Global Optimization for First Order Markov Random Fields with Submodular Priors

Jérôme Darbon

UCLA Mathematics Department, Los Angeles, USA

Abstract

This paper copes with the global optimization of Markovian energies. Energies are defined on an arbitrary graph and pairwise interactions are considered. The label set is assumed to be linearly ordered and of finite cardinality, while each interaction term (prior) shall be a submodular function. We propose an algorithm that computes a *global* optimizer under these assumptions. The approach consists of mapping the original problem into a combinatorial one that is shown to be globally solvable using a maximum-flow/s-t minimum-cut algorithm. This restatement relies on considering the level sets of the labels (seen as binary variables) instead of the label values themselves. The submodularity assumption of the priors is shown to be a necessary and sufficient condition for the applicability of the proposed approach. Eventually, some numerical results are presented.

Key words: Maximum-flow/s,t minimum-cut, global optimization, submodular functions

1 Introduction

Many image processing and computer vision problems are formulated as a discrete optimization problem. Among many available discrete frameworks, Maximum *a posteriori* (MAP) estimators for Markov Random Fields (MRFs) models have been increasingly used [8,18,33,47,63] since the seminal work of Geman and Geman [32]. Markovian energies of interest are generally a weighted combination of a fidelity term and a prior. The former measures the distance of the reconstructed solution to the observed data while the latter

Email address: jerome@math.ucla.edu (Jérôme Darbon).
URL: http://jerome.berbiqui.org (Jérôme Darbon).

embeds some *a priori* knowledge on the result. Unfortunately global optimization of these energies is generally difficult. For some particular cases, computations are tractable using dynamic programming [4] and shortest path algorithms [19]. However, energies of interest generally remain difficult to globally optimize. Besides, these optimization problems can even be NP-hard as shown in [12,31,39]. Thus stochastic sampling and simulated annealing techniques [32] are sometimes used to compute the MAP although it might be slow in practice [26,39,41,45,65]. Another approach consists of searching for approximate solutions as proposed in [6,12] for instance. This paper aims at finding the MAP of Markovian energies that involves pairwise interactions and any separable data fidelity. An algorithm that computes a global minimizer of a subclass of these energies in more generality that it was previously possible is presented. The proposed approach shall be seen as a complementary computational point of view to the theoretical work on the properties of global minimizers of Nikolova [52,53], and Durand and Nikolova [28,29].

Let us define the problem of minimizing a first-order MRF. Images are defined on a discrete lattice $\mathcal{V}$ whose cardinality is $|\mathcal{V}|$. Let us denote by u_p the value of the image u at the site $p \in \mathcal{V}$. In this paper, it is assumed that u_p takes value in the finite, linearly ordered, discrete set $\mathcal{L} \subset \mathbb{R}$ of cardinality $|\mathcal{L}| = L$, i.e., $\mathcal{L} = \{l_0, \dots l_{L-1}\}$ with $l_i < l_{i+1}\, \forall i \in [\![0, L-2]\!]$. The lattice $\mathcal{V}$ is endowed with a neighbourhood system. The adjacency relationship of two adjacent sites p and q is referred to as $p \sim q$. In this paper, only pairwise interactions are considered, and such a clique is referred to as (p, q), with $p \sim q$. The set of all cliques is denoted by $\mathcal{E}$. The goal of this paper is to compute a global minimizer of the following first order Markovian energy:

$$E(u|v) = \sum_{p \in \mathcal{V}} f_p(u_p|v_p) + \sum_{(p,q) \in \mathcal{E}} g_{pq}(u_p, u_q) \ , \tag{1}$$

where v is the observed image. The family of functions $\{f_p\}$ and $\{g_{pq}\}$ are respectively referred to as the fidelity terms and the priors. It is also assumed that the functions $\{f_p\}$ and $\{g_{pq}\}$ take values in $\mathbb{R}$ and are respectively defined on the discrete sets $\mathcal{L}$ and $\mathcal{L}^2$. Such functions can thus be considered as discrete functions [51]. In this paper, the priors $\{g_{pq}\}$ shall be submodular functions. For any positive integer k, a function $h : \mathcal{L}^k \mapsto \mathbb{R}$ is said submodular if and only if it satisfies the following inequality [51]:

$$\forall (x, y) \in \mathcal{L}^{2k} \quad h(x \vee y) + h(x \wedge y) \leq h(x) + h(y) \ , \tag{2}$$

where $(x \vee y)$ and $(x \wedge y)$ respectively corresponds to the component-wise maximum and minimum between x and y, i.e., we have for any $p \in \mathcal{V}$ $(x \vee y)_p = \max\{x_p, y_p\}$ and $(x \wedge y)_p = \min\{x_p, y_p\}$. A function is said supermodular if and only if its negation is submodular. Submodularity and supermodularity

can be seen as a general property of discrete functions that are analogous to convexity of functions defined on continuous domain [49,51].

The main contributions of this paper are the following: first, an algorithm that computes a global minimizer for first order MRFs with submodular priors is proposed. No assumption is required on data fidelity terms. The approach relies on mapping the original problem into a binary optimization problem using the level sets of the labels (see Section 3 for a definition). It is shown that the global optimization of this binary energy can be performed by computing a s-t minimum-cut (or by duality a maximum-flow [1,43]) on a graph associated to this binary energy following the approach described in [5,12,35,45,55]. Second, it is shown that the submodularity of the priors is a sufficient and necessary condition for the applicability of the proposed approach. To our knowledge, these results are new and considerably extend previous approaches for global energy optimization such as those of Ahuja et al. [3], Darbon and Sigelle [26], Ishikawa [41] and Zalesky [65]. This work can also be seen as a natural extension of the work of Picard and Ratliff [55], Kolmogorov and Zabih [45] for global binary energies minimization via "graph-cuts" [12]. We also note the independent work of [57] that seems to describe similar results.

Although computing a maximum-flow can be performed in polynomial time, our algorithm has only a pseudo-polynomial time complexity [1] since the number of nodes of the graph grows linearly with respect to the number of label $|\mathcal{L}|$. To be polynomial, this number shall be $O(\log |\mathcal{L}|)$ [1,31]. Although these non-polynomial time and space complexities might not be attractive for applications when the number of labels is large, it is worth to note, as in [35] and [63, p. 136], that this exact optimization scheme allows to study the quality of a model and the influence of its parameters *independently* of the algorithm to compute the MAP estimates. Besides, a ground truth allows to measure the practical performance of an approximation algorithm [7,24,27,48,60,61]. These considerations motivates the development of global minimization algorithms.

The remainder of this paper is organized as follows. Related works are presented in section 2. The level-set based restatement of the data fidelities and priors in terms of binary variables is described in section 3. This restatement is the core of our mapping of the original minimization problem into a binary one. The proposed algorithm for minimizing first order Markovian energies with submodular priors is described in section 4. Some experiments for image processing purposes are presented in section 5. Finally, some conclusions are drawn in section 6.

2 Related Works

Due to the difficulty, in general, of minimizing the discrete energy defined by Equation (1), many suboptimal schemes have been proposed [6,8,32]. In this paper, we focus on algorithms that compute a global minimizer. Most of these global optimization schemes rely on combinatorial optimization algorithms [59]. In particular among them, s-t minimum cuts or by duality maximum flows [1,43] are now popular techniques to exactly optimize Markovian energies [12,26,41,45]. This technique consists of defining a graph such that its s-t minimum-cut/maximum flow [1,43] yields an optimal labeling of the discrete energy. Recall that computing a maximum flow can be performed in polynomial time [1,43]. Thus the main challenge consists of defining a graph construction associated with the discrete energy. In the following, we review the schemes available in the literature. We first begin with energies involving binary labels before considering the case of linearly ordered labels.

2.1 Binary energies

In [55], Picard and Raliff characterize a class of boolean energies, i.e., $\mathcal{L} = \{0, 1\}$, that can be minimized exactly in polynomial time via computing a s-t minimum-cut on a related graph. More precisely, they consider binary quadratic polynomials of the following form:

$$E(u) = \sum_{p \in \mathcal{V}} a_p u_p + \sum_{(p,q) \in \mathcal{E}} w_{pq} u_p u_q \ , \tag{3}$$

where a_p and w_{pq} are real values and with $w_{pq} \leq 0$. The latter assumption $w_{pq} \leq 0$ ensures that all capacities in the related graph are nonnegative and thus, allows the minimum-cut to be efficiently computed [1,43]. This seminal work has been firstly used by the statistical physics community to study ground states of binary Markov Random Fields. In [5], Barahona applies this approach to compute global minimizer of Markovian energies with ferromagmetic Ising priors, i.e., priors that takes the following form $g_{pq}(x, y) = |x - y|$ where $(x, y) \in \{0, 1\}^2$. Indeed, by noticing that the following equality holds for binary variables $|x - y| = x + y - 2xy$, then it is readily seen that it fits the Picard and Ratliff framework. In [54], Ogielski uses this graph construction to study the phase transition for the ferromagnetic Ising MRF model in 3D. In [38], Hartmann and Usadel propose an approach that also relies on the theoretical work of Picard and Ratliff to compute *all* optimal labelings of ferromagnetic Ising MRF. The authors also note that the approach works for all energies that they can be mapped to ferromagnetic Ising energies with local changes of variables, also called local Gauge transformation in physics.

From an image processing point of view, Greig et al. were the first to use in [35] the work of Picard and Ratliff for studying binary image restoration with ferromagnetic Ising-based models. This prior mainly corresponds to penalize objects with large perimeters [63].

In [10], the authors use a similar technique to perform image segmentation with hard constraints. We refer the reader to the review [9] for such approaches for image segmentation and computer vision problems. In [45], Kolmogorov and Zabih propose a graph construction for classes of binary Markovian energies where priors correspond to pairwise or triplewise interactions. Besides, it is shown that interactions should be binary submodular.

The next proposition establishes the equivalence between the work of Picard and Ratlif [55] and the one of Kolmogorov and Zabih [45] for pairwise interactions.

Proposition 1 [PR-KZ condition] *Assume E is a binary energy with pairwise interactions, i.e.,*

$$E(u) = \sum_p f_p(u_p) + \sum_{(p,q)} g_{pq}(u_p, u_q) \ ,$$

with $\forall p \in S \, u_p \in \{0,1\}$. The following two assertions are equivalent:

- *(Picard and Ratliff [55]), each pairwise interaction writes as $g_{pq}(x,y) = w_{pq}xy$ with $w_{pq} \leq 0$,*
- *(Kolmogorov and Zabih [45]), each pairwise interaction is submodular.*

Besides, if the binary energy E satisfies these conditions, then it can be optimized in polynomial time via computing a maximum flow on a graph associated with E as shown in [12,45,55].

The proof of the equivalence is given in Appendix A. These conditions for maximum flow-based exact optimization will be referred to as the PR-KZ condition in the remainder of this paper.

Energies involving triplewise interactions can also be exactly minimized via graph-cuts as long as they are submodular, as shown by Kolmogorov and Zabih in [45]. Some examples of submodular functions with higher order interactions are described by Zalesky in [65]. In [30], some sufficient conditions for higher order interaction energies (i.e., strictly more than triplewise interactions) are given so that they can be globally minimized via maximum flows. We also note that message passing approaches, such as those of [46,62], exactly optimize binary MRF with submodular priors .

Note that all these approaches yield polynomial algorithms since all graph constructions described in these papers are linear with respect to the number of

variables and that there exists polynomial algorithms to solve the maximum flow problem [1,43]. However, the use of this binary optimization approach for solving image processing and computer vision problem is made possible mainly because of the very efficient maximum-flow algorithm of Boykov and Kolmogorov [11]. Although the latter does not have a theoretical polynomial time complexity, its empirical computational time behaves quasi-linearly with respect to the number of nodes for graphs that have small connectivity (which is the case for most of image processing and computer vision problems). Eventually, let us note that minimizing a non-submodular function is NP-hard in general as shown in [45].

2.2 Energies with Linear ordered labels

Extensions of these approaches for exact optimization of MRFs involving more than two labels have been tackled by some authors [3,26,41,65]. All approaches assume that labels can be linearly ordered and there are no assumptions on fidelity terms. In [41], Ishikawa proposes a graph construction for MRFs where the priors are convex functions of the difference of labels. In [65], Zalesky study the class of MRFs whose energies can be rewritten as submodular Boolean polynomials (i.e., without any restriction on the degree of the interactions). The author also proposes an algorithm based on submodular function minimization algorithms [42,58] to optimize these Boolean energies. In [26], Darbon and Sigelle consider the same class of Markovian energies dedicated to image processing purposes. This assumption allows the authors to propose a graph construction scheme for which a minimum s,t-cut yields a global minimizer.

Markovian energies with priors that are convex functions of the difference of the labels, i.e., $g_{pq}(x, y) = h_{pq}(x - y)$ where the family $\{h_{pq}\}$ are unary convex functions, can be minimized via minimum cost flow as shown in [3,26,41]. To our knowledge, these approaches cope with the most general class of first order MRFs. Recall that the approach proposed in this paper includes this class and also cope with more general cases.

Efficient approximation algorithms are proposed by Boykov et al. in [12] for MRFs with pairwise priors that are semi-metric of the difference of the labels. Moreover, the authors are able to estimate the quality of the local minimizer computed via this algorithm. Such an approach is iterative and relies on the ability to globally optimize a binary Markovian energy with submodular priors.

Finally, we note that efficient exact minimization schemes have been developed for MRFs where data fidelity terms are convex functions and where the priors are convex functions of the difference of the labels, i.e., $g_{pq}(\cdot - \cdot)$. Such

approaches are described in [2,7,22,44,51]. Among these models, the particular case of the Total Variation [56] minimization has received a lot of attention. Very efficient algorithms are available and presented in [14,15,25,34,39,64]. We do not detail these approaches here since we focus, in this paper, on exact optimization of Markovian energies that are generally non-convex.

The proposed approach presented in this paper copes with all Markovian optimization problems presented in this section. We need to introduce some notations and useful results before describing our approach.

3 Development through Lower Level Sets

This section presents the rewriting of all unary data fidelity term $\{f_p\}$ and all pairwise prior terms $\{g_{pq}\}$ appearing in the energy E defined by equation (1), as a linear combination of binary energies that involves the level sets of the variables. More precisely, a label is rewritten as a sum of binary variables. Note that such a representation is indeed widely used in the community of integer programming [20,21] and has been used successfully for many combinatorial problems [3,21,40] as well as some image processing problems [14,16,17,25,26,64,65]. Our approach for optimizing exactly Markovian energies with submodular priors relies on these developments. The notion of level sets is firstly defined. Then the developments on level sets for functions of one and two variables are given.

Let us introduce the characteristic function of a lower level set $[x]_\lambda$ of a variable $x \in \mathcal{L}$ at a level $\lambda \in \mathcal{L}$ as follows:

$$[x]_\lambda = \begin{cases} 0 \quad \text{if} \quad x \leq \lambda, \\ 1 \quad \text{if} \quad x > \lambda \ . \end{cases}$$

Note that the upper level set $[x]^\lambda$, of a variable x could be used in the sequel instead of the lower ones since we have: $[x]^\lambda = 1 - [x]_\lambda$. Also, note that inequality and strict inequality can be interchanged in the definition of a level set without altering the nature of the results presented in this paper. It is straightforward to see that lower level sets of a variable x satisfies a monotone property:

$$\forall \lambda \geq \mu \qquad [x]_\lambda \geq [x]_\mu \ , \tag{4}$$

Also, one can reconstruct the the original gray-level value from its lower level

sets using the following reconstruction formula as shown in [36,50]:

$$x = \max\{\lambda \in \mathcal{L}, [x]^\lambda = 0\} \ . \tag{5}$$

Conversely, it is shown in [36,50] that any family of binary variables $\{b_i\}_{i=1\ldots n}$ that satisfies the monotone property, given by equation (4), defines a label. Note that the same result holds for level sets defined for continuous labels [36,37] (up to a null measure set). In other words, it is equivalent to have the knowledge of a label x or of its level sets $\{[x]_\lambda\}$. The idea is to look for on optimal solution, not in terms of the labels but using a series of binary variables that represent the level sets of the labels. We now rewrite the fidelity terms and the priors through the level sets of the labels.

The next proposition gives a development for data fidelity term as a summation on the level sets of its variable as already proposed in [25]. It is based on a "discrete" integration of the variations of f_p over its lower or upper level sets.

Proposition 2 *Any data fidelity term $f_p : \mathcal{L} \mapsto \mathbb{R}$ rewrites on its lower or upper sets as follows:*

$$f_p(x) \ = \ \sum_{i=0}^{L-2} D_p(i)[x]_{l_i} + f_p(l_0) \ , \tag{6}$$

where $\forall i \in [\![0, L-2]\!] \ D_p(i) = f_p(l_{i+1}) - f_p(l_i) \ .$

The proof is straightforward and is a simple adaptation of the one given in [25]. Note that the summation goes only up to $(L-2)$ because for any label x that lives in $\mathcal{L}$ we have $[x]_{l_{L-1}} = 0$.

Next, we need to cope with functions of two variables. A natural way to achieve this consists in applying the previous developments firstly on the first variable and then on the second one. It yields the following proposition that relates to the submodularity of a function to the PR-KZ conditions for global optimization.

Proposition 3 *Any prior term $g_{pq} : \mathcal{L}^2 \mapsto \mathbb{R}$ rewrites on its level sets as follows:*

$$g_{pq}(x, y) = \sum_{i=0}^{L-2} \sum_{j=0}^{L-2} R_{pq}(i, j)[x]_{l_i}[y]_{l_j} \tag{7}$$

$$+\sum_{i=0}^{L-2}\left(D^1_{pq}(i)[x]_{l_i}+D^2_{pq}(i)[y]_{l_i}\right)+C\ \ ,$$

where

$$\forall i\in[\![0,L-2]\!]\ D^1_{pq}(l_i)=g(l_{i+1},l_0)-g(l_i,l_0)\ \ ,$$

and

$$\forall i\in[\![0,L-2]\!]\ D^2_{pq}(l_i)=g(l_0,l_{i+1})-g(l_0,l_i)\ \ ,$$

and $C=g_{pq}(l_0,l_0)$ *and more importantly where*

$$\forall(i,j)\in[\![0,L-2]\!]^2\ \ R_{pq}(i,j)=g(l_{i+1},l_{j+1})-g(l_{i+1},l_j)-g(l_i,l_{j+1})+g(l_i,l_j) \quad (8)$$

Again more details on this calculus can be found in [26]. So far, no assumptions have been set neither on data fidelity terms, nor on priors. In other words, claims stated in proposition 2 and proposition 3 hold for any function of one and two variables, respectively. The next section specializes these results for globally optimizing first order Markovian energies with submodular priors.

4 Markovian Energies with Submodular Priors

Recall that it is assumed that all priors $\{g_{pq}\}$ are submodular functions. In this section we show that, under this assumption, a first order Markovian energy can be exactly optimized via computing a maximum flow on an associated graph [12,45,55]. Recall that for any positive integer k a function $h:\mathcal{L}^k\mapsto\mathbb{R}$, is said submodular if and only if it satisfies the following inequality:

$$\forall(x,y)\in\mathcal{L}^{2k}\ \ h(x\vee y)+h(x\wedge y)\leq h(x)+h(y)\ \ . \quad (9)$$

First, let us give some examples of submodular priors:

- $g_{pq}(x,y)=g(x-y)$ with g a unary convex function. This kind of prior is widely used models in image processing for restoration purposes.
- $g_{pq}(x,y)=-xy$, i.e., minus of a scalar product. Such a prior can be used for vector diffusion for instance.
- $g_{pq}(x,y)=g(x+y)$ with g a unary concave function.

Our goal is to map the original Markovian energy given by equation (1), to a binary one. For this purpose, all data fidelity and prior terms are rewritten using the level sets-based expansions given by Proposition 2 and Proposition 3, respectively. Thus the energy $E(\cdot|v)$ is to be rewritten such that it only involves the level sets of the variable to optimize, i.e., $\{[\cdot]_{l_i}\}_{i=0\ldots L-1}$. We have:

$$E(u|v) = \sum_{(p,q)\in\mathcal{E}} \left\{ \sum_{i=0}^{L-2}\sum_{j=0}^{L-2} R_{pq}(i,j)[u_p]_{l_i}[u_q]_{l_j} + \sum_{i=0}^{L-2} D^1_{pq}(i)[u_p]_{l_i} + D^2_{pq}(i)[u_q]_{l_i} \right\} + \sum_{p\in\mathcal{V}}\sum_{i=0}^{L-2} D_p(i)[u_p]_{l_i} + K \ ,$$

where the constant K comes from the constant terms in Propositions 2 and 3. We can now define a new energy $\tilde{E}(\cdot|v)$, whose variables are $|L|$ binary images $\{b^i\}_{i=0\dots L-1}$, as the following:

$$\tilde{E}(\{b^i\}_{i=0\dots L-1}|v) = \sum_{(p,q)\in\mathcal{E}} \left\{ \sum_{i=0}^{L-2}\sum_{j=0}^{L-2} R_{pq}(i,j)b^i_p b^j_q + \sum_{i=0}^{L-2} D^1_{pq}(i)b^i_p + D^2_{pq}(i)b^i_q \right\} + \sum_{p\in\mathcal{V}}\sum_{i=0}^{L-2} D_p(i)b^i_p + K \ .$$

Intuitively, the two energies $E(\cdot|v)$ and $\bar{E}(\cdot|v)$ are equal provided the binary variables involves in $\bar{E}$ defines an image. More formally, if for all sites $p \in \mathcal{V}$, the families of binary images $\{b^i\}_{i=0\dots L-1}$ satisfy the monotone property given by equation (4), then this family defines an image using the reconstruction given by equation (5). Let us emphasize that if any of the families $\{[b^i_p]_\lambda\}_{\lambda=0\dots l_{L-1}}$ violates the monotone property, then a gray level image cannot be defined and thus one cannot easily relate $\bar{E}(\{[b^i_p]_\lambda\}_{\lambda=0\dots l_{L-1}}|v)$ to some $E(\cdot|v)$. Besides, since for any image u we have $E(u|v) = \tilde{E}(\{[u]_\lambda\}_{\lambda=0\dots l_{L-1}}|v)$, it means that if we are able to minimize the energy $\bar{E}(\{[\cdot]_\lambda\}_{\lambda=l_0\dots l_{L-1}}|v)$ while preserving the monotone property, then we get a global minimizer of $E(\cdot|v)$. In order to force the monotone property to hold during the optimization process we define the following new energy:

$$\tilde{E}_\alpha(\{b_i\}_{i=0\dots L-1}|v) = \tilde{E}(\{b_i\}_{i=0\dots L-1}|v) + \sum_{p\in\mathcal{V}} \alpha \sum_{i=0}^{L-2} H(b^{i+1}_p - b^i_p) \ , \qquad (10)$$

where $H : \mathbb{R} \mapsto \mathbb{R}$ is the Heaviside function defined as $H(x) = 0$ if $x \leq 0$ and 1 else. The right hand side of equation 10 corresponds to a penalization term that increases the energy by α every time two consecutive binary variables violate the monotone property. Moreover, it is shown in [26] that if α is set to a sufficiently large finite value, then we are assured that any global minimizer of $\bar{E}_\alpha(\{\cdot\}_{i=0\dots L-1}|v)$ never violates the monotone property give by Eq. (4). The latter proof assumes that the set of labels $\mathcal{L}$ is discrete.

So far we have reformulated the original energy in terms of binary variables. Our goal is now to show that than the energy $E_\alpha(\{[\cdot]_\lambda\}|v)$ can be optimized by maximum flow. We first show in the next proposition that the submodularity of the priors yields binary terms that satisfy the PR-KZ condition.

Proposition 4 *Assume $g : \mathcal{L}^2 \mapsto \mathbb{R}$. The following two assertions are equivalent:*

(1) g is submodular,
(2) g writes as

$$g(x,y) = \sum_{i=0}^{L-2}\sum_{j=0}^{L-2} R(i,j)[x]_{l_i}[y]_{l_j} + \sum_{i=0}^{L-2}\Big(D^+(i)[x]_{l_i} + D^-(i)[y]_{l_j}\Big) + C \ , \tag{11}$$

where $\forall (i,j) \in [\![0, L-2]\!]^2 \ \ R(i,j) \leq 0$, D^+ and D^- are two functions and C is a constant.

Proof. Case 1) $\Rightarrow$ 2) We apply Proposition 3 to g and we get the form given in 2). It is straightforward to see that any unary function is submodular. The submodularity condition given by Eq. (13) applied for the remaining terms $R(i,j)[x]_{l_i}[y]_{l_j}$, reduces to show that $\forall (i,j) \in [\![0, L-2]\!]^2 \ \ R(i,j) \leq 0$.

Recall that Eq. (8) of Proposition 3 also states that

$$R(i,j) = g(l_{i+1}, l_{j+1}) - g(l_{i+1}, l_j) - g(l_i, l_{j+1}) + g(l_i, l_j) \ .$$

Now let us introduce the pairs $a = (l_i, l_{j+1})$ and $b = (l_{i+1}, l_j)$. Then it is readily seen that $R(i,j)$ rewrites as follows:

$$R(i,j) = g(a \wedge b) - g(a) - g(b) + g(a \vee b) \ .$$

The latter is non-positive due to the submodularity of g. This concludes the proof for the first case.

Case 2) $\Rightarrow$ 1): Let $a \in \mathcal{L}^2$ and $b \in \mathcal{L}^2$. Note that the only interesting case happens when $a \notin \{(a \vee b), (a \wedge b)\}$ (otherwise the submodularity property is obviously satisfied).

Let us denote by $(x_m, y_m) = (a \wedge b)$ and $(x_M, y_M) = (a \vee b)$. We need to show that $g((x_m, y_m)) + g((x_M, y_M)) - g((x_m, y_M)) - g((x_M, y_m)) \leq 0$.

To prove this inequality we write each term in the level-set development form given by Eq. (11). One sees that the constant C and the terms involving the

single summation $\left(\sum_{i=0}^{L-2} \cdot\right)$ cancel each other. Thus, only the double summation terms remain, i.e., we need to show:

$$\sum_{i=0}^{L-2}\sum_{j=0}^{L-2} R(i,j)\left([x_m]_{l_i}[y_m]_{l_j} + [x_M]_{l_i}[y_M]_{l_j} - [x_M]_{l_i}[y_m]_{l_j} - [x_m]_{l_i}[y_M]_{l_j}\right) \leq 0 \ ,$$

which is equivalent to

$$\sum_{i=0}^{L-2}\sum_{j=0}^{L-2} R(i,j)\left([x_M]_{l_i} - ([x_m]_{l_i})\left([y_M]_{l_j} - ([y_m]_{l_j}\right) \leq 0 \ . \tag{12}$$

Since $x_M \geq x_m$ and $y_M \geq y_m$ get that

$$\forall i \in \mathcal{L} \ \ ([x_M]_{l_i} \geq ([x_m]_{l_i}) \wedge ([y_M]_{l_i} \geq ([y_m]_{l_i}) \, ,$$

and thus every term in the double summation in (12) are non-positive since $R(i,j) \leq 0$. This concludes the proof. □

The next proposition claims than the energy $E_\alpha(\{[\cdot]_\lambda\}|v)$ can be optimized thanks to the maximum flow algorithm and that it is equivalent to the submodularity of the prior terms $\{g_{pq}\}$.

Proposition 5 *The following two assertions are equivalent:*

(1) all prior terms of energy $E(\cdot|v)$ given (1) are submodular functions,
(2) the associated energy $E_\alpha(\{[\cdot]_\lambda\}|v)$, given by equation (10), is binary submodular, i.e., all binary pairwise interaction terms satisfy the PR-KZ condition.

Proof. Case 1) ⇒ 2) Following Proposition 1, it is enough to show for $\bar{E}_\alpha$, that each single pairwise interaction terms of *binary* variables is submodular. Specializing the definition of submodularity, given by Equation (2), for a binary function f of two variables, i.e., $f : \{0,1\}^2 \to \mathbb{R}$, we get that:

$$f(0,0) + f(1,1) \leq f(0,1) + f(1,0) \ . \tag{13}$$

For the case we are considering, we shall check the submodularity of the terms $H([u_p]_{l_{i+1}} - [u_p]_{l_i})$ and $R_{pq}(i,j)b_p^i b_q^j$. It is easily seen that the terms $H(b_p^{i+1} - b_p^i)$ satisfy the submodular property; see also [26] for further details. Thus it remains to show the submodularity of the terms $R_{pq}(i,j)b_p^i b_q^j$. Using the inequality (13) we need to show that $\forall (i,j) \in [\![0, L-2]\!]^2 \ \ R(i,j) \leq 0$. This property is assured by the submodularity assumption of the priors, as stated in Proposition 4.

Case 2) ⇒ 1) This is straightforward by considering the converse part of Proposition (4) which states that states that any series of non-positive $R(\cdot,\cdot)$

terms defines a submodular function up to some unary terms and a constant. □

So, provided the priors are submodular, a global minimizer of $E(\cdot|v)$ can be computer by minimizing its associated energy $\bar{E}_\alpha(\{[\cdot]_\lambda\}|v)$ with the maximum-flow approach approach [12,35,45,55] (recall that we set α to a sufficiently finite large value). The submodularity assumption is necessary and sufficient. To our knowledge, this result highly generalizes previous ones [3,26,41,65]. Indeed, our approach both includes the class of "levelable" priors described in [26] and [65] (that mainly corresponds to non-convex priors) and the class of priors defined as a convex function of the difference of the labels [3,26,41]. The next proposition relates convexity and submodularity of the priors when they are based on the difference of labels.

Proposition 6 *Assume $g : \mathcal{L}^2 \rightarrow \mathbb{R}$ is submodular and takes the following form $g(x, y) = \tilde{g}(x - y)$ then $\tilde{g}$ is a unary convex function.*

Proof. First we apply Proposition 3. Now, due to the form of $\tilde{g}$ we have that $R(i, j) = 2\tilde{g}(i - j) - \tilde{g}(i - j + 1) - \tilde{g}(i - j - 1)$. We also have $R(i, j) \leq 0$ by the submodularity of g. By letting $k = i - j$ we get that $2\tilde{g}(i - j) \leq \tilde{g}(i - j + 1) + \tilde{g}(i - j - 1)$ which is exactly the discrete second variation convexity criteria for a unary function [51] applied for $\tilde{g}$. □

In other words, the proposed approach reduces to the convex cases presented [3,26,41] when the priors are unary functions of the difference of the labels. Note that the latter priors are widely used in image analysis because it corresponds to the regularization of the gradient of an image. The most well-known example of such a prior is most probably the Total Variation [56].

Eventually, let us note that our approach consists in computing a s-t minimum-cut on graph composed of $|\mathcal{V}|\,|\mathcal{L}|$ nodes (one node per pixel and per gray level value), and $O(|\mathcal{E}||\mathcal{L}|^2)$ edges. From a practical point of view, a large value of $|\mathcal{L}|$ may correspond to a huge amount of memory that is potentially not available. This behavior is not favorable for practical purposes but it allows to compare an approximate solution (computed for instance by the approaches described in [12,61,48]) with a global one. Eventually, recall that a global optimization process, as the one proposed in this paper, allows to intrinsically study the quality of a model [63, p. 136].

5 Experiments

In this Section we illustrate our approach for different MRF models. Our first experiments cope with image restoration where MRFs have submodular priors.

Then we present a Markovian model for image dithering purposes.

5.1 Image Restoration

We present some results for image restoration when images are highly corrupted by impulsive noise with parameter P, which corresponds to the following data fidelity term:

$$f_p(u_p|v_p) = \begin{cases} -\log\left((1-P)+\frac{P}{L}\right) & \text{if } u_p = v_p \\ -\log\frac{P}{L} & \text{else.} \end{cases},$$

In other words, a pixel keeps its original value with a probability $(1-P)$ or it takes a new one uniformly in $\mathcal{L}$. Note that this kind of noise is extremely difficult to remove because the original information a pixel carries is totally lost as soon as it is corrupted. Note that this noise behavior is much more destructive than additive noise for instance. Figures 1-(a) and 2-(a) respectively depicts the image *plane* (of size 250×366) and *squirrel* (of size 209×288) while some highly corrupted versions are presented in Figure 1-(b), and in Figure 2-(b) and Figure 3-(a). Note that it is almost impossible to guess the content of the noisy image.

We consider a Discrete Total Variation (DTV) prior [15] as the prior of the MRF. More formally, the considered neighborhood system on the grid is the 4-nearest neighbors and we set $g_{pq}(u_p, u_q) = \beta|u_p - u_q|$. The non-negative parameter β corresponds to the power of the filtering, i.e., the higher β the more the image is regularized. For these experiments, we set $\mathcal{L} = [\![0, 255]\!]$. Besides, β is adjusted so that the best visual result is achieved. Figure 1-(c) depicts the result for the *plane* image. Note that although 70 percent of the content of the original image is lost, one can recover most of the content of the image. It takes 115s on an Intel Q9650 3GHz processor to perform this minimization.

Restoration result for the corrupted *squirrel* image with $P = 0.7$ is depicted in Figure 2-(c). We also depict in Figure 2-(d) the image obtained by median filtering. The latter is able to recover globally the structure of the image but some artifacts remain. Figure 2-(e) presents the result obtained by minimizing the energy using the $\alpha-$expansion algorithm described in [12]. It produces a solution whose energy is within a known factor of the optimal one. It is clear seen that this procedure is able to recover part of the optimal solution while it fails for other areas by producing constant zones. We have checked that 49% of the pixels of the result obtained by the $\alpha-$expansion algorithm are identical to the ones of the global solution we have computed, and 69% pixels differ with an absolute difference that is than 10. Computing a global minimizer for

this image takes 81s on an Intel Q9650 3GHz.

The restoration for the corrupted *squirrel* image with $P = 0.9$ is depicted in Figure 3-(b). Again, note that many pixels are recovered even in the hardest case where only 10 percent of the information has survived and without knowing where the non-corrupted pixels are. Figure 3-(c) and (d) respectively presents the result for the median filter and the TV minimization using the $\alpha-$expansion algorithm. Comparing the $\alpha-$expansion based result with the global minimizer, we have found that there are 21% of the pixels that are identical, while 40% differ by a magnitude of at most 10. For this experience, the global optimization process takes 152s on an Intel Q9650 3GHz processor.

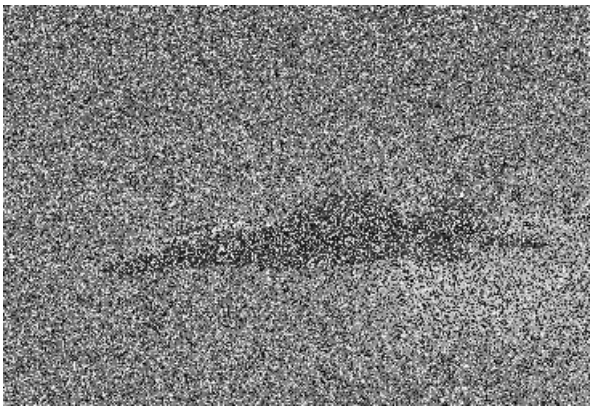
Let us recall that the graphs we need to built for these experiments contain $|\mathcal{V}|\,|\mathcal{L}|$ nodes and $|\mathcal{E}||\mathcal{L}|$ edges (for the special case of TV). In practice, this behavior prevents us from applying the approach on very large images and in particular 3D volumes.

5.2 Halftoning

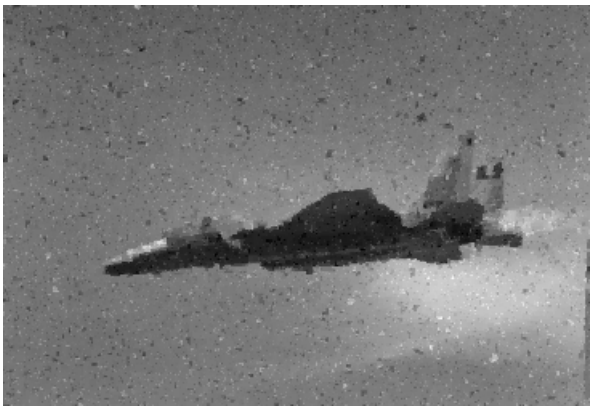
We now present a Markovian energy for performing non-binary halftoning. The goal is to approximate an image with a very little number of gray level values. The use of MRFs for binary halftoning has originally been proposed by Carnaveli et al. in [13]. Their approach relies on the use of a anti-ferromagnetic Ising model. Note that the latter favors signal that are oscillating. Such a behavior is of interest for halftoning purposes since such oscillations are blended into smooth tone by human eyes. This optical effect is thus used to approximate continuous smooth tone.

We propose to extend their approach to the gray level case by considering the *maximization* of the total variation (which also favours oscillations). More precisely, the observed image takes values in the discrete set $[\![0, 255]\!]$ while each pixel of its approximation lives in $\mathcal{L} = \{0, 31, 62, 93, 124, 155, 186, 217, 248\}$. We use the following data fidelity term: $f_p(u_p, v_p) = (31u_p - v_p)^2$. The discrete TV we use here is the same as the one for the restoration experiments. Since we wish to maximize TV, we set $g_{pq} = -\beta|u_p - u_q|$, with β non-negative. Note that this prior is *not submodular* but it can be mapped to a submodular one following the local change of variables strategies of [38]. Since, the nearest neighbors connectivity is considered, the graph is bipartite (equivalent to the 2-colorability), i.e., the sites $\mathcal{V}$ can be decomposed into a partition $\mathcal{V} = \mathcal{V}^1 \cup \mathcal{V}^2$ such that for any couple of interacting variables (u_p, u_q) with $s \sim t$, we have either $(p, q) \in \mathcal{V}^1 \times \mathcal{V}^2$ or $(p, q) \in \mathcal{V}^2 \times \mathcal{V}^1$. One can thus perform the following one-to-one mapping for the sites that belongs to S^1: $u_p \leftarrow (248 - u_p)$. It is straightforward to see that this maps the original prior into a sum of unary

(a)



(b)



(c)

Fig. 1. The *plane* original image is depicted in (a) and its noisy version corrupted by impulsive noise with $P = 0.7$ in (b). The restoration with a Total Variation prior is depicted in (c).

concave functions of the sum of two labels, and is thus submodular. Thus one can globally optimize this new energy by the maximum-flow approach and then get a global minimizer of the original one by inverting the mapping.

Figure 4 depicts both original *lena* image (of size 512^2) and the minimization

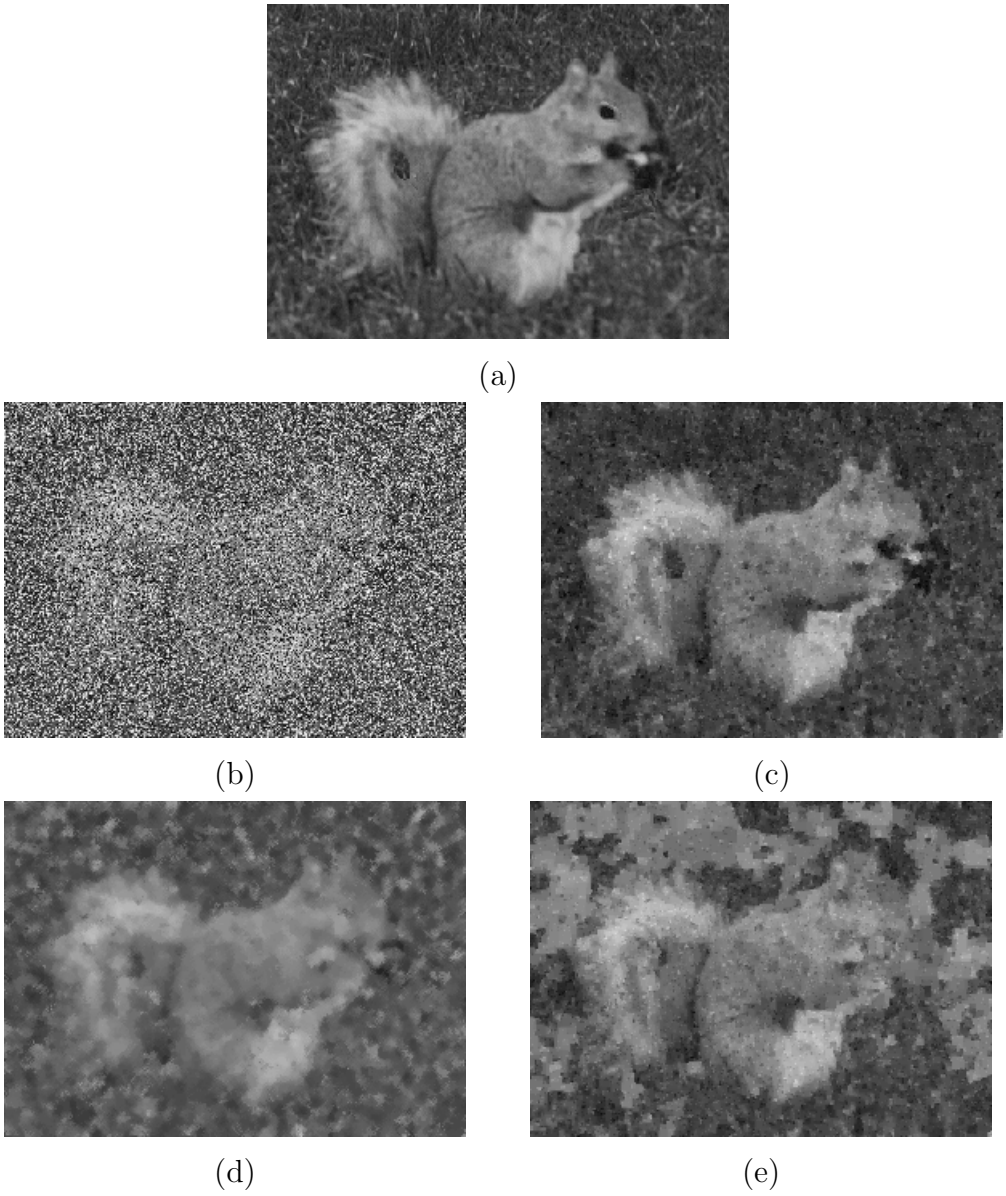


Fig. 2. (a): the *squirrel* original image . (b) and (c): impulsive noisy version with $P = 0.7$ and its TV restoration. (d): result obtained by median filtering. (e): result obtained by minimizing TV with the $\alpha-$expansion algorithm.

result when the weight coefficient β is set to $\beta = 120$. To better observe the behavior of the model, a zoom on both the original image and on the results are shown in Figure 5. It takes 2s to perform the minimization on an Intel Q9650 processor. Let us note that there are many available and well known techniques for dithering purposes. We do not claim that the simple model proposed here is a state-of-the-art approach.

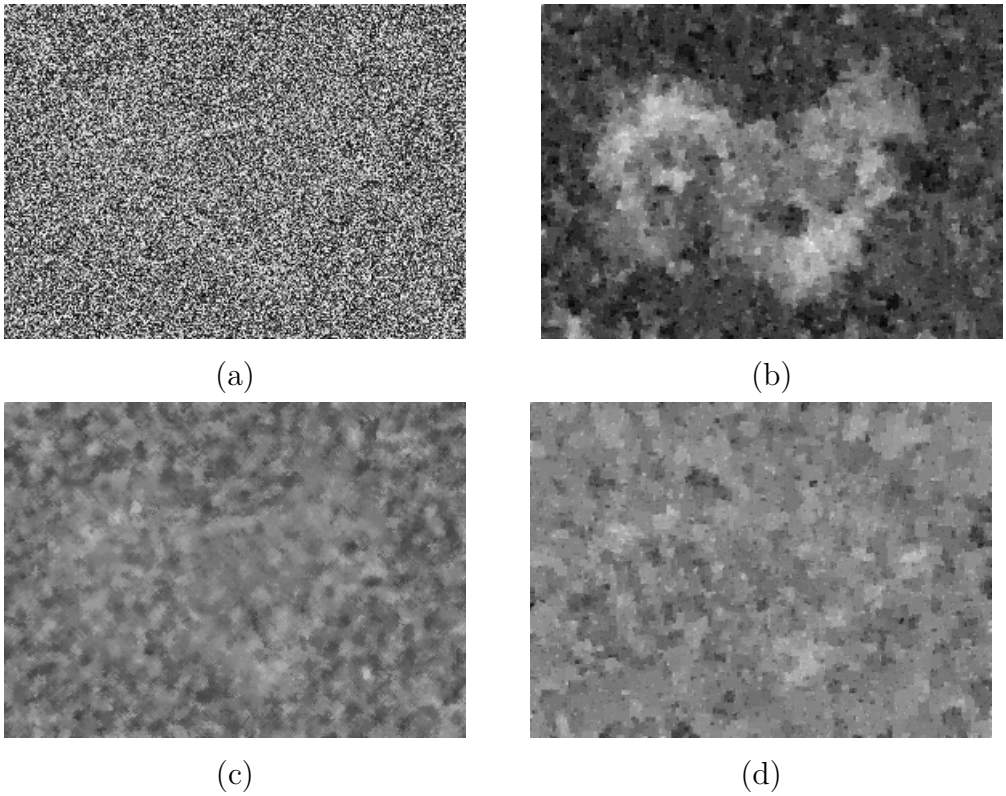


Fig. 3. The *squirrel* image corrupted with an impulsive noise with $P = 0.9$ is depicted in (a) while the restorations by global TV minimization,median filtering and TV minimization with by the $\alpha-$expansion algorithm are respectively shown in (b), (c) and (d).

6 Conclusion

We have presented a method to globally optimize a Markovian energy with pairwise interactions whose priors are submodular functions. The original problem is restated as as binary optimization for which, an efficient graph-based approach can be used. This mapping relies on considering the level-sets of the labels. The submodularity of the priors has been shown to be a necessary and sufficient condition for the applicability the proposed approach. As future work, we plan to iteratively use this approach on a restricted set of labels in order to perform the minimization while reducing the size of the induced graph [12,61,48].

A Proof of Proposition 1

Recall the binary case is considered, i.e. $\forall p \in \mathcal{V}\ u_p \in \{0, 1\}$. Let us write a binary function of two variables g_{pq} as follows:

(a)

(b)

Fig. 4. The original image *Lena* is depicted in (a) while the halftoning result is shown on (b). The weight coefficient β is set to $\beta = 150$.

$$g_{pq}(u_p, u_q) = g_{pq}(0,0)(1-u_p)(1-u_q) + g_{pq}(1,1)u_p u_q \\ + g_{pq}(0,1)(1-u_p)u_q + g_{pq}(1,0)u_p(1-u_q) \quad .$$

(a)

(b)

Fig. 5. Zoom of the original *lena* and of the halftoning result are respectively depicted in (a) and (b).

A simple rewriting gives an expansion on the base $(1, u_p, u_q, u_p u_q)$ so that we get:

$$\begin{aligned} g_{pq}(u_p, u_q) = &\left(g_{pq}(1,0) - g_{pq}(0,0)\right) u_p + \left(g_{pq}(0,1) - g_{pq}(0,0)\right) u_q \\ &+ \left(g_{pq}(0,0) + g_{pq}(1,1) - g_{pq}(0,1) - g_{pq}(1,0)\right) u_p u_q \\ &+ g_{pq}(0,0) \ . \end{aligned}$$

The Picard and Ratliff condition [55] states that the coefficient of the pairwise interaction shall be non positive, that is:

$$g_{pq}(0,0) + g_{pq}(1,1) - g_{pq}(0,1) - g_{pq}(1,0) \leq 0 \ .$$

The latter is the submodular inequality for the function g_{pq} defined on the lattice $\{0,1\}^2$ endowed with the usual order. This concludes the proof.

Acknowledgements

A preliminary version of the paper was presented at the 12th International Workshop on Combinatorial Image Analysis, Buffalo, NY, USA, April 2008 NJ [23]. The author deeply thanks Marc Sigelle (Telecom ParisTech) for fruitful discussions and careful proofreading, Bülent Sankur (Bogazici Univeristy) and TUBITAK for funding his stay at Bogazici University, Istanbul, in April and May 2006 during which which most of this paper was written. The author also thanks Olga Veksler for pointing out the reference [57]. Part of this work was done while the author was with the Research and Development Laboratory of EPITA (LRDE), Paris, France. This work was also supported by grants from the ONR under contract ONR N000140710810 and also partially through NSF DMS-0610079 and ONR N00014-06-1-0345,

References

[1] R. Ahuja, T. Magnanti, J. Orlin, Network Flows: Theory, Algorithms and Applications, Prentice Hall, 1993.

[2] R. K. Ahuja, D. S. Hochbaum, J. B. Orlin, Solving the convex cost integer dual network flow problem, Management Science 49 (7) (2003) 950–964.

[3] R. K. Ahuja, D. S. Hochbaum, J. B. Orlin, A cut-based algorithm for the nonlinear dual of the minimum cost network flow problem, Algorithmica 39 (3) (2004) 189–208.

[4] A. Amini, T. Weymouth, R. Jain, Using dynamic programming for solving variational problems in vision, IEEE Transactions on Pattern Analysis and Machine Interaction 12 (9) (1990) 855–867.

[5] F. Barahona, Finding ground states in random-field ising ferromagnets, Journal of Physics A 18 (1985) 673–675.

[6] J. Besag, On the statistical analysis of dirty pictures, Journal of the Royal Statistics Society 48 (1986) 259–302.

[7] J. Bioucas-Dias, G. Valado, Phase unwrapping via graph cuts, IEEE Transactions on Image processing 16 (3) (2007) 698–709.

[8] A. Blake, A. Zisserman, Visual Reconstruction, MIT Press, 1987.

[9] Y. Boykov, G. Funka-Lea, Graph cuts and efficient n-d image segmentation, International Journal of Computer Vision 70 (2) (2006) 109–131.

[10] Y. Boykov, M.-P. Jolly, Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images, in: in the Proceedings of the IEEE Tenth International Conference on Computer Vision (ICCV), 2001.

[11] Y. Boykov, V. Kolmogorov, An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision, IEEE Transactions on Pattern Analysis and Machine Interaction 26 (9) (2004) 1124–1137.

[12] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Transactions on Pattern Analysis and Machine Interaction 23 (11) (2001) 1222–1239.

[13] P. Carnevali, L. Coletti, S. Patarnello, Image porcessing by simulated annealing, IBM Journal of Research and Development 29 (6) (1985) 569–579.

[14] A. Chambolle, Total variation minimization and a class of binary mrf models, in: Energy Minimization Methods in Computer Vision and Pattern Recognition, vol. LNCS 3757, St. Augustine, Florida, USA, 2005.

[15] A. Chambolle, J. Darbon, On total variation minimization and surface evolution using parametric maximum flows, Tech. Rep. 08-19, UCLA (2008).

[16] T. Chan, S. Esedoglu, Aspect of Total Variation regularized L^1 Function Approximation, SIAM journal of Applied Mathematics 65 (5) (2005) 1817–1837.

[17] T. Chan, S. Esedoglu, M. Nikolova, Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models, Tech. rep., UCLA (Sep. 2004).

[18] R. Chellappa, A. Jain (eds.), Markov Random Fields: Theory and Application, Academic Press, 1993.

[19] L. Cohen, R. Kimmel, Global minimum for active contour models: A minimal path approach, International Journal of Computer Vision 24 (1) (1997) 57–78.

[20] K. Croxton, B. Gendron, T. Magnanti, A comparison of mixed-integer programming models for nonconvex piecewise linear cost minimization problem, Management Science 49 (9) (2003) 1268–1273.

[21] G. Dantzig, Linear Programming and Extensions, Princeton University Press, 1963.

[22] J. Darbon, Composants logiciels et algorithmes de minimisation exacte d'nergies ddis au traitement des images, Ph.D. thesis, Ecole Nationale Suprieure des Tlcommunications (October 2005).

[23] J. Darbon, Global optimization for first order Markov Random Fields with submodular priors, in: V. Brimkov, R. Barneva, H. Hauptman (eds.), Combinatorial Image Analysis, Lecture Notes in Computer Science Series, Vol. 4958, Springer, Berlin, 2008.

[24] J. Darbon, S. Peyronnet, A vectorial self-dual morphological filter based on total variation minimization, in: In the proceedings of the International Symposium on Visual Computing (ISVC 2005), Springer-Verlag, 2005.

[25] J. Darbon, M. Sigelle, Image restoration with discrete constrained Total Variation part I: Fast and exact optimization, Journal of Mathematical Imaging and Vision 26 (3) (2006) 261–276.

[26] J. Darbon, M. Sigelle, Image restoration with discrete constrained Total Variation part II: Levelable functions, convex priors and non-convex cases, Journal of Mathematical Imaging and Vision 26 (3) (2006) 277–291.

[27] L. Denis, F. Tupin, J. Darbon, M. Sigelle, Sar image regularization with fast approximate discrete minimization,, Tech. Rep. CAM-07-38, UCLA (September 2007).

[28] S. Durand, M. Nikolova, Stability of the minimizers of least squares with a non-convex regularization. part I: Local behavior, Journal of Applied Mathematics and Optimization 53 (2) (2006) 185–208.

[29] S. Durand, M. Nikolova, Stability of the minimizers of least squares with a non-convex regularization. part II: Global behavior, Journal of Applied Mathematics and Optimization 53 (3) (2006) 259–277.

[30] D. Freedman, P. Drineas, Energy minimization via graph cuts: settling what is possible, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, 2005.

[31] M. R. Garey, D. S. Johnson, Computers and Intractability, Freeman, 1979.

[32] S. Geman, D. Geman, Stochastic relaxation, gibbs distributions, and the bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Interaction 6 (6) (1984) 721–741.

[33] G. L. Gimel'farb, Image Textures and Gibbs Random Fields, Kluwer Academic Publishers, 1999.

[34] D. Goldfarb, Y. Yin, Parametric maximum flow algorithms for fast total variation minimization, Tech. rep., Rice University (2007).

[35] D. Greig, B. Porteous, A. Seheult, Exact maximum a posteriori estimation for binary images, Journal of the Royal Statistics Society 51 (2) (1989) 271–279.

[36] F. Guichard, J.-M. Morel, Image Iterative Smoothing and PDE s., downloadable manuscript : please write email to fguichard@poseidon-tech.com, 2000.

[37] F. Guichard, J.-M. Morel, Mathematical morphology "almost everywhere", in: Proceedings of International Symposium on Mathematical Morphology, Csiro Publishing, 2002.

[38] A. Hartmann, K. Usadel, Exact determination of all ground states of random field systems in polynomial time, Physica A 214 (2) (1995) 141–152.

[39] D. Hochbaum, An efficient algorithm for image segmentation, Markov Rand om Fields and related problems, Journal of the ACM 48 (2) (2001) 686–701.

[40] D. Hochbaum, J. Shanthikumar, Convex separable optimization is not muchharder than linear optimization, Journal of the ACM 37 (1990) 843–862.

[41] H. Ishikawa, Exact optimization for markov random fields with convex priors, IEEE Transactions on Pattern Analysis and Machine Interaction 25 (10) (2003) 1333–1336.

[42] S. Iwata, L. Fleischer, S. Fujishige, A combinatorial strongly polynomial algorithm for minimizing submodular functions, Journal of the ACM 48 (4) (2001) 761–777.

[43] D. Jungnickel, Graphs, Networks and Algorithms, Springer-Verlag, 2005.

[44] V. Kolmogorov, primal-dual algorithm for convex markov random fields, Tech. rep., Microsoft Research (2005).

[45] V. Kolmogorov, R. Zabih, What energy can be minimized via graph cuts?, IEEE Transactions on Pattern Analysis and Machine Interaction 26 (2) (2004) 147–159.

[46] V. N. Kolmogorov, M. J. Wainwright, On optimality of tree-reweighted max-product message-passing, in: Proceedings of the 21st conference on Uncertainty in Artificial Intelligence, Edinburgh, Scotland., 2005.

[47] S. Z. Li, Markov Random Field Modeling in Image Analysis, 2nd ed., Springer-Verlag, 2001.

[48] X. Liu, O. Veksler, J. Samarabandu, Graph cut with ordering constraints on labels and its applications, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2007.

[49] L. Lovász, Mathematical Programming – The state of the Art, chap. Submodular functions and convexity, Springer-Verlag, 1983, pp. 235–257.

[50] P. Maragos, R. Ziff, Threshold superposition in morphological image analysis systems, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (5) (1990) 498–504.

[51] K. Murota, Discrete Convex Analyis, SIAM Society for Industrial and Applied Mathematics, 2003.

[52] M. Nikolova, Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares, SIAM Journal on Multiscale Modeling and Simulation 4 (3) (2005) 960–991.

[53] M. Nikolova, Model distortions in bayesian MAP reconstruction, AIMS Journal on Inverse Problems and Imaging 1 (2) (2007) 399–422.

[54] A. Ogielski, Integer optimization and zero-fixed point in ising random-field systems, Physical Review Letters 57 (10) (1986) 1251–1254.

[55] J. Picard, H. Ratlif, Minimum cuts and related problem, Networks 5 (1975) 357–370.

[56] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D. 60 (1992) 259–268.

[57] D. Schlesinger, B. Flach, Transforming an arbitrary minsum problem into a binary one, Tech. rep., Dresden University of Technology (2008).

[58] A. Schrijver, A combinatorial algorithm minimizing submodular functions in strongly polynomial time, Journal of Combinatorial Theory Series B 80 (2) (2000) 346–355.

[59] L. Schrijver, Combinatorial Optimization - Polyhedra and Efficiency, Springer-Verlag, 2003.

[60] A. B. V. Lempitsky, C. Rother, Logcut - efficient graph cut optimization for markov random fields, in: In the Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2007.

[61] O. Veksler, Graph cut based optimization for mrfs with truncated convex priors, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2007.

[62] M. J. Wainwright, T. S. Jaakkola, A. S. Willsky, MAP estimation via agreement on (hyper)trees: Message-passing and linear-programming approaches, IEEE Transactions on Information Theory 51 (11) (2005) 3697–3717.

[63] G. Winkler, Image Analysis, Random Fields and Dynamic Monte Carlo Methods. A Mathematical Introduction, Applications of mathematics, 3rd ed., Springer-Verlag, 2006.

[64] B. Zalesky, Network flow optimization for restoration of images, J. Appl. Math. 2 (4) (2002) 199–218.

[65] B. Zalesky, Efficient determination of gibbs estimator with submodular energy functions, Tech. rep., United Institution of Information Problem (2005).