Pseudo-bound Optimization for Binary Energies

Meng Tang[1], Ismail Ben Ayed[1,2], and Yuri Boykov[1]

[1] University of Western Ontario, Canada
[2] GE Healthcare, Canada

Abstract. High-order and non-submodular pairwise energies are important for image segmentation, surface matching, deconvolution, tracking and other computer vision problems. Minimization of such energies is generally *NP-hard*. One standard approximation approach is to optimize an *auxiliary function* - an upper bound of the original energy across the entire solution space. This bound must be amenable to fast global solvers. Ideally, it should also closely approximate the original functional, but it is very difficult to find such upper bounds in practice.

Our main idea is to relax the upper-bound condition for an auxiliary function and to replace it with a family of pseudo-bounds, which can better approximate the original energy. We use fast polynomial parametric maxflow approach to explore all global minima for our family of submodular pseudo-bounds. The best solution is guaranteed to decrease the original energy because the family includes at least one auxiliary function. Our Pseudo-Bound Cuts algorithm improves the state-of-the-art in many applications: appearance entropy minimization, target distribution matching, curvature regularization, image deconvolution and interactive segmentation.

Keywords: Binary energy minimization, high-order and non-submodular functions, auxiliary functions, parametric maxflow, pseudo-bounds.

1 Introduction

Recently high-order [2,3,12,13,15,26,29,37] and non-submodular pairwise [11,16,21,19] energy minimization have drawn tremendous research interests. Those energy functions arise naturally in many computer vision and image processing applications. Examples of high-order functions include but are not limited to constraints on segment volume [12,37], clique labeling consistency [13,18,35] and matching target distributions [2,3,12,29]. Pairwise non-submodular energies occur in deconvolution [11], curvature regularization [8,11,28], inpainting [16] and surface registration [16].

In general, optimization of high-order or non-submodular pairwise energy is *NP-hard*. Existing approximation methods make optimization tractable either by *global* or *local* linearization. Well established LP relaxation methods such as QPBO [4,30] and TRWS [19] are examples of global linearization techniques for solving non-submodular energies in vision [16]. By relaxing the integrality constraints, they globally transform the original function into a linear function

with extra variables and linear constraints. Unlike global linearization, local techniques iteratively approximate the original energy around current solution, for instance, using Taylor approximations [11,12,17,24] or auxiliary functions [2,3,11,27,29,31]. The recent Fast Trust Region (FTR) method [12] finds the optimal solution of a local approximation within a trust region, *i.e.* a region near current solution where the approximation can be trusted. The trust region size is adaptively adjusted depending on the quality of current approximation using well-known trust region paradigms [38]. The recent studies [11,12] have shown that FTR achieved the state of the art performance in many applications. Our work is more closely related to bound optimizers, which take auxiliary functions [2,3,11,22,27,29,31] as local approximations and were recently shown to yield competitive performances in several vision problems [3,11,29].

1.1 Bound Optimization

We tackle binary energy functions $E(S)$ where $S = \{s_p \,|\, p \in \Omega\}$ is a vector of binary variables for pixels $p \in \Omega$. *Bound optimizers* iteratively minimize an *auxiliary function* bounding the original energy across the entire solution space.

Definition 1 (Auxiliary function). $A_t(S)$ *is an auxiliary function of $E(S)$ at current solution S_t if it satisfies the following conditions:*

$$E(S) \leq A_t(S), \forall S \tag{1a}$$
$$E(S_t) = A_t(S_t). \tag{1b}$$

Then, the current solution S_t is updated to the global optimum of the auxiliary function:

$$S_{t+1} = \arg\min_S A_t(S), \quad t = 1, 2, \ldots \tag{2}$$

Ideally optimization of the auxiliary function is easier than that of the original energy. Bound optimizers guarantee not to increase the original energy at each iteration since we have

$$E(S_{t+1}) \leq A_t(S_{t+1}) \leq A_t(S_t) = E(S_t). \tag{3}$$

Examples of well-known bound optimizers include mean-shift [9], difference of convex functions (DC) programming techniques [1], expectation maximization (EM) and submodular-supermodular procedures [27]. Besides, bound optimizers successfully tackled various problems in machine learning [39], computational statistics [22] and nonnegative matrix factorization [23].

In vision, bound optimizers were recently used for high-order or non-submodular pairwise energies [3,29,11,31]. The recent Auxiliary Cuts [3] work derived bounds for certain class of high-order functions. A variant of Auxiliary Cuts (LSA-AUX) is proposed in [11] for quadratic pseudo-boolean optimization.
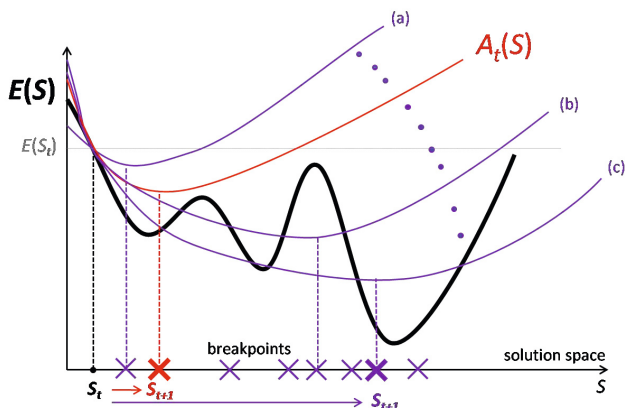
Fig. 1. Illustration of how pseudo-bound optimization framework updates current solution from S_t to S_{t+1}. Instead of optimizing only one auxiliary function $A_t(S)$, we explore a parametric family of pseudo-bounds. The best solution with the minimum original energy is chosen among the set of all global minima for the family. They correspond to breakpoints of parametric maxflow method [20] that can be efficiently explored in polynomial time. By optimizing the family of pseudo-bounds, larger decrease in energy is achieved ($S_t \rightarrow S_{t+1}$). [Best viewed in color]

1.2   Motivation and Contributions

Typically, for bound optimizers, one auxiliary function is chosen and optimized at each iteration. Furthermore, such an auxiliary function has to be an upper bound for the original energy $E(S)$ across the entire solution space, see $A_t(S)$ in Fig.1. However, in practice, it is difficult to find bounds that approximate well the original energy while being amenable to fast global solvers. Although working well for some applications, *auxiliary cuts* [3,11] may converge to undesirable solutions for several types of functions, see a representative example in Fig. 2.

   Our main idea is to relax the bound condition for an auxiliary function replacing it with pseudo-bounds, which may better approximate the original energy. Consider the example in Fig.1. Auxiliary function $A_t(S)$ does guarantee that its global minimum decreases the original energy $E(S)$, see Sec.1.1. However, there are many other approximation functions whose global minimum also decrease the original energy. For example, optimal solutions for (a) and (c) decrease $E(S)$ because these functions are local upper bounds for $E(S)$ around their global minima. Function (b) does not have this local bound property, but its global minimum still decreases the original energy. Moreover, solutions obtained by minimizing other approximation functions, e.g. (c), could be better than the one from the upper bound, *i.e.* auxiliary function $A_t(S)$. In that sense, relaxing the upper bound constraint allows better approximations of $E(S)$.

   We want to design an optimization algorithm using a larger class of relaxed bounds, which could give better solutions than proper auxiliary functions. The key challenge is choosing such *pseudo-bounds* so as to guarantee the original

energy decrease; note that the upper bound constraint was used when proving (3). One way to proceed could be to design some specific relaxed bounds that guarantee the decrease for $E(S)$ by construction. For example, in some applications it might be possible to design a particular approximation function that is guaranteed to locally dominate the original energy only around its global optimum, as in Fig.1 (a) or (c), which is sufficient to prove the decrease of $E(S)$.

This paper follows an alternative approximation approach. Instead of a single auxiliary function, at each iteration we optimize a family of *pseudo-bounds* that includes only one proper bound, while the bound constraint is relaxed for the other functions. As shown in Sec.2.1, inclusion of one proper bound is sufficient to guarantee the original energy decrease when the best solution is selected among global minima for the whole family. As illustrated in Fig.1 and confirmed by our practical experiments, relaxation of the bound constraint allows to significantly improve the quality of optimization compared to auxiliary functions, even when pseudo-bounds come from the same class of globally optimizable functionals.

A parametric family of pseudo-bounds is built as follows. We start from a known optimizable, *i.e.* submodular, auxiliary function and add a unary *bound relaxation* term weighted by a parameter. In order to explore all global minima for the whole parametric family efficiently, we propose parametric maxflow [20,14,10], reviewed in Sec. 2.2. To find all global minima for the whole family in polynomial time, the unary bound relaxation term must be monotone w.r.t. parameter. This practical consideration is important when selecting parametric pseudo-bound families for specific applications, e.g (12), (16), or (19). Note that parametric maxflow can be easily parallelized to further accelerate our algorithm.

Our contributions can be summarized as follows.

- This paper proposes a new general *pseudo-bound optimization* paradigm for approximate iterative minimization of high-order and non-submodular binary energies. It is a generalization of the standard *majorize-minimize* principle relaxing the bound constraint for an auxiliary function.
- We optimize a parametric family of pseudo-bounds at each iteration. To guarantee the energy decreases we include one proper bound in the family.
- In the context of discrete optimization, we propose parametric maxflow technique [20,14,10] to explore all global minima for the whole family in low-order polynomial time. To guarantee this complexity, we can choose families of pseudo-bounds with *monotone* dependence on parameter.
- We propose and discuss several examples of pseudo-bound families for different high-order and non-submodular pairwise energies.
- Our *parametric Pseudo-Bound Cuts* algorithm (pPBC) improves the-state-of-the-art in many energy minimization problems, e.g. entropy based image segmentation, target distributions matching, curvature regularization and image deconvolution. In particular, we outperform the standard GrabCut algorithm [32] both in terms of energy and segmentation error statistics. Our pseudo-bound approach is more robust to initialization and binning. Our pPBC algorithm also gives lower energy than *Auxiliary Cuts* [3] and *Fast Trust Region* [12] for distribution matching, see Fig. 2, and other challenging optimization problems in computer vision, see Section 4.
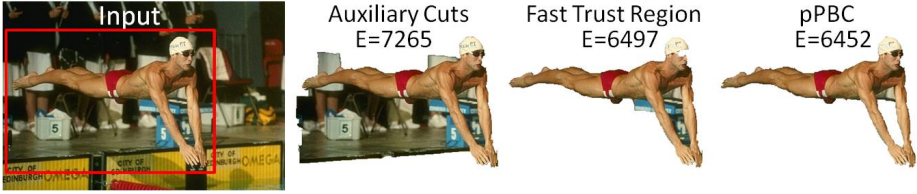
Fig. 2. Matching target foreground color distribution using auxiliary cuts [3], fast trust region [12] and pPBC. pPBC achieves the lowest energy.

2   Parametric Pseudo-bound Cuts (pPBC)

2.1   Our Pseudo-bound Framework

First, we define a family of *pseudo-bounds* for a scalar parameter λ with values in some set $\Lambda \subseteq \mathbb{R}$, for example $\Lambda = [\lambda^{min}, \lambda^{max}]$.

Definition 2 (Pseudo-Bound). *Assume energy $E(S)$, some current solution $S_t \in \{0,1\}^\Omega$ and parameter $\lambda \in \Lambda$. Then, function $\mathcal{F}_t(S, \lambda) : \{0,1\}^\Omega \times \Lambda \to \mathbb{R}$ is called a* pseudo bound *for energy $E(S)$ if there exists $\lambda' \in \Lambda$ such that $\mathcal{F}_t(S, \lambda')$ is an auxiliary function for $E(S)$ at current solution S_t.*

We may informally refer to pseudo-bound function $\mathcal{F}_t(S, \lambda)$ as a family of pseudo-bounds or a parametric family.

Our goal is to iteratively update current solution S_t for energy $E(S)$. Instead of bound optimization discussed in Sec.1.1, we propose Algorithm 1 that computes new better solution S_{t+1} by optimizing pseudo-bound $\mathcal{F}_t(S, \lambda)$ as follows.

Proposition 1. *Assume energy $E(S)$, current solution S_t and a pseudo-bound family $\mathcal{F}_t(S, \lambda)$ over parameter $\lambda \in \Lambda$. Let S^λ denote an optimal solution for $\mathcal{F}_t(S, \lambda)$ at any particular λ:*

$$S^\lambda = \arg\min_S \mathcal{F}_t(S, \lambda). \tag{4}$$

Then, $\lambda^* = \arg\min_\lambda E(S^\lambda)$ gives solution $S_{t+1} := S^{\lambda^*}$ reducing original energy

$$E(S_{t+1}) = E(S^{\lambda^*}) \leq E(S_t).$$

Proof. Pseudo-bound family $\mathcal{F}_t(S, \lambda)$ contains an auxiliary function $\mathcal{F}_t(S, \lambda')$ for some λ'. Optimization over the whole family should give better solution than one particular function $E(S^{\lambda^*}) \leq E(S^{\lambda'})$. Then, the proposition follows from the property of auxiliary functions $E(S^{\lambda'}) \leq E(S_t)$, see (3).   □

We construct a pseudo-bound family at current solution S_t by augmenting some auxiliary function $A_t(S)$ with a weighted *bound relaxation* term $R_t(S)$:

$$\mathcal{F}_t(S, \lambda) = A_t(S) + \lambda\, R_t(S). \tag{5}$$

Algorithm 1. PARAMETRIC PSEUDO-BOUND CUTS (pPBC)

---

1 $S_0 \longleftarrow S_{init}$
2 For $t = 0, 1, 2, ...$, **repeat until convergence**
3   Construct an auxiliary function $A_t(S)$ at current solution S_t;
4   Combine $A_t(S)$ with unary relaxation term $R_t(S)$ to form pseudo-bound
$$\mathcal{F}_t(S, \lambda) = A_t(S) + \lambda\, R_t(S)$$
5   **//Optimize the parametric family of pseudo-bounds**
$$S^\lambda = \arg\min_S \mathcal{F}_t(S, \lambda), \text{ for } \lambda \in \Lambda$$
6   **//Score candidate solutions and update**
$$\lambda^* = \arg\min_\lambda E(S^\lambda), \qquad S_{t+1} \longleftarrow S^{\lambda^*}$$

Note that for $\lambda = 0$ our pseudo-bound $\mathcal{F}_t(S, \lambda)$ in (5) reduces to auxiliary function $A_t(S)$. Starting from the same current solution S_t, our pseudo-bound optimization is guaranteed to find at least as good or better solution than optimization of bound $A_t(S)$. While pseudo-bound may not be a proper bound for $\lambda \neq 0$, it may better approximate the original energy $E(S)$, see Fig.1.

In the context of binary energies $E(S)$ we typically choose some submodular $A_t(S)$ and modular (unary) $R_t(S)$. The resulting pseudo-bound family (5) is of the form (6) that allows to efficiently explore the whole set of solutions S^λ with standard *parametric maxflow* techniques reviewed in Sec. 2.2. The next solution $S_{t+1} = S^{\lambda^*}$ can be computed by selecting S^λ with the lowest value of original energy $E(S)$, as summarized in Alg.1.

2.2   Overview of *Parametric maxflow*

Parametric maxflow technique [20,14,10] is a building block in our proposed algorithm. For all λ in some interval $\Lambda = [\lambda^{min}, \lambda^{max}]$, parametric maxflow can efficiently generate a (finite) set of all distinct solutions $S^\lambda \in \{0, 1\}^\Omega$ minimizing energy $E(S, \lambda)$ of form

$$S^\lambda = \arg\min_S \overbrace{\sum_{p \in \Omega} (a_p + \lambda b_p)s_p + \sum_{(p,q) \in \mathcal{N}} \phi_{pq}(s_p, s_q)}^{E(S,\lambda)} \qquad (6)$$

where ϕ_{pq} are submodular pairwise terms for a set of pairwise factors $\mathcal{N}$. Note that the unary terms in (6) *linearly* depend on parameter λ.

As discussed in [7,20], interval Λ can be broken into a finite set of subintervals between *breakpoints* $\lambda_1 < \lambda_2 < ... < \lambda_k \in \Lambda$ such that any λ inside each given interval $[\lambda_i, \lambda_{i+1}]$ gives the same solution $S^\lambda = S^i$. Parametric maxflow identifies all breakpoints and solutions S^i by making a finite number of calls to the maxflow procedure, see [20,7] for details. Importantly, in *monotonic* case when coefficients b_p in (6) are either all non-negative or all non-positive, optimal solutions S^i have a *nestedness* property leading to guaranteed *polynomial* complexity. This necessitates our choice of relaxation term $R_t(S)$ to have unary coefficients of the same sign.

(a) Volumetric prior, Sec.3.1.1   (b) Bhattacharyya prior, Sec.3.1.3
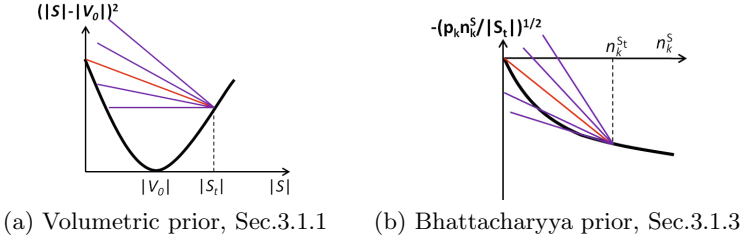
Fig. 3. Pseudo-bound families for two cardinality functions. Auxiliary functions are red.

3   Examples of Pseudo-bounds

Algorithm 1 for minimizing energy $E(S)$ depends on pseudo-bound (5) and requires specific choices of a submodular auxiliary function $A_t(S)$ and a unary relaxation term $R_t(S)$. This section provides practical pseudo-bound examples for a wide range of high-order and non-submodular pairwise energies $E(S)$.

3.1   High-Order Energies

3.1.1   Volumetric Potential [12,37] like $\psi(|S| - V_0)$ for convex symmetric function $\psi(\cdot)$ penalize deviation of segment size $|S|$ from target volume V_0. For example, if $\psi(x) = x^2$ and $S \subset S_t$ we can use the following pseudo bound family[1] illustrated in Fig.3(a)

$$\mathcal{F}_t(S, \lambda) = \underbrace{(|S_t| - 2V_0)|S| + V_0^2}_{A_t(S)} + \lambda \underbrace{(|S| - |S_t|)}_{R_t(S)}. \qquad (7)$$

3.1.2   Appearance Entropy was proposed for image segmentation in [35] as a general color consistency criterion that can be combined with other standard terms, e.g. boundary smoothness, as in the following binary segmentation energy

$$E(S) \;=\; |S| \cdot H(S) + |\bar{S}| \cdot H(\bar{S}) \;+\; |\partial S| \qquad (8)$$

where $H(S)$ and $H(\bar{S})$ are entropies of color histograms inside foreground S and background $\bar{S}$ and $|\partial S| = \sum_{\{p,q\} \in \mathcal{N}} \omega_{pq}|s_p - s_q|$ is segmentation boundary length. Indirectly, color entropy was also used for segmentation in [40,32,6]. Entropy can also be used as a clustering criterion for any image features that can be binned. In fact, entropy and related *information gain* criterion are widely used in learning, e.g. for contextual clustering [25] or decision trees [34].

---

[1] For $\psi(x) = x^2$ our volumetric potential is non-submodular pairwise, see also Sec.3.2.

Entropy-based energy (8) from [35] is related to a well-known *minimum description length* (MDL) functional [40,6] for color model fitting. In particular, for two segments it reduces to a color model fitting energy in GrabCut [32]

$$E(S, \theta^1, \theta^0) = -\sum_{p \in \Omega} \log Pr(I_p | \theta^{s_p}) + |\partial S| \tag{9}$$

where θ^1 and θ^0 are variables corresponding to unknown color models for foreground S and background $\bar{S}$. As shown in [35], globally optimal S for high-order energy (8) and mixed optimization functional (9) coincide if color models θ are represented by histograms. Since (9) is known to be *NP-hard* [36], it follows that high-order entropy energy in (8) is also *NP-hard*.

Equivalence of global solutions for entropy (8) and color-model fitting (9) suggests that (8) is minimized indirectly when applying standard *block-coordinate descent* (BCD) techniques [40,32,6] to energy (9) separately optimizing variables S and θ at each iteration. Below, we show that BCD in [32] can be seen as a bound optimization method for entropy (8). Then, we use the corresponding auxiliary function to build a family of pseudo-bounds that generate significantly better results, as shown by our experiments in Sec.4.1.

Proposition 2. *Assume fixed histograms θ_t^1 and θ_t^0 computed from the colors of current solution S_t (foreground) and its complement $\bar{S}_t$ (background). Then,*

$$A_t(S) := E(S, \theta_t^1, \theta_t^0), \tag{10}$$

with E as in (9), is an *auxiliary function for entropy-based energy (8) at S_t.*

Proof. It follows from a cross entropy discussion in [35]. Indeed, as easy to check

$$E(S, \theta_t^1, \theta_t^0) = |S| \cdot H(S|S_t) + |\bar{S}| \cdot H(\bar{S}|\bar{S}_t) + |\partial S| \tag{11}$$

where $H(\cdot|\cdot)$ is a cross-entropy of color distributions in two sets of pixels. Inequality $H(A|B) \geq H(A|A) = H(A)$ for $\forall A, B \subset \Omega$ implies

$$E(S, \theta_t^1, \theta_t^0) \geq E(S)$$

where $E(S)$ is from (8). It is also easy to check that $E(S_t, \theta_t^1, \theta_t^0) = E(S_t)$.   □

Corollary 1. *Block-coordinate descent (BCD) for mixed functional (9), as in GrabCut [32], is a bound optimization for entropy-based energy (8), see Sec.1.1.*

Proof. Two steps during each iteration of BCD in GrabCut are (I) optimization of segment S by applying graph cuts to energy (9) with fixed color models, as in Boykov-Jolly [5], and (II) optimization of color models θ^1, θ^2 in energy (9) with fixed segmentation. Prop. 2 implies that the segmentation step optimizes auxiliary function $A_t(S)$ for energy (8) at S_t and gives
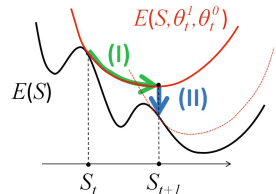
next solution S_{t+1}, as illustrated on the right. Color model re-estimation step gives new auxiliary function $A_{t+1}(S)$ at S_{t+1}. □

Our proposed pPBC method for entropy-based segmentation energy (8) augments the auxiliary function $A_t(S)$ in (10) with weighted bound relaxation term $\lambda(|S| - |S_t|)$ giving the following family of pseudo-bounds:

$$\mathcal{F}_t(S, \lambda) = E(S, \theta_t^1, \theta_t^0) + \lambda \left(|S| - |S_t|\right). \tag{12}$$

3.1.3   Matching Color Distributions [3,12,29,2] One way of matching target color distributions is to minimize Bhattacharyya measure:

$$Bha(S) = -\sum_k \sqrt{p_k n_k^S / |S|}, \tag{13}$$

where n_k^S is the number of foreground pixels in color bin k and $\sum_k p_k = 1$ is the target distribution. For $S \subset S_t$, a family of pseudo-bounds (Fig. 3) is given as:

$$\mathcal{F}_t(S, \lambda) = \underbrace{-\sum_k \sqrt{\frac{p_k}{n_k^{S_t} |S_t|}} n_k^S}_{A_t(S)} + \lambda \underbrace{\left(|S| - |S_t|\right)}_{R_t(S)} \tag{14}$$

Another option for matching distributions is to use the KL divergence [3]:

$$KL(S) = \sum_k p_k \log \frac{p_k}{n_k^S / |S| + \epsilon} = \sum_k p_k \log p_k - \sum_k p_k \log \left(\frac{n_k^S}{|S|} + \epsilon\right), \tag{15}$$

where ϵ is a small constant used to avoid numerical issue. In this case, for $S \subset S_t$, we have the following family of pseudo-bounds (omitting constant $\sum_k p_k \log p_k$):

$$\mathcal{F}_t(S, \lambda) = \underbrace{\sum_k \frac{p_k}{n_k^{S_t}} \left(\log \frac{\epsilon}{\frac{n_k^{S_t}}{|S_t|} + \epsilon}\right) n_k^S - \log \epsilon}_{A_t(S)} + \lambda \underbrace{\left(|S| - |S_t|\right)}_{R_t(S)}, \tag{16}$$

where $A_t(S)$ is the auxiliary function derived recently in [3].

3.2   Non-submodular Pairwise Energies

We consider a general class of binary pairwise non-submodular energies, which are useful in various vision applications [16,11], e.g., segmentation, stereo, inpainting, deconvolution, and many others. Such energies can be expressed as:

$$E(S) = \sum_{(p,q) \in \mathcal{N}} m_{pq} s_p s_q = S^T M S, \qquad S \in \{0, 1\}^\Omega \tag{17}$$

where $M = \{m_{pq} \in \mathbb{R} \,|\, p, q \in \Omega\}$ is a symmetric matrix containing pairwise potentials. if $m_{pq} \leq 0 \;\forall(p, q)$, energy (17) is *submodular* and, therefore, global
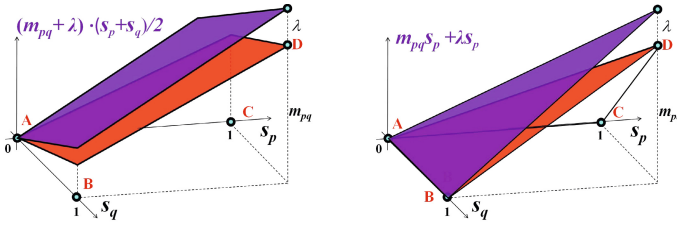
Fig. 4. pPBC-T: Pseudo-bounds (purple) and auxiliary functions (red) of non-submodular potential $m_{pq}s_ps_q$ for current configuration $s_{p,t} = 0, s_{q,t} = 0$ (left) and $s_{p,t} = 0, s_{q,t} = 1$ (right).

Table 1. Auxiliary functions [11] and weighted bound relaxation term for pPBC-T

$(s_{p,t}, s_{q,t})$	Auxiliary function	relaxation term (pPBC-T)
$(0,0)$	$m_{pq}(s_p + s_q)/2$	$\lambda(s_p - s_{p,t} + s_q - s_{q,t})$
$(0,1)$	$m_{pq}s_p$	$\lambda(s_p - s_{p,t})$
$(1,0)$	$m_{pq}s_q$	$\lambda(s_q - s_{q,t})$
$(1,1)$	$m_{pq}(s_p + s_q)/2$	$\lambda(s_p - s_{p,t} + s_q - s_{q,t})$

optima can be reached in a low-order polynomial time using graph cuts [4]. The general non-submodular case is *NP-hard*. In the following, we propose three different pseudo-bounds families for (17) for non-submodular pairs ($m_{pq} > 0$).

pPBC-T(touch) gives pseudo-bounds for each non-submodular potential $m_{pq}s_ps_q, m_{pq} > 0$. Depending on the current configuration $s_{p,t}$ and $s_{q,t}$ for s_p and s_q, we augment the bound recently proposed in [11] with the relaxation terms specified as in Table 1. Fig. 4 shows the auxiliary functions in red and pseudo-bounds in purple for current configuration $(0,0)$ and $(0,1)$. Note that the bound relaxation term for current configuration $(0,1)$ and $(1,0)$ is different from that of $(0,0)$ and $(1,1)$. This relaxation allows the pseudo-bounds to *touch* the original energy at as many points as possible, yielding better approximation.

pPBC-B(ballooning) This option uses the auxiliary function in Table 1 augmented with a linear ballooning term $\lambda (|S| - |S_t|)$.

pPBC-L(Laplacian) We derive the third pseudo-bounds family based on the *Laplacian* matrix. Let $d(p) = \sum_q m_{pq}$ and D be the diagonal matrix having d on its diagonal. Notice that, in the case of supermodular terms ($m_{pq} \geq 0$), D is diagonally positive and, therefore, positive semidefinite. With symmetric matrix M, it is well known that the corresponding *Laplacian* matrix $L = D - M$ is positive semidefinite [33]. Now we write (17) as follows for $\lambda \in \Lambda$:

$$E(S) = \underbrace{S^T(M - \lambda D)S}_{G(S)} + \lambda \underbrace{S^T D S}_{H(S)}. \tag{18}$$

H is a unary potential for binary variables: $H(S) = \sum_p d(p)s_p^2 = \sum_p d(p)s_p$. Also, notice that $\forall \lambda \geq 1$, G is concave w.r.t S because $M - \lambda D$ is negative semidefinite (as it is the sum of two negative semidefinite matrices:
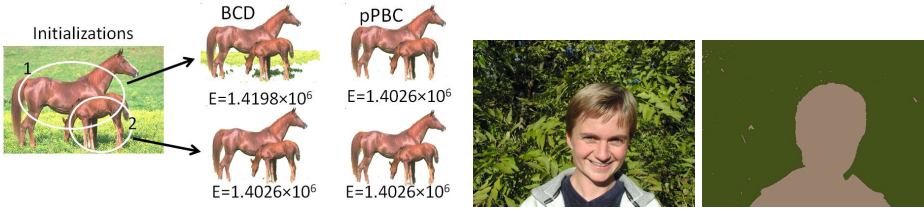
Fig. 5. Left: interactive segmentations with BCD (GrabCut) or pPBC from different initialization (ellipses). Proposed pPBC method is more robust to inferior initialization. Right: unsupervised figure-ground segmentation with pPBC. Average color is shown.

$M - \lambda D = -L + (1 - \lambda)D$). Therefore, let ∇ denotes the gradient, we have the following pseudo-bounds at current solution S_t which includes bounds of (17) for $\lambda \geq 1$.

$$\mathcal{F}_t(S, \lambda) = G(S_t) + \nabla G(S_t)^T (S - S_t) + \lambda H(S)$$
$$= \underbrace{G(S_t) - \nabla G(S_t)^T S_t}_{\text{Constant}} + \underbrace{2[(M - \lambda D)S_t]^T S + \lambda H(S)}_{\text{Unary potential}} \qquad (19)$$

4   Experiments

### 4.1   Appearance Entropy Based Segmentation

Robustness w.r.t Initialization and Binning. We use GrabCut and BCD interchangeably for the rest of the paper. Left part of Fig. 5 depicts an example of interactive segmentation with BCD or our proposed pPBC, and shows that BCD is sensitive to initializations, unlike pPBC. pPBC can even tolerate trivial initialization, see an un-supervised segmentation example in the right of Fig. 5.

Furthermore, we observed that with more appearance model variables, namely the number of color bins, BCD is more likely to get stuck in weak local minima. We randomly generated 500 box-like initializations for an input image, and run BCD and pPBC for different numbers of color bins, ranging from 16^3 to 128^3. From the solutions we obtained with BCD or pPBC, we computed the corresponding error rates and energies. Fig. 6 depicts the scatter plots of error rates
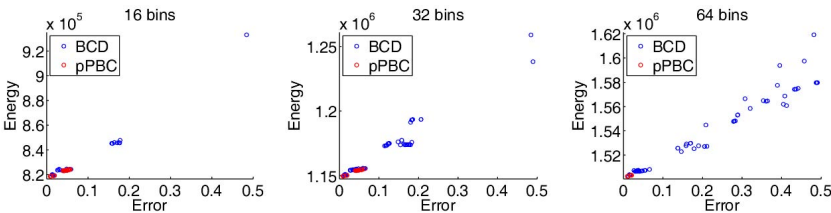
Fig. 6. Scatter plots; error rates versus energies for 500 solutions of BCD and pPBC

	Error rate	Time
GrabCut (16^3 bins)	$7.1\%^2$	1.78 s
GrabCut (32^3 bins)	8.78%	1.63 s
GrabCut (64^3 bins)	9.31%	1.64 s
GrabCut (128^3 bins)	11.34%	1.45 s
DD (16^3 bins)	10.5%	576 s
One-Cut (16^3 bins)	8.1%	5.8 s
One-Cut (32^3 bins)	6.99%	2.4 s
One-Cut (64^3 bins)	6.67%	1.3 s
One-Cut (128^3 bins)	6.71%	0.8 s
pPBC (16^3 bins)	5.80%	11.7 s
pPBC (32^3 bins)	5.60%	11.9 s
pPBC (64^3 bins)	5.56%	12.3 s
pPBC (128^3 bins)	7.51%	15.9 s
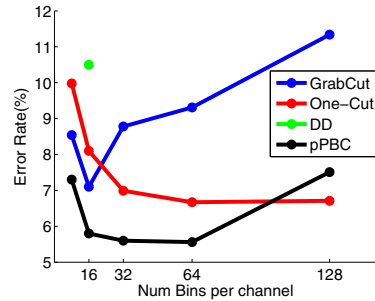
Fig. 7. Error rates and speed on GrabCut dataset for GrabCut [32], Dual Decomposition (DD) [36], One-Cut [35] and our pPBC method.

versus energies for the 500 solutions. Points on bottom-left give low energy and small error rate. The wider these dots spread across the plane, the more local minima the algorithm converged to. pPBC works much better than BCD for finer binning and is more robust to initializations.

Comparisons with the State of the Art [35,36] We compare with Grab-Cut, which as demonstrated in Sec.3.1.2, can be viewed as a bound optimizer. We run both algorithms on the GrabCut dataset [32] (The cross image excluded for comparison with [36]). We set the weight of the 8-connected contrast-sensitive smoothness term to 15 and vary number of color bins. As shown in Tab.3, pPBC consistently gives lower energies and misclassification errors. Our current implementation does not use a straightforward multi-core CPU parallelization of parametric maxflow by breaking the range of λ into intervals. Thus, significant speed up of our pPBC algorithm is possible. In the next experiment we tuned the smoothness term weight for pPBC and other methods [32,35] to obtain the best error statistics for each. Fig. 7 shows a competitive performance of pPBC.

Table 3. Statistics of pPBC and GrabCut [32] over the GrabCut database

	Mean Energy	# of lower energy	Mean time(s)
GrabCut [32] - 16^3 bins	1.2349×10^6	1	**1.0s**
pPBC - 16^3 bins	$\mathbf{1.2335 \times 10^6}$	**38**	11.2s
GrabCut [32] - 32^3 bins	1.7064×10^6	2	**0.9s**
pPBC - 32^3 bins	$\mathbf{1.7029 \times 10^6}$	**37**	11.7s
GrabCut [32] - 64^3 bins	2.2408×10^6	1	**0.9s**
pPBC - 64^3 bins	$\mathbf{2.2361 \times 10^6}$	**47**	14.1s

Table 4. Matching color distribution (KL or Bhattacharyya distance) with Auxiliary Cuts [3], FTR [12], pPBC and its limited version with $\lambda \leq 0$ for the pseudo-bounds.

Method	KL divergence (15)			Bhattacharyya distance (13)		
	Mean energy	Mean error	Time	Mean energy	Mean error	Time
Auxiliary Cuts [3]	6189	16.54%	**1.8s**	-12402	24.1%	**1.7s**
pPBC($\lambda \leq 0$)	6150	14.88%	N/A	-12451	23.7%	N/A
FTR [12]	5868	7.70%	4.40s	-14499	3.2%	2.71s
pPBC($\lambda \in [-\infty, +\infty]$)	**5849**	**3.63%**	2.98s	**-14504**	**2.9%**	1.99s

4.2   Matching Color Distributions

We experiment on the database [32], and used the bounding boxes as initializations. Similar to [3,12], the target distribution is learned from the ground truth. We compared pPBC with auxiliary cuts [3] and FTR [12]. We also tested a limited version of pPBC where only non-positive λ's were explored within the family of pseud-bounds. Note that, when λ is non-positive, the parametric family includes only auxiliary functions. The mean error rate, energy and running time are reported in Table 4. Exploring only a family of auxiliary functions ($\lambda \leq 0$) did not improve the results. pPBC with parameter $\lambda \in \mathbb{R}$ yielded the best performance, while being slightly slower than auxiliary cuts (even though pPBC explores a family of functions instead of only one). FTR yielded comparable mean energy to pPBC, but is slower. Fig. 2 depicts typical examples.

4.3   Curvature Regularization

We applied our framework to the curvature model proposed in [8], which penalizes 90 degree angles in a 4-connect neighborhood system. We also compare pPBC to the recent algorithms (LSA-AUX and LSA-TR) in [11], which were shown to outperform standard state-of-the-art methods such as QPBO [30] and TRWS [19]. Fig. 8 plots the energies of the solutions with different weights of the curvature term. pPBC-T gives the lowest energy among all methods. We also observed that the best λ for pPBC-T often does not make the pseudo-bound an auxiliary function, which means the bounding constraint is violated.

Table 5. Average energy with 10 random noisy images

Noise σ	LSA-AUX [11]	LSA-TR [11]	pPBC-L	pPBC-B	pPBC-T
0.05	40.34	30.83	39.49	39.65	**30.81**
0.10	130.24	**119.84**	128.68	128.06	121.20
0.15	277.27	**263.06**	275.89	276.62	266.35
0.20	482.54	**451.78**	480.80	482.09	471.11
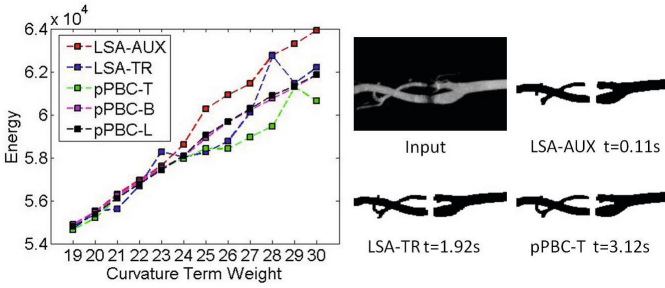
Fig. 8. Segmentation with the curvature regularization model in [8]

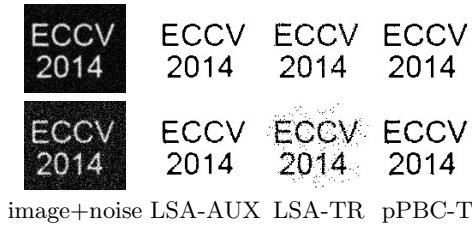image+noise LSA-AUX LSA-TR pPBC-T

Fig. 9. Deconvolution results. Top row: noise $\sigma = 0.10$, bottom row: $\sigma = 0.2$.

4.4   Deconvolution

Fig. 9 depicts a binary image convolved with a mean 3×3 filter, with a Gaussian noise added. The purpose is to recover the original image via optimizing the energy: $E(S) = \sum_{p \in \Omega} (I_p - \frac{1}{9} \sum_{q \in \mathcal{N}_p} s_q)^2$, where $\mathcal{N}_p$ is a 3×3 neighborhood window centered at pixel p. In this energy, all pairwise interactions are supermodular. We compared our pPBC-B, L or T to the recent algorithms in [11] (LSA-AUX and LSA-TR). Table 5 shows average energy of those methods. Note that LSA-TR achieves lower energy but visually worse deconvolution. For $\sigma = 0.05$ noise, LSA-AUX takes 0.12s, LSA-TR 0.73s and pPBC-T 1.46s.

5   Conclusion

This paper proposes a new general *pseudo-bound optimization* paradigm for approximate iterative minimization of high-order and non-submodular binary energies. It generalizes the standard *majorize-minimize* principle relaxing the bound constraint for an auxiliary function. We propose to optimize a family of pseudo-bounds at each iteration. To guarantee the energy decreases we include at least one bound in the family. We propose parametric maxflow [20,14,10] to explore all global minima for the whole family in low-order polynomial time.

To guarantee polynomial time complexity, pseudo-bounds families with *monotone* dependence on parameter are chosen. We propose and discuss several options of pseudo-bound families for various high-order and non-submodular

pairwise energies. Our *parametric Pseudo-Bound Cuts* algorithm (pPBC) improves the-state-of-the-art in many energy minimization problems, e.g. entropy based segmentation, target distributions matching, curvature regularization and deconvolution. In particular, we show that the well-known GrabCut algorithm [32] is a bound optimizer. Our pseudo-bound approach is more robust to inferior initialization and finer binning for image segmentation. Our pPBC algorithm also gives lower energy than *Auxiliary Cuts* [3] and *Fast Trust Region* [12] for distribution matching and other challenging optimization problems in vision.

References

1. An, L.T.H., Tao, P.D., Canh, N.N., Thoai, N.V.: Dc programming techniques for solving a class of nonlinear bilevel programs. J. Global Optimization 44(3), 313–337 (2009)
2. Ben Ayed, I., Chen, H.M., Punithakumar, K., Ross, I., Li, S.: Graph cut segmentation with a global constraint: Recovering region distribution via a bound of the bhattacharyya measure. In: CVPR, pp. 3288–3295 (2010)
3. Ben Ayed, I., Gorelick, L., Boykov, Y.: Auxiliary cuts for general classes of higher order functionals. In: IEEE Conference on Computer Vision and Pattern Recognition (2013)
4. Boros, E., Hammer, P.L.: Pseudo-boolean optimization. Discrete Applied Mathematics 123, 2002 (2001)
5. Boykov, Y., Jolly, M.-P.: *Interactive graph cuts* for optimal boundary & region segmentation of objects in N-D images. In: ICCV, vol. I, pp. 105–112 (July 2001)
6. Delong, A., Osokin, A., Isack, H., Boykov, Y.: Fast approximate energy minimization with label costs. International Journal of Computer Vision 96, 1–27 (2012)
7. Eisner, M.J., Severance, D.G.: Mathematical techniques for efficient record segmentation in large shared databases. J. ACM 23(4), 619–635 (1976)
8. El-Zehiry, N., Grady, L.: Fast global optimization of curvature. In: Proc. of CVPR 2010. IEEE Computer Society. IEEE (June 2010)
9. Fashing, M., Tomasi, C.: Mean shift is a bound optimization. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 471–474 (2005)
10. Gallo, G., Grigoriadis, M.D., Tarjan, R.E.: A fast parametric maximum flow algorithm and applications. SIAM J. Comput. 18(1), 30–55 (1989)
11. Gorelick, L., Boykov, Y., Veksler, O., Ayed, I.B., Delong, A.: Submodularization for binary pairwise energies. In: IEEE conference on Computer Vision and Pattern Recognition (CVPR), Columbus, Ohio (June 2014)
12. Gorelick, L., Schmidt, F.R., Boykov, Y.: Fast trust region for segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, Oregon, pp. 1714–1721(June 2013)
13. Gould, S.: Max-margin learning for lower linear envelope potentials in binary markov random fields. In: ICML (2011)
14. Hochbaum, D.S.: Polynomial time algorithms for ratio regions and a variant of normalized cut. IEEE Trans. Pattern Anal. Mach. Intell. 32(5), 889–898 (2010)
15. Jiang, H.: Linear solution to scale invariant global figure ground separation. In: CVPR. pp. 678–685 (2012)
16. Kappes, J.H., Andres, B., Hamprecht, F.A., Schnorr, C., Nowozin, S., Batra, D., Kim, S., Kausler, B.X., Lellmann, J., Komodakis, N., Rother, C.: A comparative

study of modern inference techniques for discrete energy minimization problem. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1328–1335 (2013)

17. Kim, J., Kolmogorov, V., Zabih, R.: Visual correspondence using energy minimization and mutual information. In: Int. Conf. on Comp. Vision (ICCV), vol. 2, pp. 1033–1040 (October 2003)

18. Kohli, P., Ladický, L., Torr, P.H.: Robust higher order potentials for enforcing label consistency. Int. J. Comput. Vision 82(3), 302–324 (2009)

19. Kolmogorov, V., Schoenemann, T.: Generalized sequential tree-reweighted message pass. In: arXiv:1205.6352 (2012)

20. Kolmogorov, V., Boykov, Y., Rother, C.: Applications of parametric maxflow in computer vision. In: IEEE International Conference on Computer Vision (2007)

21. Kolmogorov, V., Rother, C.: Minimizing non-submodular functions with graph cuts - a review. IEEE transactions on Pattern Analysis and Machine Intelligence 29(7), 1274–1279 (July 2007)

22. Lange, K., Hunter, D.R., Yang, I.: Optimization transfer using surrogate objective functions. Journal of Computational and Graphical Statistics 9(1), 1–20 (2000)

23. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: NIPS, pp. 556–562 (2000)

24. Leordeanu, M., Hebert, M., Sukthankar, R.: An integer projected fixed point method for graph matching and map inference. In: NIPS, pp. 1114–1122 (2009)

25. Li, T., Ma, S., Ogihara, M.: Entropy-based criterion in categorical clustering. In: Proc. of Intl. Conf. on Machine Learning (ICML), pp. 536–543 (2004)

26. Mukherjee, L., Singh, V., Dyer, C.R.: Half-integrality based algorithms for cosegmentation of images. In: CVPR. pp. 2028–2035 (2009)

27. Narasimhan, M., Bilmes, J.A.: A submodular-supermodular procedure with applications to discriminative structure learning. In: UAI, pp. 404–412 (2005)

28. Nieuwenhuis, C., Toeppe, E., Gorelick, L., Veksler, O., Boykov, Y.: Efficient squared curvature. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, Ohio (June 2014)

29. Pham, V.Q., Takahashi, K., Naemura, T.: Foreground-background segmentation using iterated distribution matching. In: CVPR. pp. 2113–2120 (2011)

30. Rother, C., Kolmogorov, V., Lempitsky, V., Szummer, M.: Optimizing binary mrfs via extended roof duality. In: IEEE CVPR (2007)

31. Rother, C., Minka, T.P., Blake, A., Kolmogorov, V.: Cosegmentation of image pairs by histogram matching - incorporating a global constraint into mrfs. In: CVPR. pp. 993–1000 (2006)

32. Rother, C., Kolmogorov, V., Blake, A.: Grabcut - interactive foreground extraction using iterated graph cuts. In: ACM Tran. on Graphics, SIGGRAPH (2004)

33. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 22(8), 888–905 (2000)

34. Shotton, J., Sharp, T., Kohli, P., Nowozin, S., Winn, J., Criminisi, A.: Decision jungles: Compact and rich models for classification. In: Advances in Neural Information Processing Systems (NIPS), pp. 234–242 (2013)

35. Tang, M., Gorelick, L., Veksler, O., Boykov, Y.: Grabcut in one cut. In: International Conference on Computer Vision (ICCV), Sydney, Australia (2013)

36. Vicente, S., Kolmogorov, V., Rother, C.: Joint optimization of segmentation and appearance models. In: IEEE ICCV (2009)

37. Woodford, O.J., Rother, C., Kolmogorov, V.: A global perspective on map inference for low-level vision. In: ICCV, pp. 2319–2326 (2009)
38. Yuan, Y.: A review of trust region algorithms for optimization. In: Proceedings of the Fourth International Congress on Industrial and Applied Mathematics, ICIAM (1999)
39. Zhang, Z., Kwok, J.T., Yeung, D.Y.: Surrogate maximization/minimization algorithms and extensions. Machine Learning 69, 1–33 (2007)
40. Zhu, S.C., Yuille, A.: Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. IEEE Transactions on PAMI 18(9), 884–900 (1996)