Pseudo-Boolean Optimization: Theory and Applications in Vision

Strandmark, Petter; Kahl, Fredrik

2012

Pseudo-Boolean Optimization: Theory and Applications in Vision

Petter Strandmark, Fredrik Kahl
Centre for Mathematical Sciences, Lund University
Email: {petter, fredrik}@maths.lth.se

Abstract—**Many problems in computer vision, such as stereo, segmentation and denoising can be formulated as pseudo-boolean optimization problems. Over the last decade, graphs cuts have become a standard tool for solving such problems. The last couple of years have seen a great advancement in the methods used to minimize pseudo-boolean functions of higher order than quadratic. In this paper, we give an overview of how one can optimize higher-order functions via generalized roof duality and how it can be applied to problems in image analysis and vision.**

I. Pseudo-boolean Optimization

A pseudo-boolean function is a function from the set of boolean vectors of dimension n, denoted $\mathbf{B}^n = \{0,1\}^n$, to the reals. Any pseudo-boolean f can be uniquely represented by a multilinear polynomial of the form

$$f(\boldsymbol{x}) = \sum_i a_i x_i + \sum_{i<j} a_{ij} x_i x_j + \sum_{i<j<k} a_{ijk} x_i x_j x_k + \dots$$

The degree m of f is equal to the degree of the corresponding polynomial. In this paper, we are interested in efficient ways of solving the following optimization problem:

$$\min_{\boldsymbol{x} \in \mathbf{B}^n} f(\boldsymbol{x}), \qquad (1)$$

in order to improve performance for many basic applications in computer vision.

Application problems that can be turned into such an optimization problem abound. For example, state-of-the-art methods for stereo, segmentation and image denoising are often formulated as the inference of the maximum a posteriori estimate in a Markov Random Field (MRF) and such problems can be formulated as energy minimization problems where the energy function is given by a pseudo-boolean function.

In general, the minimization problem in (1) is NP-hard so approximation algorithms are necessary. For the quadratic case ($m = 2$), one of the most popular and successful approaches is based on the roof duality bound [1], [2]. We generalize the roof duality framework for higher-order pseudo-boolean functions. Our main contributions are (i) how one can define a general bound for any order (for which the quadratic case is a special case), (ii) how one can efficiently compute solutions that attain this bound in polynomial time and (iii) to give example applications in vision for which performance is improved by this technique. Our results rely on many previous significant contributions.

A. Related work and Applications

Graph cuts is by now a standard tool for many vision problems, in particular, for the minimization of quadratic and cubic submodular pseudo-boolean functions [3], [4]. The same technique can be used for non-submodular functions in order to compute a lower bound [1].

In recent years, there has been an increasing interest in higher-order models and approaches for minimizing the corresponding energies. For example, in [5], approximate belief propagation is used with a learned higher-order MRF model for image denoising. Similarly, in [6], an MRF model is learned for texture restoration, but the model is restricted to submodular energies which can be optimized exactly with graph cuts. Curvature regularization requires higher-order models [7], [8]. Even global potentials defined over all variables in the MRF have been considered, e.g., in [9] for ensuring connectedness, in [10] to model co-occurrence statistics of objects. Another state-of-the-art example is [11] where second-order surface priors are used for stereo reconstruction. The optimization strategies rely on dual decomposition [12], [13], move-making algorithms [14], [15], linear programming [16], belief propagation [5] and, of course, graph cuts.

B. Submodularity

There is a large subset of pseudo-boolean which contains functions which are easy to minimize: the set of submodular functions. If we let $\boldsymbol{x} \wedge \boldsymbol{y}$ and $\boldsymbol{x} \vee \boldsymbol{y}$ mean element-wise $\min$ and $\max$, respectively, the set can be defined as the set of all functions f satisfying

$$f(\boldsymbol{x}) + f(\boldsymbol{y}) \geq f(\boldsymbol{x} \wedge \boldsymbol{y}) + f(\boldsymbol{x} \vee \boldsymbol{y}) \qquad (2)$$

for all points $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{B}^n$. In the quadratic ($m = 2$) and cubic ($m = 3$) cases, a submodular function f can be minimized very efficiently via graph cuts and in polynomial time for any degree. However, for $m \geq 4$, it is a very diffcult problem (co-NP-complete) to determine whether a given function is submodular. [17]

As an example, the function

$$f(\boldsymbol{x}) = x_1 - 4x_2 + 2x_3 - 5x_1 x_2 \qquad (3)$$

is submodular but the function

$$f(\boldsymbol{x}) = x_1 - 4x_2 + 2x_3 + 5x_1 x_2 \qquad (4)$$

is not.

C. Roof Duality

The roof duality bound is an efficiently computable lower bound to the minimum of f in the quadratic case. While a good lower bound can be interesting in many applications, e.g. when using branch and bound, another property of roof duality is arguably even more useful: *persistency*. Each variable in the solution obtained via roof duality is equal to one of three possibilities: $\{0, 1, ?\}$. It is guaranteed that a global solution exists corresponding to the parts of the solution not equal to '?.' Therefore, the roof duality solution is said to be partially optimal.

The fastest method of computing the roof dual for quadratic pseudo-boolean polynomials is by using graph cuts. Reductions have been explored for higher-order polynomials ($m > 2$), e.g., [18]–[21]. These methods all convert a higher-order function to a quadratic one.

Our framework builds on [22] using submodular relaxations directly on higher-order terms. We define *optimal relaxations* to be those that give the tightest lower bound. As an example, consider the problem of minimizing the following cubic polynomial f over $\mathbf{B}^3$:

$$f(\boldsymbol{x}) = -2x_1 + x_2 - x_3$$
$$+ 4x_1x_2 + 4x_1x_3 - 2x_2x_3 - 2x_1x_2x_3. \quad (5)$$

The standard reduction scheme [21] would use the identity $-x_1x_2x_3 = \min_{z \in \mathbf{B}} z(2 - x_1 - x_2 - x_3)$ to obtain a quadratic minimization problem with one auxiliary variable z. Roof duality gives a lower bound of $f_{\min} \geq -3$, but it does not reveal how to assign any of the variables in $\boldsymbol{x}$. However, there are many possible reduction schemes from which one can choose. Another possibility is $-x_1x_2x_3 = \min_{z \in \mathbf{B}} z(-x_1 + x_2 + x_3) - x_1x_2 - x_1x_3 + x_1$. For this reduction, the roof duality bound is tight and the optimal solution $\boldsymbol{x}^* = (0, 1, 1)$ is obtained. This simple example illustrates two facts: (i) different reductions lead to different lower bounds and (ii) it is not an obvious matter how to choose the optimal reduction.

II. GENERALIZED ROOF DUALITY

A. Problem Formulation

Consider the optimization problem in (1) where f has degree m. In this paper, we will investigate the cases when $m = 3$ and $m = 4$. By enlarging the domain, we will relax the problem and look at the following tractable problem:

$$\min_{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbf{B}^{2n}} g(\boldsymbol{x}, \boldsymbol{y}), \quad (6)$$

where $g : \mathbf{B}^{2n} \mapsto \mathbf{R}$ is a pseudo-boolean function that satisfies the three conditions

$$g(\boldsymbol{x}, \bar{\boldsymbol{x}}) = f(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathbf{B}^n, \quad \text{(A)}$$
$$g \text{ submodular}, \quad \text{(B)}$$
$$g(\boldsymbol{x}, \boldsymbol{y}) = g(\bar{\boldsymbol{y}}, \bar{\boldsymbol{x}}), \quad \forall (\boldsymbol{x}, \boldsymbol{y}) \in \mathbf{B}^{2n} \text{ (symmetry)}. \quad \text{(C)}$$

For a point $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbf{B}^n$, we denote $\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n) = (1 - x_1, 1 - x_2, \ldots, 1 - x_n)$. The reason for requirement (A) is that if the range of f is included in the range of g then the minimum of g is a lower bound to the minimum of f. If the computed minimizer $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ of the relaxation g happens to fulfill $\boldsymbol{x}^* = \bar{\boldsymbol{y}}^*$ then, of course, $\boldsymbol{x}^*$ is a minimizer of f as well. Even if it is not the case that $\boldsymbol{x}^* = \bar{\boldsymbol{y}}^*$, we still obtain a lower bound on f and as we shall see, it is possible to extract a partial solution for a minimizer of f.

Requirement (B) is also fairly obvious. Since we must be able to minimize g, requiring that g is submodular is natural. The last requirement is more technical and is required in order to prove persistency. It also turns out that including (C) does not give worse lower bounds.

Let $f_{\min}$ denote the unknown minimum value of f, that is, $f_{\min} = \min f(\boldsymbol{x})$. Ideally, we would like $g(\boldsymbol{x}, \boldsymbol{y}) \geq f_{\min}$ for all points $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbf{B}^{2n}$. This is evidently not possible in general. However, one could try to maximize the lower bound of g, $\max \min_{\boldsymbol{x}, \boldsymbol{y}} g(\boldsymbol{x}, \boldsymbol{y})$, that is,

$$\max_g \quad \ell$$
$$\text{such that} \quad g(\boldsymbol{x}, \boldsymbol{y}) \geq \ell, \quad \forall (\boldsymbol{x}, \boldsymbol{y}) \in \mathbf{B}^{2n}, \quad (7)$$
$$g \text{ satisfies (A)–(C)}.$$

A relaxation g that provides the maximum lower bound will be called *optimal*. As we have proved in [23], when $m = 2$, the lower bound coincides with the roof duality bound [1] and therefore this maximum lower bound will be referred to as *generalized roof duality*.

B. Properties of the Solution

By construction, the solution to (7) gives a lower bound to the optimal value of f. Another important property is that it satisfies persistency in the following way: Let g^* be the solution to (7) and $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in \arg\min_{\boldsymbol{x}, \boldsymbol{y}} g^*(\boldsymbol{x}, \boldsymbol{y})$. If we then let I be the set of all i such that $\boldsymbol{x}_i^* \neq \boldsymbol{y}_i^*$, then there exists a global minimum $\hat{\boldsymbol{x}}$ to our original function f such that $\hat{\boldsymbol{x}}_i = \boldsymbol{x}_i^*$ for all $i \in I$. See [23] for a proof.

In other words, the minimizer to g^* sometimes gives us partial information about a global minimizer of f, just as roof duality does.

C. Solving the Problem

While problem (7) is a linear program in the coefficients of g, the number of constraints is exponential in the number of variables n. Nevertheless, it is possible to attain the roof duality bound in polynomial time [23]. Requirements (A) and (C) never give any problems. Requirement (B) (submodularity) can be represented with polynomially many constraints only when $m \leq 3$. Recall that determining whether a function f is submodular is co-NP-complete. For $m = 4$ we therefore have to look at smaller sets than the entire set of submodular functions, as no reasonably sized set of linear constraint can describe all submodular functions.

The final hurdle is the exponentially many constraints involving ℓ in (7). We solved this in [23] using an iterative approach. Instead of maximizing $\min_{\boldsymbol{x}, \boldsymbol{y}} g(\boldsymbol{x}, \boldsymbol{y})$ it is enough to maximize $g(\mathbf{0}, \mathbf{0})$, which eliminates all constraints. Then, when a solution is obtained, g can be minimized. If any persistencies are available, they can be used to simplify f and the process starts over. Otherwise, the process terminates. This will give a bound which is at least as good as the bound from (7).

III. Experiments

We have performed experiments comparing the proposed generalized roof duality to the current state of the art. The methods we are considering are:

GRD	Generalized roof duality over all sub-modular functions for $m = 3$ and a subset for $m = 4$.
GRD-gen	Same as GRD, except over a larger subset for $m = 4$ [24] (which results in a larger linear program).
GRD-heuristic	Instead of solving a linear program (which can be expensive), a very fast heuristic is used to give an approximate solution.
Fix et al.	The reductions proposed in [25].
HOCR	The reductions proposed in [21].

A. Random Polynomials

In the first experiment, we generated polynomials with random coefficients:

$$f(\boldsymbol{x}) = \sum_{(i,j,k) \in T} f_{ijk}(x_i, x_j, x_k), \qquad (8)$$

where $T \subseteq \{1 \ldots n\}^3$ is a random set of triplets and each f_{ijk} is a cubic polynomial in x_i, x_j and x_k with all its coefficients picked uniformly in $\{-100, \ldots, 100\}$. The results for all methods can be see in Fig. 2.

B. Image Denoising

Ishikawa [21] used image denoising as a benchmark problem for higher-order pseudo-boolean minimization. In each iteration proposals are generated in two possible ways which are alternated: by blurring the current image and picking all pixels at random. Each pixel then have a choice of staying the same or switching to the proposal, resulting in a pseudo-boolean optimization problem. The smoothness term consists of a Fields of Experts (FoE) model using patches of size 2×2. Thus, quartic polynomials are needed to formulate the image restoration task as a pseudo-boolean minimization problem.

Figure 3 shows a comparison between the different methods for this problem. Generalized roof duality performed very well, often labeling very close to 100% of the problem variables. Figure 1 shows the results for a 160×240 image, where all three methods used have comparable running time.

References

[1] E. Boros and P. Hammer, "Pseudo-boolean optimization," *Discrete Appl. Math.*, vol. 123, pp. 155–225, 2002.

[2] V. Kolmogorov and C. Rother, "Minimizing nonsubmodular functions with graph cuts - a review," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 7, pp. 1274–1279, 2007.

[3] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[4] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147–159, 2004.

[5] X. Lan, S. Roth, D. Huttenlocher, and M. Black, "Efficient belief propagation with learned higher-order Markov random fields," in *European Conf. Computer Vision*, Graz, Austria, 2006.

[6] D. Cremers and L. Grady, "Statistical priors for efficient combinatorial optimization via graph cuts," in *European Conf. Computer Vision*, Graz, Austria, 2006.

[7] T. Schoenemann, F. Kahl, and D. Cremers, "Curvature regularity for region-based image segmentation and inpainting: A linear programming relaxation," in *Int. Conf. Computer Vision*, Kyoto, Japan, 2009.

[8] P. Strandmark and F. Kahl, "Curvature regularization for curves and surfaces in a global optimization framework," in *EMMCVPR*, St Petersburg, Russia, 2011.

[9] S. Nowozin and C. Lampert, "Global interactions in random field models: A potential function ensuring connectedness," *SIAM J. Imaging Sciences*, vol. 3, no. 4, pp. 1048–1074, 2010.

[10] L. Ladicky, C. Russell, P. Kohli, and P. Torr, "Graph cut based inference with co-occurrence statistics," in *European Conf. Computer Vision*, Crete, Greece, 2010.

[11] O. Woodford, P. Torr, I. Reid, and A. Fitzgibbon, "Global stereo reconstruction under second-order smoothness priors," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2115–2128, 2009.

[12] N. Komodakis and N. Paragios, "Beyond pairwise energies: Efficient optimization for higher-order MRFs," in *Conf. Computer Vision and Pattern Recognition*, Miami, USA, 2009.

[13] P. Strandmark and F. Kahl, "Parallel and distributed graph cuts by dual decomposition," in *Conf. Computer Vision and Pattern Recognition*, San Francisco, USA, 2010.

[14] P. Kohli, M. Kumar, and P. Torr, "P³ & beyond: Move making algorithms for solving higher order functions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1645–1656, 2009.

[15] V. Lempitsky, C. Rother, S. Roth, and A. Blake, "Fusion moves for Markov random field optimization," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1392–1405, 2010.

[16] T. Werner, "High-arity interactions, polyhedral relaxations, and cutting plane algorithm for MAP-MRF," in *Conf. Computer Vision and Pattern Recognition*, Anchorage, USA, 2008.

[17] Y. Crama, "Recognition problems for special classes of polynomials in 0-1 variables," *Math. Programming*, vol. 44, no. 1–3, pp. 139–155, 1989.

[18] S. H. Lu and A. C. Williams, "Roof duality for polynomial 0-1 optimization," *Math. Programming*, vol. 37, no. 3, pp. 357–360, 1987.

[19] D. Freedman and P. Drineas, "Energy minimization via graph cuts: Settling what is possible," in *Conf. Computer Vision and Pattern Recognition*, San Diego, USA, 2005.

[20] C. Rother, P. Kohli, W. Feng, and J. Jia, "Minimizing sparse higher order energy functions of discrete variables," in *Conf. Computer Vision and Pattern Recognition*, Miami, USA, 2009.

[21] H. Ishikawa, "Transformation of general binary MRF minimization to the first order case," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no. 6, pp. 1234–1249, 2011.

[22] V. Kolmogorov, "Generalized roof duality and bisubmodular functions," in *Neural Information Processing Systems*, Vancouver, Canada, 2010.

[23] F. Kahl and P. Strandmark, "Generalized roof duality for pseudo-boolean functions," in *Int. Conf. Computer Vision*, Barcelona, Spain, 2011.

[24] S. Živný and P. G. Jeavons, "Which submodular functions are expressible using binary submodular functions?" Oxford University Computing Laboratory, Tech. Rep. CS-RR-08-08, 2008.

[25] A. Fix, A. Grubner, E. Boros, and R. Zabih, "A graph cut algorithm for higher-order markov random fields," in *Int. Conf. Computer Vision*, 2011.
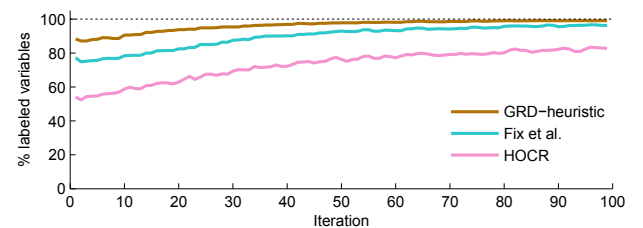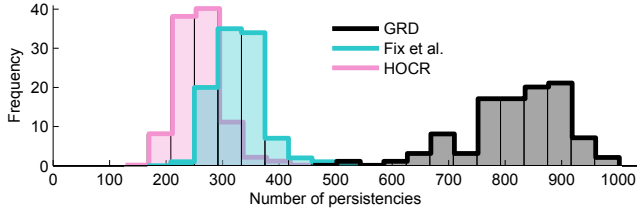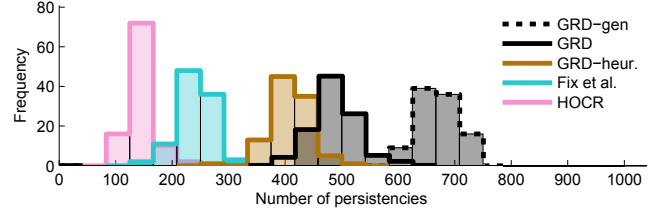
Fig. 1: Restoring a big image. See Fig. 3. Linear programming was not used for this experiment.

|	Rel. bounds			Time (ms)		
	Min	Med.	Max	Min.	Med.	Max
GRD	0	0	0	422	514	749
GRD-heur.	0.00	0.00	0.00	31	63	125
Fix et al.	0.05	0.09	0.13	15	16	32
HOCR	0.10	0.14	0.20	0	15	31

(a) Cubic polynomials with $n = 1000$ and $|T| = 1000$. GRD-heuristic is not shown in the histogram because it is almost indistinguishable from GRD.

	Rel. bounds			Time (ms)		
	Min	Med.	Max	Min.	Med.	Max
GRD-gen.	-0.10	-0.07	-0.05	360	438	560
GRD	0	0	0	599	810	1478
GRD-heur.	0.04	0.06	0.09	60	100	160
Fix et al.	0.13	0.16	0.20	0	10	20
HOCR	0.42	0.48	0.55	0	10	17

(b) Quartic polynomials with $n = 1000$ and $|T| = 300$.

Fig. 2: Number of persistencies, relative bounds and running time for 100 random polynomials. The set of coefficients T was drawn uniformly after making sure that all variables were used once.
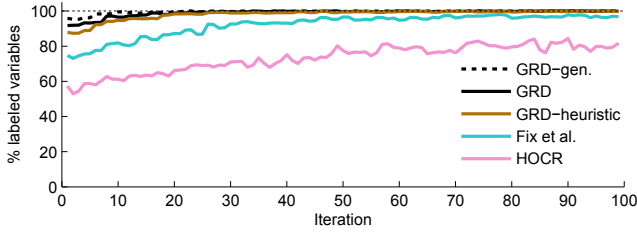
(a) Each method progresses independently
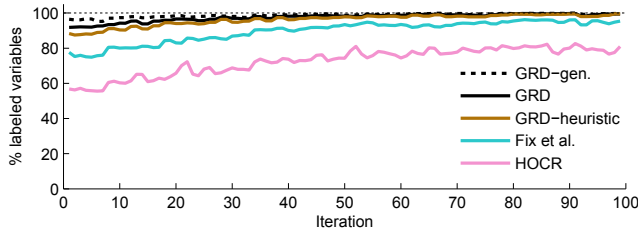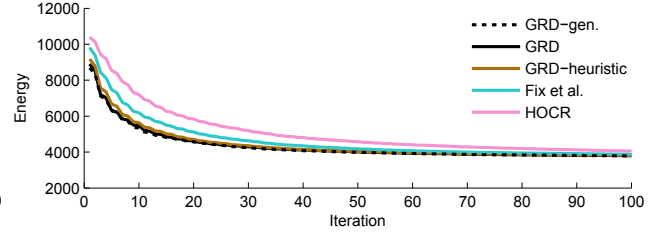
(b) Each method solves the same problem in each iteration

(c) Noisy image

(d) Restored image

Fig. 3: Restoring a small image. In each iteration a proposal is generated and each pixel can either stay the same or switch to the proposal. A quartic smoothness function is used.