Auxiliary Cuts for General Classes of Higher Order Functionals

Ismail Ben Ayed [1,2]
ismail.benayed@ge.com

Lena Gorelick [2]
lenagorelick@gmail.com

Yuri Boykov [2]
yuri@csd.uwo.ca

[1] GE Healthcare
London, ON, Canada

[2] University of Western Ontario
London, ON, Canada

Abstract

Several recent studies demonstrated that higher order (non-linear) functionals can yield outstanding performances in the contexts of segmentation, co-segmentation and tracking. In general, higher order functionals result in difficult problems that are not amenable to standard optimizers, and most of the existing works investigated particular forms of such functionals. In this study, we derive general bounds for a broad class of higher order functionals. By introducing auxiliary variables and invoking the Jensen's inequality as well as some convexity arguments, we prove that these bounds are auxiliary functionals for various non-linear terms, which include but are not limited to several affinity measures on the distributions or moments of segment appearance and shape, as well as soft constraints on segment volume. From these general-form bounds, we state various non-linear problems as the optimization of auxiliary functionals by graph cuts. The proposed bound optimizers are derivative-free, and consistently yield very steep functional decreases, thereby converging within a few graph cuts. We report several experiments on color and medical data, along with quantitative comparisons to state-of-the-art methods. The results demonstrate competitive performances of the proposed algorithms in regard to accuracy and convergence speed, and confirm their potential in various vision and medical applications.

1. Introduction

A large class of segmentation, co-segmentation and tracking problems in computer vision and medical imaging seek image-domain segments that optimize functionals of the following general form:

$$E(S) = R(S) + Q(S), \tag{1}$$

where term $R(S)$ describes some regional properties of the segments and $Q(S)$ is a regularization term, which enforces segment-boundary smoothness. Unfortunately, most of the existing powerful and global optimizers, e.g., discrete graph cuts [5] or continuous convex-relaxation techniques [6], are restricted to special linear (or unary) forms of regional functionals, where R can be expressed as the sum (or integral in the continuous setting) of individual-pixel penalties [4, 21, 16].

1.1. Linear functionals

Let $I_p = I(p) : \Omega \subset \mathbb{R}^2 \rightarrow \mathcal{Z} \subset \mathbb{R}^n, n \in \mathbb{N}^*$, be an image-feature function defined over a domain Ω. $\mathcal{Z}$ is a space of a feature variable such as intensity, color or texture. Linear functionals can be expressed in the following general form [4, 21, 16]:

$$\sum_{p \in S} g(p) = \sum_{\Omega} g(p) \chi_S(p) =: \langle g, S \rangle \tag{2}$$

where χ_S is the characteristic function of segment $S \subset \Omega$:

$$\chi_S(p) = \begin{cases} 1 & \text{if } \; p \in S \\ 0 & \text{if } \; p \in \Omega \setminus S, \end{cases} \tag{3}$$

and $g : \Omega \rightarrow \mathbb{R}$ is an arbitrary scalar function. The log-likelihood energy commonly used in image segmentation [4, 21, 16] corresponds to $g(p) = \log \frac{M_f(I_p)}{M_b(I_p)}$, where M_f and M_b are fixed foreground/background models.

Linear functionals are frequently used in vision and medical imaging because they are easily amenable to powerful global optimizers. However, as discussed and demonstrated in several recent studies [9, 20], such linear terms can impose only a limited set of characteristics on the solutions. Several recent research efforts have demonstrated that higher order (non-linear) functionals can enforce much more powerful constraints in various vision [8, 18, 20, 1, 11] and medical-imaging applications [2, 12].

1.2. Prior art on non-linear functionals

Useful non-linear functionals include but are not limited to several affinity measures on the distributions or moments of segment appearance [8, 20, 1, 11, 18] and shape

[12, 2], as well as soft constraints on segment volume [8, 24]. For instance, distribution (or histogram) matching formulations, which enforce some consistency between the feature distribution of the target segment and a given model, have recently sparked a significant research effort [8, 11, 20, 1, 18]. These studies proved that, in the context of interactive segmentation [20, 1], tracking [11] and co-segmentation of image pairs [18], optimizing some non-linear measures of affinity between distributions/histograms (e.g, the Kullback-Leibler divergence, Bhattacharyya coefficient or $\mathcal{L}_j$*-distances*) can yield outstanding performances unattainable with standard linear terms, and can be very useful in image retrieval [22].

In general, non-linear functionals result in difficult optimization problems that are not amenable to standard global optimization techniques. Some notable prior-art studies investigated specialized optimizers for particular forms of non-linear functionals. For instance, Rother et *al.* [22] pioneered optimization of the $\mathcal{L}_1$*-distance* between histograms via graph cuts. They combined some approximation of the functional with a heuristic linear term to damp the Newton's steps. As discussed in [9], their steps could be in the direction opposite to the gradient descent. Consequently, [22] may converge to solutions far from the local minima and, therefore, it is very sensitive to initializations [23]. For better initialization, they used a submodular-supermodular procedure [19] that assumes the functional is supermodular, and proved that this holds for the $\mathcal{L}_1$*-distance*. In [18], Mukherjee et al. suggested to replace the $\mathcal{L}_1$ by the $\mathcal{L}_2$*-distance*, arguing that the latter affords some interesting combinatorial properties, which befit graph cut optimization. After linearization of the $\mathcal{L}_2$*-distance*, the problem is rewritten as the optimization of a quadratic pseudo-boolean function (QPBF). Although QPBFs are non-submodular, the ensuing optimization problems allow roof-duality relaxation that can be solved by graph cuts [13]. Such relaxation yields, however, only a partial solution with some pixels left unlabeled. Furthermore, the algorithm in [18] is based on specific properties of the $\mathcal{L}_2$*-distance* and, therefore, cannot be applied to other non-linear functionals. In [1], the authors built a bound optimizer specific for the Bhattacharyya measure, which yielded very rapid convergences and competitive accuracies/optima. Unfortunately, the bound derivation in [1] relies on very specific properties of the Bhattacharyya coefficient, which precludes its use for any of the other important non-linear functionals.

Active-curve/level-set techniques can address arbitrary differentiable functionals [17]. By computing a curve evolution equation via a local gradient descent and the Gateâux derivative, the ensuing solutions may correspond to weak local minima [1]. Furthermore, in many instances of non-linear functionals, curve evolution yields computationally expensive algorithms [1, 17]. For example, the curve flow optimizing the Bhattacharyya measure is incremental [1, 17], and requires a large number of iterative updates of computationally costly distributions. As is the case of many continuous-optimization approaches, the robustness of curve-evolution algorithms inherently relies on the choice an approximating numerical solution that is controlled by several crucial parameters (e.g., step size).

In the recent study in [8], Gorelick et *al.* proposed an iterative trust-region method, which can address a fairly general class of differentiable functionals. At each iteration, they used graph cuts to optimize some second-order approximation of the functional within adaptively selected step size. This method can make larger moves than standard curve evolution techniques by finding an optimal combination of gradient descent and Newton's steps. In some practical examples, the adaptive schemes in [8] may use a relatively large number of iterations/cuts to converge. In the case of distribution (or histogram) based functionals, each iteration requires a distribution update that can be computationally expensive, more so when the number of bins is high. Therefore, a large number of iterations/cuts may lead a heavy computational burden. It is also worth noting that the method in [8] is not applicable to non-differentiable functionals, e.g., the $\mathcal{L}_1$*-distance* between histograms [22].

1.3. Bound optimization and auxiliary functionals

Let $E(u)$ be a functional to be optimized over some variable function u. *Bound-optimization* algorithms proceed by constructing and optimizing upper bounds of E, which serve as *auxiliary functionals* whose optimization is easier than the original functional:

Definition 1. *Given an auxiliary variable u^i, $A(u, u^i)$ is an auxiliary functional of E if it satisfies:*

$$E(u) \leq A(u, u^i) \tag{4a}$$

$$E(u) = A(u, u) \tag{4b}$$

Rather than optimizing directly E, we optimize iteratively a sequence of auxiliary functionals:

$$u^{i+1} = \min_u A(u, u^i), \quad i = 1, 2, \ldots \tag{5}$$

Using the conditions in (4a) and (4b), and by definition of minimum in (5), we can show that the solutions in (5) yield a decreasing sequence of E: $E(u^i) = A(u^i, u^i) \geq A(u^{i+1}, u^i) \geq E(u^{i+1})$. Therefore, if E is lower bounded, sequence $E(u^i)$ converges to a minimum of E.

Bound optimizers are *derivative- and parameter-free* algorithms (They do not require the functional to be differentiable, neither do they use optimizer parameters such as step sizes), and enjoy a strong guarantee: they never worsen the functional. They can be very efficient because they turn difficult and, in some instances, very complex optimization

problems into easier ones [25]. Bound optimization has yielded efficient solutions for difficult problems in Nonnegative Matrix Factorization [15] and computational statistics [14], and is gaining interest in machine learning [25]. For instance, the study [25] demonstrated the power of bound optimization in solving AdaBoost and logistic regression models.

The key difficulty in bound optimization is in constructing an appropriate auxiliary functional. On the one hand, the bound should be close enough to the original functional. On the other hand, a good auxiliary functional should be amenable to fast and global solvers.

1.4. Contributions of this study

In this study, we derive general bounds for a broad class of higher order functionals, which can be expressed as non-linear combinations of linear terms and their ratios. By introducing auxiliary variables and invoking the Jensen's inequality as well as some convexity arguments, we prove that these bounds are auxiliary functionals for various non-linear terms, which include but are not limited to several affinity measures on the distributions or moments of segment appearance and shape, as well as soft constraints on segment volume. From these general-form bounds, we state various non-linear problems as the optimization of auxiliary functionals by graph cuts. The proposed bound optimizers are derivative-free, and consistently yield very steep functional decreases, thereby converging within a few graph cuts (typically less than 5). We report several experimental evaluations on color and medical data, along with quantitative comparisons of the proposed bound optimizers to the methods in [8, 22]. The results demonstrate competitive performances of the proposed algorithms in regard to accuracy and convergence speed, and confirm their potential in various vision and medical applications.

2. Non-linear functions of linear terms

In this section, we focus on functionals $E = R + Q$, where R is a non-linear term of the following general form:

$$R(S) = \sum_{z \in Z} F_z(\langle g_z, S \rangle) \tag{6}$$

where $g_z : \Omega \rightarrow \mathbb{R}^+$, $z \in Z$, is a family of non-negative scalar functions defined over the image domain, and $F_z : \mathbb{R} \rightarrow \mathbb{R}^+$, $z \in Z$, is a family of non-negative scalar functions. The following lists some examples of very useful non-linear regional functionals of the form (6).

2.1. The $\mathcal{L}_j$-*distance* between histograms

This type of terms enforces an $\mathcal{L}_j$-*distance* consistency between the image histogram of the target segment S and a given model histogram $\{h_z, z \in Z\}$, and is very useful in co-segmentation [22, 18, 23] and tracking [11]:

$$\sum_{z \in Z} \left| h_z - \sum_{p \in S} \delta(I_p - z) \right|^j , j \in \mathbb{N}^* \tag{7}$$

where δ is given by:

$$\delta(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases} \tag{8}$$

Summation $\sum_{p \in S} \delta(I_p - z)$ counts the number of segment pixels which fall within bin z. The $\mathcal{L}_j$-*distance* in (7) belongs to the general from in (6), with the following particular functions: $g_z(p) = \delta(I_p - z)$, $z \in Z$, and $F_z(t) = |t - h_z|^j$, $j \in \mathbb{N}^*$, $z \in Z$.

2.2. The volume/area penalty function

The volume/area function penalizes a segment S when its size deviates from a given volume/area v_1 [8]:

$$\left(v_1 - \sum_{p \in S} 1 \right)^2 = (v_1 - \langle 1, S \rangle)^2 \tag{9}$$

The volume penalty can be written in the form of (6), with the following particular functions: $g_z = 1$, $z \in \{1\}$, and $F_z(t) = (v_z - t)^2$, $z \in \{1\}$. Note that other interesting shape-moment penalties can also be written in the form of (6) [8].

2.3. A general auxiliary functional

In the following, we will derive a family of general bounds (auxiliary functionals) for non-linear terms of the form in (6) by introducing auxiliary variables and invoking the Jensen's inequality as well as convexity arguments.

Proposition 1. *If F_z is convex $\forall z \in Z$, and given an auxiliary segment S^i, the following functional is an upper bound of the general non-linear form $R(S)$ in (6), $\forall \alpha \in \mathbb{R}^+$ and for any segment S verifying $S \subset S^i$:*

$$A_\alpha(S, S^i) = (1+\alpha) R(S^i) - \alpha \left\langle b(S^i), S \right\rangle + \left\langle \sum_{z \in Z} a_z^i, S^- \right\rangle \tag{10}$$

with $S^- = S^i \setminus S$, and functions a_z^i and b given by:

$$a_z^i(p) = \frac{g_z(p) \left[F_z(0) - F_z\left(\langle g_z, S^i \rangle \right) \right]}{\langle g_z, S^i \rangle} \tag{11a}$$

$$b(S^i) = \frac{R(S^i)}{\langle 1, S^i \rangle} \tag{11b}$$

Proof. Let us first recall the Jensen's inequality:

Lemma 1. ***The Jensen's inequality****: Let $F : D \subset \mathbb{R} \to \mathbb{R}$ be a convex function defined over a convex domain D, and P a finite set. If $x_p \in D \quad \forall p \in P$, and $\alpha_p \geq 0 \quad \forall p \in P$ with $\sum_{p\in P} \alpha_p = 1$, then we have [7]:*

$$F\left(\sum_{p\in P} \alpha_p x_p\right) \leq \sum_p \alpha_p F(x_p) \tag{12}$$

Now, for $S \subset S^i$, let us re-write $\langle g_z, S\rangle$ as follows:

$$\langle g_z, S\rangle = \sum_{p\in S} g_z(p)\chi_S(p) = \sum_{p\in S^i} g_z(p)\chi_S(p) \tag{13}$$

The second equality in (13) is due to the fact that, when $S \subset S^i$, we have $\chi_S(p) = 0 \quad \forall p \in S^i \setminus S$. Using (13), we can further re-write $F_z(\langle g_z, S\rangle)$ in the following form:

$$\begin{aligned} F_z\left(\langle g_z, S\rangle\right) =& F_z\left(\sum_{p\in S^i} g_z(p)\chi_S(p)\right) \\ =& F_z\left(\sum_{p\in S^i} \underbrace{\frac{g_z(p)}{\langle g_z, S^i\rangle}}_{\alpha_p} \underbrace{\langle g_z, S^i\rangle \chi_S(p)}_{x_p}\right) \end{aligned} \tag{14}$$

Notice that for $\alpha_p = \frac{g_z(p)}{\langle g_z, S^i\rangle}$, we have:

$$\sum_{p\in S^i} \alpha_p = \sum_{p\in S^i} \frac{g_z(p)}{\langle g_z, S^i\rangle} = \frac{\langle g_z, S^i\rangle}{\langle g_z, S^i\rangle} = 1 \tag{15}$$

Therefore, for F_z convex, applying the Jensen's inequality to (14) with $x_p = \langle g_z, S^i\rangle \chi_S(p)$ and $\alpha_p = \frac{g_z(p)}{\langle g_z, S^i\rangle}$ yields the following bound on $F_z\left(\langle g_z, S\rangle\right)$:

$$\begin{aligned} F_z\left(\langle g_z, S\rangle\right) \leq& \sum_{p\in S^i} \frac{g_z(p)}{\langle g_z, S^i\rangle} F_z\left(\langle g_z, S^i\rangle \chi_S(p)\right) \\ =& \sum_{p\in S} \frac{g_z(p)}{\langle g_z, S^i\rangle} F_z\left(\langle g_z, S^i\rangle\right) \\ &+ \sum_{p\in S^-} \frac{g_z(p)}{\langle g_z, S^i\rangle} F_z(0) \\ =& \frac{\langle g_z, S\rangle}{\langle g_z, S^i\rangle} F_z\left(\langle g_z, S^i\rangle\right) + \frac{\langle g_z, S^-\rangle}{\langle g_z, S^i\rangle} F_z(0) \end{aligned} \tag{16}$$

where $S^- = S^i \setminus S$.

Also, because $S = S^i \setminus S^-$, we have the following equality:

$$\frac{\langle g_z, S\rangle}{\langle g_z, S^i\rangle} = \frac{\langle g_z, S^i\rangle - \langle g_z, S^-\rangle}{\langle g_z, S^i\rangle} = 1 - \frac{\langle g_z, S^-\rangle}{\langle g_z, S^i\rangle} \tag{17}$$

Plugging (17) in the bound in (16), and after some manipulations, we obtain:

$$F_z(\langle g_z, S\rangle) \leq F_z(\langle g_z, S^i\rangle) + \langle a_z^i, S^-\rangle \tag{18}$$

Summing both sides of inequality (18) over z, we obtain the following bound of the general non-linear functional in (6):

$$R(S) \leq R(S^i) + \left\langle \sum_{z\in Z} a_z^i, S^- \right\rangle \tag{19}$$

Now notice the following inequality $\forall \alpha \in \mathbb{R}^+$

$$\begin{aligned} R(S^i) =& (1+\alpha)R(S^i) - \alpha R(S^i) \\ \leq& (1+\alpha)R(S^i) - \alpha R(S^i)\frac{\langle 1, S\rangle}{\langle 1, S^i\rangle} \\ =& (1+\alpha)R(S^i) - \alpha\left\langle b(S^i), S\right\rangle \end{aligned} \tag{20}$$

The inequality in (20) holds when R is nonnegative (this is the case when F_z is nonnegative), and is due to the fact that $\langle 1, S\rangle \leq \langle 1, S^i\rangle$ for $S \subset S^i$.

Finally, combining (19) and (20) proves proposition 1. □

Proposition 2. *If F_z is convex $\forall z \in Z$, $A_\alpha(S, S^i) + Q(S)$ is an auxiliary functional of $E(S) = R(S) + Q(S)$ for non-linear terms $R(S)$ of the form in (6), $\forall \alpha \in \mathbb{R}^+$ and for any segment S verifying $S \subset S^i$*

Proof. To prove that $A_\alpha + Q$ is an auxiliary function of $E = R + Q$, we need to have two conditions: (i) $A_\alpha(S, S^i) + Q(S)$ obeys the bound condition of the form in (4a), i.e., $E(S) \leq A_\alpha(S, S^i) + Q(S)$; and (ii) $A_\alpha(S, S) + Q(S)$ verifies the equality condition of the form in (4b), i.e., $A_\alpha(S, S) + Q(S) = E(S)$. The bound condition follows directly from the result in proposition 1. For the equality condition, it suffices to see that when $S^i = S$, we have $S^- = S^i \setminus S = \emptyset$, i.e., $\chi_{S^-}(p) = 0 \quad \forall p \in \Omega$. Therefore, we have $\langle \sum_{z\in Z} a_z^i, S^-\rangle = 0$. Now writing $A_\alpha(S, S^i)$ for $S^i = S$ and adding $Q(S)$ to both sides, we obtain:

$$\begin{aligned} A_\alpha(S, S) + Q(S) =& (1+\alpha)R(S) - \alpha\left\langle \frac{R(S)}{\langle 1, S\rangle}, S\right\rangle + Q(S) \\ =& (1+\alpha)R(S) - \alpha R(S)\frac{\langle 1, S\rangle}{\langle 1, S\rangle} + Q(S) \\ =& R(S) + Q(S) = E(S) \end{aligned} \tag{21}$$

□

3. Non-linear functions of the ratio of linear terms

In this section, we focus on functionals $E = J + Q$, where J is a non-linear, general-form functional containing the ratio of linear terms:

$$J(S) = \sum_{z \in Z} F_z \left(\frac{\langle g_z, S \rangle}{\langle h_z, S \rangle} \right) \tag{22}$$

where $g_z, h_z : \Omega \to \mathbb{R}^+$, $z \in Z$, are two families of non-negative scalar functions defined over the image domain, and $F_z : C \subset \mathbb{R} \to \mathbb{R}$, $z \in Z$, is a family of scalar functions defined over a convex domain C. High-order functionals of the form (22) are very useful in vision [8, 20, 1] and medical applications [2, 12]. They include but are not limited to several affinity measures on the distributions or moments of segment appearance [8, 20, 1] and shape [12, 2]. Following are some examples of functionals of the form (22).

3.1. Probability product kernels

Here we introduce the probability product kernel [10] as a segmentation constraint that can be viewed as a generalization of the Bhattacharyya measure, which has recently led to competitive performances in segmentation [20, 1]:

$$-\sum_{Z} \left(\frac{\langle k_z, S \rangle}{\langle 1, S \rangle} \right)^{\rho} M_z^{\rho}, \quad \rho \in]0, 1] \tag{23}$$

where $\{M_z, z \in Z\}$ is a model probability distribution learned *a priori* (or interactively), and ratio $\frac{\langle k_z, S \rangle}{\langle 1, S \rangle}$ is the kernel density estimate of the distribution of image data within segment S, with k_z the Gaussian kernel (σ the width of the kernel):

$$k_z(p) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} exp^{-\frac{\|z - I_p\|^2}{2\sigma^2}} \tag{24}$$

Minimization of (23) aims at finding a segment S whose distribution most closely matches model $\{M_z, z \in Z\}$. Notice that the Bhattacharyya segmentation measure in [1, 20, 8] is a particular case of (23), which corresponds to $\rho = \frac{1}{2}$.

3.2. Kullback-Leibler divergence

Minimization of this type of constraints enforces a *Kullback-Leibler* consistency between the distribution of the target segment S and a given model $\{M_z, z \in Z\}$ [17]:

$$\sum_{z \in Z} M_z \log \left(\frac{M_z}{\frac{\langle k_z, S \rangle}{\langle 1, S \rangle} + \epsilon} \right) = \underbrace{\sum_{z \in Z} M_z \log M_z}_{Constant} - \underbrace{\sum_{z \in Z} M_z \log \left(\frac{\langle k_z, S \rangle}{\langle 1, S \rangle} + \epsilon \right)}_{Variable} \tag{25}$$

ϵ is a small positive constant that avoids division by 0.

3.3. A general auxiliary functional

Proposition 3. *If $\forall z \in Z$ we have F_z convex, monotonically decreasing and defined at 0, and given an auxiliary segment S^i, the following functional is an upper bound of the general form $J(S)$ in (22), $S \subset S^i$, $\forall \alpha \in \mathbb{R}^+$ if J is non-negative and for $\alpha = 0$ if J is negative:*

$$B_\alpha(S, S^i) = (1+\alpha) J(S^i) - \alpha \left\langle d(S^i), S \right\rangle + \left\langle \sum_{z \in Z} c_z^i, S^- \right\rangle \tag{26}$$

$$c_z^i(p) = \frac{g_z(p) \left[F_z(0) - F_z \left(\frac{\langle g_z, S^i \rangle}{\langle h_z, S^i \rangle} \right) \right]}{\langle g_z, S^i \rangle} \tag{27a}$$

$$d(S^i) = \frac{J(S^i)}{\langle 1, S^i \rangle} \tag{27b}$$

Proof. Because h_z is nonnegative and $S \subset S^i$, we have $\langle h_z, S \rangle \leq \langle h_z, S^i \rangle$. Therefore, when F_z is monotonically decreasing and using the fact that $\langle g_z, R \rangle$ is nonnegative, we have the following upper bound on the general non-linear term in (22):

$$J(S) = \sum_{z \in Z} F_z \left(\frac{\langle g_z, S \rangle}{\langle h_z, S \rangle} \right) \leq \sum_{z \in Z} F_z \left(\frac{\langle g_z, S \rangle}{\langle h_z, S^i \rangle} \right) = \sum_{z \in Z} G_z^i \left(\langle g_z, S \rangle \right) \tag{28}$$

where $G_z^i(t)$ is given by: $G_z^i(t) = F_z \left(\frac{t}{\langle h_z, S^i \rangle} \right)$. Now because F_z is convex, G_z^i is also convex and the upper bound in (28) reduces to the general non-linear term of the type in (6). Therefore, we can apply the result we obtained earlier in Eq. (19) to $\sum_{z \in Z} G_z^i \left(\langle g_z, S \rangle \right)$, thereby getting (after some manipulations):

$$J(S) \leq J(S^i) + \left\langle \sum_{z \in Z} c_z^i, S^- \right\rangle \tag{29}$$

The rest of the proof follows the same steps as the proof of proposition 1. □

Proposition 4. *If $\forall z \in Z$ we have F_z convex, monotonically decreasing and defined at 0, $B_\alpha(S, S^i) + Q(S)$ is an auxiliary functional of $E(S) = J(S) + Q(S)$ (with S^i auxiliary segment and $S \subset S^i$) for non-linear terms $J(S)$ of the form in (22), $\forall \alpha \in \mathbb{R}^+$ if J is non-negative and for $\alpha = 0$ if J is negative*

Proof. The proof follows the same steps as the proof of proposition 2. □

Note on the Bhattacharyya bound in [1]: Although the derivation in [1] is completely different from our general-form derivation (The derivation in [1] relies on very specific properties of Bhattacharyya coefficient, and is not applicable to other non-linear terms), we can show that the Bhattacharyya bound in [1] can be obtained easily as a special case of $B_\alpha(S, S^i)$.

4. A general bound-optimization algorithm

The results in propositions 2 and 4 instruct us to derive the following general-purpose bound-optimization algorithm for non-linear functionals of the forms $E(S) = R(S) + Q(S)$ with R of the type (6) and $E(S) = J(S) + Q(S)$ with J of the type (22).

Algorithm 1 Bound optimization cuts

1. Iter. $i = 0$: Initialize the auxiliary segment by S^0; S^0 can be a trivial segment corresponding to the whole image domain, i.e., $S^0 = \Omega$.
2. For each iter. $i, i = 0, 1, 2, \ldots$, repeat the following steps until convergence:
   (a) **Update step:** Update the auxiliary functional by computing
   $$a_z^i \text{ and } b(S^i) \text{ if } E = R + Q$$
   $$c_z^i \text{ and } d(S^i) \text{ if } E = J + Q$$
   (b) **Optimization step:** optimize the auxiliary functional over S with a graph cut (S^i is fixed):
   $$S^{i+1} = \min_{S \subset S^i} A_\alpha(S, S^i) + Q(S) \text{ if } E = R + Q$$
   $$S^{i+1} = \min_{S \subset S^i} B_\alpha(S, S^i) + Q(S) \text{ if } E = J + Q$$

Optimization in step 2.b with a graph cut: Auxiliary functionals $A_\alpha + Q$ and $B_\alpha + Q$ are in the form of a sum of *unary* and *pairwise (submodular)* terms. In combinatorial optimization, a global optimum of such sum can be computed efficiently in low-order polynomial time with a graph cut [5]. Furthermore, condition $S \subset S^i$ can be easily imposed on the solution by adding a hard constraint [3]. We used the well-known Boykov-Kolmogorov (BK) algorithm [5] for the optimization in *step 2.b*. Here we omit the details of the BK algorithm and hard constraints. Such details are well-known in prior art, and can be found in [5, 3].

5. Experiments

We report several experiments on color and medical data, along with quantitative comparisons to the methods in [8, 22] in regard to convergence speed and accuracy. Each of the functionals is combined with a standard 16-neighborhood boundary smoothness term $Q(S) = \lambda \sum_{(p,q)} w_{pq}[1 - \delta(\chi_S(p) - \chi_S(q))]$ [8], with w_{pq} some contrast-sensitive weights.

Convergence-speed example ($\mathcal{L}_2$-distance): The example in Fig. 1 depicts a comparison to the recent method in [8], and demonstrates that auxiliary cuts can lead to a competitive performance in regard to convergence speed and accuracy. We used the same functional ($E = R + Q$, with R the $\mathcal{L}_2$-distance between histograms and smoothness weight $\lambda = 10$) and initialization for both methods. Similarly to the experiments in [8], the model histogram was obtained from the ground-truth segment. The feature is color specified in RGB coordinates. The plots show the functional and error (number of misclassified pixels) as functions of the iteration number, and the arrows in the plots indicate the convergence points. The bound optimizer yielded a faster decrease of the functional and converged after 4 iterations (it requires only 1 graph cut per iteration), whereas the method in [8] performed 10 iterations to reach convergence (and may require more that one graph cut per iteration). Notice that the proposed method yielded a lower error, which indicates that the non-linear term R obtained at convergence is lower. The overall functional (i.e., $R + Q$) is approximately the same as [8].

Medical image example ($\mathcal{L}_1$-distance): Fig. 2 depicts an example of segmenting the vertebral bodies (VB) in a difficult lumbar spine MRI image using a simple 4-point user input. We learned a histogram h_V of one VB from the user input, and used $7 \times h_V$ as model to segment the whole spine (lumbar spine images contain 7 VBs). We optimized the $\mathcal{L}_1$-distance[1]. The features we used for each pixel p are the means of intensities within 3 different patches centered at p (1, 3×3 and 10×1). The bound optimizer segmented successfully all the VBs, whereas the Boykov-Jolly model (B-J) [4] yielded several incorrect segments (We used the same training mask, number of bins and smoothness weight for

[1] Note that for the case of the $\mathcal{L}_1$-distance, the method of Gorelick et al. [8] is not applicable because it requires the functional to be differentiable (in order to compute a first-order approximation).

Constraint	General form	Auxiliary functional	$F_z(t)$	$g_z(p)$	$h_z(p)$
Volume constraint in (9)	$E = R + Q$, with $R(S)$ given by (6)	$A_\alpha(S, S^i) + Q(S)$, $\alpha \in \mathbb{R}^+$	$(t - v_z)^2$, $z \in \{1\}$	1, $z \in \{1\}$	$--$
$\mathcal{L}_j$*-distance between histograms in* (7), $j \in \mathbb{N}^*$	$E = R + Q$, with $R(S)$ given by (6)	$A_\alpha(S, S^i) + Q(S)$, $\alpha \in \mathbb{R}^+$	$\lvert t - h_z \rvert^j$, $j \in \mathbb{N}^*$ $z \in Z$	$\delta(I_p - z)$, $z \in Z$	$--$
Probability product kernels in (23)	$E = J + Q$, with $J(S)$ given by (22)	$B_\alpha(S, S^i) + Q(S)$, $\alpha = 0$	$-(tM_z)^\rho$, $\rho \in]0, 1]$, $z \in Z$	k_z in (24), $z \in Z$	1, $z \in Z$
Kullback-Leibler divergence in (25)	$E = J + Q$, with $J(S)$ given by (22)	$B_\alpha(S, S^i) + Q(S)$, $\alpha \in \mathbb{R}^+$	$-M_z \log(t + \epsilon)$, $z \in Z$	k_z in (24), $z \in Z$	1, $z \in Z$

Table 1. Several examples of segmentation functionals containing non-linear terms of the types $R(S)$ in (6) or $J(S)$ in (22), and the corresponding auxiliary functionals. Notice that F_z is convex for all the examples, and is monotonically decreasing for the examples involving form $J(S)$ in (22).

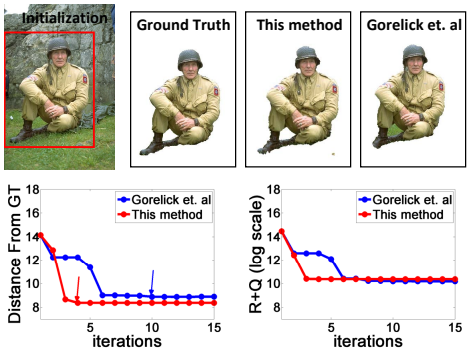

Figure 1. $\mathcal{L}_2$-distance example: log-scale plots of the functional and distance from ground truth (or error) for bound optimization and the method in [8]. The bound optimizer yielded a steeper decrease in the error/functional (the arrows indicate convergence points). Notice that we obtained a lower error, which indicates that the bound optimizer yielded a lower non-linear term R at convergence. The overall functional at convergence (i.e., $R + Q$) is approximately the same as [8]. $\lambda = 10$; Number of bins = 64^3; α=1.

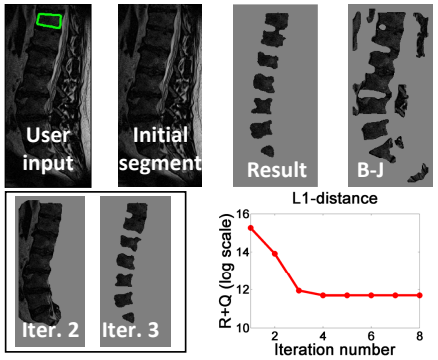

Figure 2. Spine image segmentation with two different methods: the $\mathcal{L}_1$-distance bound optimizer and the Boykov-Jolly (B-J) model [4]. $\lambda = 0.5$; Number of bins: 64^3; $\alpha = 10^{-2}$.

Method	Bound optimizer ($\mathcal{L}_2$-*distance*)	Rother et al. [22]
Error	1.09%	2.33%

Table 2. Evaluation of histogram matching on the GrabCut dataset (50 color images): average error for the bound optimizer ($\mathcal{L}_2$-*distance*) and the method of Rother et al. [22].

both methods). This is expected because the B-J model does not embed higher order information. Similarly to the previous experiment, the $\mathcal{L}_1$-*distance* bound optimizer yielded a steep decrease of the functional, using only 4 graph cuts.

Comparisons on the GrabCut data: We report a quantitative evaluation of the bound optimizer (the $\mathcal{L}_2$-*distance* version) on the GrabCut database [22] (50 color images with ground-truth segmentations) along with a comparison with the histogram-matching method of Rother et al. [22] (refer to Table 2). Similar evaluations were performed on the same data in [22]: Given a model learned from the ground truth, each image is segmented and the average error (percentage of misclassified pixels) w.r.t the ground truth was assessed. The parameters were the same for all the images: $\lambda = 8$ and $\alpha = 0.5$. The feature is color specified in RGB coordinates. The number of bins is 192 per channel. For each image, the initial segment is trivial: $S^0 = \Omega$. The results in Table 2 demonstrate that the bound optimizer obtained a significant improvement in accuracy over the method in [22].

A representative set of the results we obtained on the GrabCut data is shown in Fig. 3. Even though each of these examples contains a camouflage (an overlap between the distributions of the foreground/background segments), the bound optimizer obtained accurate image partitions, and consistently converged within less than 5 graph cuts. On a E5440 quad core 2.83 GHz Xeon machine, with a 3.25GB of RAM, the bound optimizer required between 3 and 6 seconds to process an image.

Example with the KL divergence: The example in Fig. 4 confirms the fast convergence and very steep functional

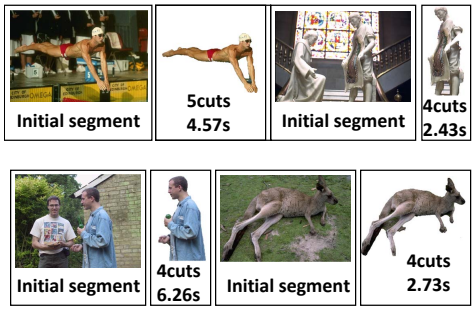

Figure 3. A representative sample of the segmentations obtained with the bound optimizer ($\mathcal{L}_2$-*distance*) using some camouflage examples in the GrabCut data. For each example, we give the number of cuts/iterations before convergence and the CPU time.

Figure 4. An example with the KL divergence: $\lambda = 10^{-3}$; Number of bins: 64^3; $\alpha = 10^{-2}$. The plot depicts the functional versus the number of cuts/iterations.

decrease obtained by the proposed optimizers in the case of the KL divergence (we used an example from the GrabCut data). We show the initial/final segments, and plot the functional versus the number of iterations. For this example, the bound optimizer required only two graph cuts.

References

[1] I. Ben Ayed, H.-M. Chen, K. Punithakumar, I. Ross, and S. Li. Graph cut segmentation with a global constraint: Recovering region distribution via a bound of the bhattacharyya measure. In *CVPR*, pages 3288–3295, 2010. 1, 2, 5, 6

[2] I. Ben Ayed, K. Punithakumar, G. J. Garvin, W. Romano, and S. Li. Graph cuts with invariant object-interaction priors: Application to intervertebral disc segmentation. In *IPMI*, pages 221–232, 2011. 1, 2, 5

[3] Y. Boykov and G. Funka Lea. Graph cuts and efficient n-d image segmentation. *International Journal of Computer Vision*, 70(2):109–131, 2006. 6

[4] Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. In *ICCV*, pages 105–112, 2001. 1, 6, 7

[5] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004. 1, 6

[6] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011. 1

[7] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991. 4

[8] L. Gorelick, F. R. Schmidt, and Y. Boykov. Fast trust region for segmentation. In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, Portland, Oregon, June 2013. 1, 2, 3, 5, 6, 7

[9] L. Gorelick, F. R. Schmidt, Y. Boykov, A. Delong, and A. Ward. Segmentation with non-linear regional constraints via line-search cuts. In *ECCV*, pages 583–597, 2012. 1, 2

[10] T. Jebara, R. I. Kondor, and A. Howard. Probability product kernels. *Journal of Machine Learning Research*, 5:819–844, 2004. 5

[11] H. Jiang. Linear solution to scale invariant global figure ground separation. In *CVPR*, pages 678–685, 2012. 1, 2, 3

[12] M. Klodt and D. Cremers. A convex framework for image segmentation with moment constraints. In *ICCV*, pages 2236–2243, 2011. 1, 2, 5

[13] V. Kolmogorov and C. Rother. Minimizing nonsubmodular functions with graph cuts-a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7):1274–1279, 2007. 2

[14] K. Lange, D. R. Hunter, and I. Yang. Optimization transfer using surrogate objective functions. *Journal of Computational and Graphical Statistics*, 9(1):1–20, 2000. 3

[15] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, pages 556–562, 2000. 3

[16] V. Lempitsky, P. Kohli, C. Rother, and T. Sharp. Image segmentation with a bounding box prior. In *ICCV*, pages 277–284, 2009. 1

[17] A. Mitiche and I. Ben Ayed. *Variational and Level Set Methods in Image Segmentation*. Springer, first edition edition, 2010. 2, 5

[18] L. Mukherjee, V. Singh, and C. R. Dyer. Half-integrality based algorithms for cosegmentation of images. In *CVPR*, pages 2028–2035, 2009. 1, 2, 3

[19] M. Narasimhan and J. A. Bilmes. Submodular-supermodular procedure with applications to discriminative structure learning. In *UAI*, pages 404–412, 2005. 2

[20] V.-Q. Pham, K. Takahashi, and T. Naemura. Foreground-background segmentation using iterated distribution matching. In *CVPR*, pages 2113–2120, 2011. 1, 2, 5

[21] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23:309–314, 2004. 1

[22] C. Rother, T. P. Minka, A. Blake, and V. Kolmogorov. Cosegmentation of image pairs by histogram matching - incorporating a global constraint into mrfs. In *CVPR*, pages 993–1000, 2006. 2, 3, 6, 7

[23] S. Vicente, V. Kolmogorov, and C. Rother. Cosegmentation revisited: Models and optimization. In *ECCV(2)*, pages 465–479, 2010. 2, 3

[24] O. J. Woodford, C. Rother, and V. Kolmogorov. A global perspective on map inference for low-level vision. In *ICCV*, pages 2319–2326, 2009. 2

[25] Z. Zhang, J. T. Kwok, and D.-Y. Yeung. Surrogate maximization/minimization algorithms and extensions. *Machine Learning*, 69:1–33, 2007. 3