

Global Optimization for First Order Markov Random Fields with Submodular Priors

Jérôme Darbon

UCLA Mathematics Department, USA
E-mail: jerome@math.ucla.edu

Abstract. This paper copes with the optimization of Markov Random Fields with pairwise interactions defined on arbitrary graphs. The set of labels is assumed to be linearly ordered and the priors are supposed to be submodular. Under these assumptions we propose an algorithm which computes an *exact* minimizer of the Markovian energy. Our approach relies on mapping the original into a combinatorial one which involves only binary variables. The latter is shown to be exactly solvable via computing a maximum flow. The restatement into a binary combinatorial problem is done by considering the level-sets of the labels instead of the label values themselves. The submodularity of the priors is shown to be a necessary and sufficient condition for the applicability of the proposed approach.

1 Introduction

Many early vision problems can be formulated as an optimization problem. In particular, Markov Random Fields (MRFs) models have been widely used [21] since the seminal work of Geman et al. [9]. These energies are generally a weighted combination of two terms: the fidelity term and the prior. The first one measures the fidelity of the reconstructed solution with the observed data while the second one contains some knowledge on the result. It is generally hard to find a global optimum since these energies are usually non-convex. For some particular cases, computations are tractable using dynamic programming [2]. However for most of problems, considered energies remain difficult to optimize in general and these optimization problems can even be NP-hard [12]. A general practice is to use Simulated Annealing [9, 21] although it may be extremely slow in practice. This paper focus on Markovian energies that involves pairwise interactions and any 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. Compared to non-global optimization algorithms, global minimization algorithms allow to study the practical performance of a model. Besides, the approach proposed in this paper can be seen as a complementary computational point of view to the theoretical work of Nikolova on the property of global minimizers[8, 17, 18].

Let us define the problem of minimizing a Markovian energy with pairwise interactions in the context of computer vision. Assume that images are defined

on a set of nodes \mathcal{V} with cardinality $|\mathcal{V}|$. The value of the image u at a site $p \in \mathcal{V}$ is denoted by u_p . The lattice is endowed with a neighboring system and the neighborhood relationship between two adjacent sites p and q is denoted by $p \sim q$. Only pairwise interactions are considered, and such a clique is referred to as (p, q) where $p \sim q$. We denote by \mathcal{E} the set of all cliques. Thus we are interested in minimizing the following Markovian energy:

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

where v is the observed image, and the functions $\{f_p\}$ and $\{g_{pq}\}$ are respectively the fidelity terms and the priors.

In the seminal work [19], Picard and Ratliff show how a subclass of this energy can be optimized by computing a maximum-flow/s-t minimum-cut [1] on a graph associated to this energy. Then Greig et. al. use this approach in [10] to study the behavior of the Ising model for binary image restoration. In [4] Boykov et al. applies this technique for computer vision applications along with an excellent approximation result for the non-binary case. In [14], Kolmogorov and Zabih give a sufficient and necessary condition for the optimization of boolean MRF with pairwise and also triplewise interactions via s-t minimum-cut.

Extension of these approaches for exact optimization of MRFs involving more than two labels have been tackled by some authors. Approaches assume that labels can be linearly ordered and there are no assumptions on fidelity terms. In [13], a graph construction is proposed for MRFs where the priors are convex functions of the difference of labels, i.e. $g_{pq}(\cdot - \cdot)$ where g_{pq} are convex functions. The convexity assumption is shown to be sufficient and necessary. In [22], a class of MRFs whose energies can be rewritten as particular boolean MRF associated to each level is studied. In [6], the two above classes of Markovian energies are considered. The above assumptions allow the authors to devise a graph construction scheme for which a s-t minimum-cut yields a global minimizer minimization. Note that the topology of the underlying graph are different for each method but the size, i.e., the number of nodes and arcs, is the same. The optimization approach we propose in this paper can cope with all the above cases.

In this paper it is assumed that u_p takes value in the discrete set discrete $\mathcal{L} \subset \mathbb{R}$ of cardinality $|\mathcal{L}| = L$. This set is assumed to be linearly ordered, i.e., $\mathcal{L} = \{l_0, \dots, l_{L-1}\}$ with $l_i < l_{i+1} \forall i \in \llbracket 0, L-2 \rrbracket$. We also assume 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 will be referred to as discrete functions. In this paper, the priors $\{g_{pq}\}$ shall be submodular functions. For any positive integer k , a function $g : \mathcal{L}^k \rightarrow \mathbb{R}$ is said submodular if and only if it satisfies the following inequality [16]:

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

where $(x \vee y)$ and $(x \wedge y)$ respectively corresponds to the component-wise minimum and maximum between x and y , i.e., $\forall p \in \mathcal{V} \quad (x \vee y)_p = \min\{x_p, y_p\}$ and $(x \wedge y)_s = \max\{x_p, y_p\}$. Submodularity can be seen as a general property of

discrete functions that are analogous to convexity of functions defined on continuous domain [16].

The main theoretical contributions of this paper are the following. First, we propose an algorithm which computes a global minimizer for MRFs with pairwise interactions where priors are submodular functions. No assumption is set on data fidelity terms. Our approach relies on restating this problem into a binary optimization problem that can be exactly solved with a maximum-flow-based approach [4, 10, 14, 19]. Our mapping to the binary formulation makes use of the level sets of the labels. Second, it is shown that submodularity of the priors is a sufficient and necessary conditions for the application is the proposed approach. To our knowledge, these results are new and considerably extend previous available approaches for global MRF optimization. The complexity of our algorithm is pseudo-polynomial [1].

The remainder of this paper is organized as follows. Section 2 describes how one can rewrite data fidelity and prior terms using the level sets of the variables. These rewritings are the core of our restatement of the original minimization problem to a binary minimization one. In Section 3 we cope with exact optimization of MRFs with submodular priors. Finally we draw some conclusions in Section 4.

2 Development through Level Sets

This Section is devoted to rewrite every single data fidelity term $f_p(\cdot)$ and all prior terms $g_{pq}(\cdot, \cdot)$ appearing in the Markovian energy E defined by equation (1), as a linear combination of binary energies. These restatements will be used for optimizing exactly first order MRFs with submodular priors. This mapping is achieved thanks to the level sets of a label. We first define the notion of level sets and then we give the developments on level sets for functions of one and two variables.

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

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

The level sets of a variable x satisfies an monotone property:

$$\forall \lambda \leq \mu \quad [x]_\lambda \geq [x]_\mu , \quad (3)$$

The original gray-level value x can be reconstructed from its level sets using the following equality as shown in [11, 15]:

$$x = \max\{\lambda \in \mathcal{L}, [x]_\lambda = 0\} . \quad (4)$$

Conversely, it is shown in [11, 15] that any family of binary variables $\{[x]_\lambda\}_{\lambda=0\dots L-1}$ which satisfies the monotone properties, given by equation (3), define a label.

In other words, knowing the label itself or its binary representation in terms of level sets are equivalent.

The next proposition gives a development for data fidelity term as a summation on the level sets of its variable. It is based on a "discrete" integration of the "discrete" variations of f_p over its level sets.

Proposition 1 *Any data fidelity term $f_p : \mathcal{L} \mapsto \mathbb{R}$ rewrites on its level sets as follows:*

$$f_p(x) = \sum_{i=0}^{L-2} D_p(i)[x]_{l_i} + f_p(l_0) , \quad (5)$$

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

The proof is a straightforward extension of a similar proposition in [5].

Next, we extend the previous result to cope with functions of two variables. A natural way to perform it consists of applying the previous development firstly on the first variable and then on the second one. By rearranging terms it yields the following level sets-based developments.

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

$$\begin{aligned} g_{pq}(x, y) = & \sum_{i=0}^{L-2} \sum_{j=0}^{L-2} R_{pq}(i, j)[x]_{l_i}[y]_{l_j} \\ & + \sum_{i=0}^{L-2} (D_{pq}^1(i)[x]_{l_i} + D_{pq}^2(i)[y]_{l_i}) + C , \end{aligned} \quad (6)$$

where

$$\forall i \in \llbracket 0, L-2 \rrbracket \quad D_{pq}^1(l_i) = g(l_{i+1}, l_0) - g(l_i, l_0) ,$$

and

$$\forall i \in \llbracket 0, L-2 \rrbracket \quad D_{pq}^2(l_i) = g(l_0, l_{i+1}) - g(l_0, l_i) ,$$

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

$$\forall (i, j) \in \llbracket 0, L-2 \rrbracket^2 \quad R_{pq}(i, j) = g(l_{i+1}, l_{j+1}) - g(l_{i+1}, l_j) - g(l_i, l_{j+1}) + g(l_i, l_j) \quad (7)$$

So far, we have made no assumptions on data fidelity terms and on priors. In other words, results given in Proposition 1 and Proposition 2 hold for any function of one and two variables, respectively. In the next section, we specialize these level sets developments in order to globally optimize MRFs with submodular priors.

3 MRFs with Submodular Priors

In this Section, we assume that all priors $\{g_{st}\}$ are submodular functions and we show that such MRFs can be exactly optimized via computing a maximum flow on an associated graph [4, 14, 19]. Our approach consist of first applying the previous proposition to restate the original energy given by Eq. (1) in terms of binary variables. So we rewrite all data fidelity and prior terms using the expansions given by Proposition 1 and Proposition 2, respectively. So we get:

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

where the constant K comes from the constant C in the previous propositions. Note that the latter rewriting of the energy $E(u|v)$ only involve the level sets of the image u , i.e., $\{[u]_{l_i}\}_{i=0 \dots L-1}$. So let us define a new energy \tilde{E} whose variables are now $|L|$ binary images $\{b^i\}_{i=0 \dots L-1}$ as follows:

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

Now 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 Eq (3), then this family defines an image using the reconstruction given by Eq (4). However if any of the families $\{[b_s^i]_\lambda\}_{\lambda=0 \dots L-1}$ violates the monotone property, then a gray level image cannot be defined. Besides note that for any image u we have $E(u|v) = \tilde{E}(\{[u]_\lambda\}_{\lambda=0 \dots L-1}|v)$. Thus, if we are able to minimize the energy $E(\{[\cdot]_\lambda\}_{\lambda=0 \dots L-1}|v)$ while preserving the monotone property, then we get a global minimizer of $E(\cdot|v)$. In order to force the monotone property to hold we define the following new energy:

$$\tilde{E}_\alpha(\{b_i\}_{i=0 \dots L-1}|v) = \tilde{E}(\{b_i\}_{i=0 \dots L-1}|v) + \sum_{p \in \mathcal{V}} \alpha \sum_{i=0}^{L-2} H(b_p^{i+1} - b_p^i) , \quad (8)$$

where $H : \mathbb{R} \mapsto \mathbb{R}$ is the Heaviside function defined as $H(x) = 0$ if $x \leq 0$ and 1 else. It is shown in [6] that if α is set to a sufficiently large finite value, then we are assured that any global minimizer of $E_\alpha(\{\cdot\}_{i=0 \dots L-1}|v)$ never violates the monotone property give by Eq. (3).

Now we show that the boolean energy (8) can be optimized via a maximum flow or by duality a s-t minimum-cut [1]. Following the seminal work of [19] or

equivalently [4, 14] it is enough to show that every pairwise interaction terms of *binary* variables are submodular. Specializing the definition of submodularity, Eq. (2) for a binary function f of two variables, i.e., $f : \{0, 1\}^2 \rightarrow \mathbb{R}$, we get that:

$$f(0, 0) + f(1, 1) \leq f(0, 1) + f(1, 0) . \quad (9)$$

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 [6] 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 (9) it means to show that $\forall (i, j) \in \llbracket 0, L-2 \rrbracket^2 \quad R(i, j) \leq 0$. This property is assured by the submodularity assumption of the priors, Eq. (2), as shown in the next proposition.

Proposition 3 *Assume $g : \mathcal{L}^2 \mapsto \mathbb{R}$. The following two assertions are equivalent:*

1. g is submodular,
2. g writes as

$$g(x, y) = \sum_{i=0}^{L-2} \sum_{j=0}^{L-2} R(i, j)[x]_{l_i}[y]_{l_j} + \sum_{i=0}^{L-2} (D^+(i)[x]_{l_i} + D^-(i)[y]_{l_j}) + C , \quad (10)$$

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

Proof. Case 1) \Rightarrow 2) We apply Proposition 2 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. (9) applied for the remaining terms $R(i, j)[x]_{l_i}[y]_{l_j}$, reduces to show that $\forall (i, j) \in \llbracket 0, L-2 \rrbracket^2 \quad R(i, j) \leq 0$.

Recall that Eq. (7) of Proposition 2 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 couples $a = (l_i, l_{j+1})$ 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 $x \in \mathcal{L}^2$ and $y \in \mathcal{L}^2$. Note that the only interesting case happens when $x \notin \{(x \vee y) \cup (x \wedge y)\}$ (otherwise the submodularity property is obviously satisfied).

Let us denote by $(x_m, y_m) = (x \wedge y)$ and $(x_M, y_M) = (x \vee y)$. 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. (10). One sees that the constant C and the terms involving the single summation $(\sum_{i=0}^{L-2} \cdot)$ cancels each other. Thus only the double summation terms remain, i.e., we need to show:

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

which is equivalent to

$$\sum_{i=0}^{L-2} \sum_{j=0}^{L-2} R(i, j) ([x_M]_{l_i} - [x_m]_{l_i}) ([y_M]_{l_j} - [y_m]_{l_j}) \leq 0 . \quad (11)$$

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

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

and thus every term in the double summation in (11) are non-positive since $R(i, j) \leq 0$. This concludes the proof. \square

So far we have shown that the binary energy 8 can be exactly optimized using a maximum flow approach [4, 10, 14, 19]. Minimizing the latter energy is equivalent to minimize a first order MRF with submodular priors. Note that Proposition 3 gives a sufficient and necessary condition for applying the proposed approach. This result highly generalizes the results presented in [6] and [13].

We now consider the case where the priors are a unary function of the difference of the labels. These are widely used in image analysis because it corresponds to regularize the gradient of an image. The most well-known example of such a prior is most probably the Total Variation [20]. Under the above assumption the next proposition shows that only a *convex* regularization of the difference of the labels can be considered using the approach presented in this paper.

Proposition 4 *Assume $g : \mathcal{L}^2 \rightarrow \mathbb{R}$ is submodular and has the following $g(x, y) = \tilde{g}(x - y)$ then \tilde{g} is a unary convex function.*

Proof. First we apply Prop 2. 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 $2\tilde{g}(k) \leq \tilde{g}(k+1) + \tilde{g}(k-1)$ which is exactly the discrete second variation convexity criteria for a unary function [16] applied for \tilde{g} . \square

Note that although computing a maximum can be performed in polynomial time [1] our approach, like those of [6] and [13], is not. Indeed, an algorithm has a polynomial time if it performs a polynomial number of operations with respect to

the number of bits required to describe the optimization problem. The necessary number of bits to describe an integer n is $\lceil \log_2 n \rceil$. The graph we built has for each pixel one node per gray level (i.e., for each pixel we have $(L-1)$ nodes) and is thus exponential with respect to $\lceil \log_2 n \rceil$. This exponential behavior prevents us from applying this approach on very large images (such as 3D volumes) because it requires too much memory. However, the proposed approach is manageable for standard size images and we refer the reader to [6] and [7] for image restoration and time results. Finally note that the maximum flow algorithm described in [3] has been shown to be extremely efficient in practice though its time complexity might not be polynomial. For image processing purposes it outperforms other polynomial algorithms as reported in [3]. This makes the method applicable as reported in [6] and [7].

4 Conclusion

In this paper we have presented a method to globally optimize a Markovian energy with pairwise interactions whose priors are submodular functions. The approach consists of restating the original problem as a binary optimization problem that can be efficiently solved using a graph approach. The binarization makes use of the level set of the image. The submodularity of the priors has been shown to be a necessary and sufficient condition for the applicability of the proposed approach.

Acknowledgements

The author deeply thanks Marc Sigelle (Télécom Paris, France) for fruitful discussions and careful proofreading, Bülent Sankur (Bogazici University, Istanbul, Turkey) and TUBITAK for funding his April and May 2006 stay at Bogazici University, during part of this paper was written, and finally the anonymous reviewers for helpful comments. This work was partly done while the author was with the Research and Development Laboratory of EPITA (LRDE), Paris, France. This research was also supported by ONR Grant N000140710810.

References

1. R.K Ahuja, T.L. Magnanti, and J.B. Orlin. *Network Flows: Theory, Algorithms and Applications*. Prentice Hall, 1993.
2. A. Amini, T. Weymouth, and R. Jain. Using dynamic programming for solving variational problems in vision. *IEEE Transactions on Pattern Analysis and Machine Interaction*, 12(9):855–867, 1990.
3. 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 Interaction*, 26(9):1124–1137, 2004.
4. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Interaction*, 23(11):1222–1239, 2001.

5. J. Darbon and M. Sigelle. Image restoration with discrete constrained Total Variation part I: Fast and exact optimization. *Journal of Mathematical Imaging and Vision*, 26(3):261–276, December 2006.
6. J. Darbon and 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):277–291, December 2006.
7. J. Darbon, M. Sigelle, and F. Tupin. The use of levelable regularization functions for mrf restoration of sar images while preserving reflectivity. In *the proceedings of the IS&T/SPIE 19th Annual Symposium Electronic Imaging (oral presentation). Conference on Computational Imaging (E112).*, San Jose, CA, USA, February 2007.
8. S. Durand and 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):259–277, 2006.
9. S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Interaction*, 6(6):721–741, 1984.
10. D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistics Society*, 51(2):271–279, 1989.
11. F. Guichard and J.-M. Morel. Mathematical morphology ”almost everywhere”. In *Proceedings of International Symposium on Mathematical Morphology*, pages 293–303. Csiro Publishing, April 2002.
12. D. S. Hochbaum. An efficient algorithm for image segmentation, markov random fields and related problems. *Journal of the ACM*, 48(2):686–701, 2001.
13. H. Ishikawa. Exact optimization for markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Interaction*, 25(10):1333–1336, October 2003.
14. V. Koltmogorov and R. Zabih. What energy can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Interaction*, 26(2):147–159, 2004.
15. P. Maragos and R.D. Ziff. Threshold superposition in morphological image analysis systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):498–504, 1990.
16. K. Murota. *Discrete Convex Optimization*. SIAM Society for Industrial and Applied Mathematics, 2003.
17. 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):960–991, 2005.
18. M. Nikolova. Model distortions in bayesian MAP reconstruction. *AIMS Journal on Inverse Problems and Imaging*, 1(2):399–422, 2007.
19. J.P. Picard and H.D. Ratlif. Minimum cuts and related problem. *Networks*, 5:357–370, 1975.
20. L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D.*, 60:259–268, 1992.
21. G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods. A Mathematical Introduction*. Applications of mathematics. Springer-Verlag, 3rd edition, 2006.
22. B. Zalesky. Efficient determination of gibbs estimator with submodular energy functions. Technical report, United Institution of Information Problem, 2005.