications Paris in 1977. In 1993 he received a Ph.D. from Ecole Nationale Supérieure des Télécommunications. He worked first at Centre National d'Etudes des Télécommunications in Physics and Computer algorithms. Since 1989 he has been working at Ecole Nationale Supérieure des Télécommunications in image and more recently in speech processing. His main subjects of interests are restoration and segmentation of signals and images with Markov Random Fields (MRF's), hyperparameter estimation methods and relationships with Statistical Physics. His interests concerned first reconstruction in angiographic medical imaging and processing of remote sensed satellital and synthetic aperture radar images, then speech and character recognition using MRF's and bayesian networks. His most recent interests concern a MRF approach to image restoration with Total Variation and its extensions.