# EXTENDING $\alpha$-EXPANSION TO A LARGER SET OF REGULARIZATION FUNCTIONS

*Mathias Paget, Jean-Philippe Tarel*[*]

Lepsis/Cosys, IFSTTAR,
14-20 Boulevard Newton,
F-77420 Champs-sur-Marne,
Université Paris-Est, France

*Laurent Caraffa*

Matis, IGN,
73 Avenue de Paris,
F-94165 Saint-Mandé, France

## ABSTRACT

Many problems of image processing lead to the minimization of an energy, which is a function of one or several given images, with respect to a binary or multi-label image. When this energy is made of unary data terms and of pairwise regularization terms, and when the pairwise regularization term is a metric, the multi-label energy can be minimized quite rapidly, using the so-called $\alpha$-expansion algorithm. $\alpha$-expansion consists in decomposing the multi-label optimization into a series of binary sub-problems called move. Depending on the chosen decomposition, a different condition on the regularization term applies. The metric condition for $\alpha$-expansion move is rather restrictive. In many cases, the statistical model of the problem leads to an energy which is not a metric. Based on the enlightening article [1], we derive another condition for $\beta$-jump move. Finally, we propose an alternated scheme which can be used even if the energy fulfills neither the $\alpha$-expansion nor $\beta$-jump condition. The proposed scheme applies to a much larger class of regularization functions, compared to $\alpha$-expansion. This opens many possibilities of improvements on diverse image processing problems. We illustrate the advantages of the proposed optimization scheme on the image noise reduction problem.

***Index Terms*—** Minimization, Discrete optimization, Regularization, Markov Random Field, $\alpha$-expansion, Graph-cuts, Denoising,Noise reduction.

## 1. INTRODUCTION

Image processing problems are usually set as the minimization of an energy with respect to the unknown variables of the problem. The Bayesian approach provides ways to derive the energy of the problem from its statistical model. This energy is a function of the observations and of the unknown variables which are both numerous. In image processing, the used statistical models are generally Markov Random Fields. If several optimization methods for large problems are available, each method has its own field of application due to restrictive conditions of use on the energy. Matching between the energy derived from the Bayesian approach and the conditions of use of the optimization method is often tricky.

$\alpha$-expansion is a discrete multi-label optimization method, introduced by [2, 3], where successive binary sub-problems are solved. The sub-problem, which is parametrized by $\alpha$, consists in deciding for every pixels if setting the current pixel label to $\alpha$ allows to decrease the total energy. When the energy of the sub-problem is sub-modular, the binary sub-problem can be minimized to a global minimum. Nevertheless the $\alpha$-expansion algorithm only converges towards a local minimum of the energy w.r.t $\alpha$-expansion move space, as a succession of decreasing steps of the energy. $\alpha$-expansion is well known as one of the fastest algorithm to optimize an energy on a graph, with unary data terms and pairwise regularization data terms. However, the condition on the energy for $\alpha$-expansion to be applied, is that the regularization term must be a metric. Examples of metric pairwise terms are concave functions of the absolute difference of the two labels.

This condition is rather restrictive. For instance, $\alpha$-expansion can not cope with the standard Gaussian prior. Indeed, when a Gaussian pairwise prior is used, the regularization term is a quadratic function of the difference of the two labels. When the pairwise term is a concave continuous function of the absolute difference of two labels, this implies that the regularization function is not smooth when the two labels are equals. This is an important limitation, since due to the presence of Gaussian noise or slow variations, it is useful to have a locally quadratic shape at the zero of the function which applies to the two labels difference. As pointed in [3], this property is also important in disparity reconstruction from stereo images of thin objects without filling the hole between them.

Our objective is thus to find ways to extend $\alpha$-expansion to a larger set of energies. In Sec. 2, from the enlightening theory summarized in [1], a simple implementation of the $\alpha$-expansion algorithm is obtained and the condition the regularization term must fulfill is again derived. Then, the same kind of derivation is performed, for a different case, where the sub-problem is now a $\beta$-jump. A sufficient condition of use

[*]Thanks to FUI AWARE project for funding.

is that the pairwise regularization term is a convex function of the label difference. Then, we propose in Sec. 3, a scheme to optimize an energy where the pairwise regularization term is neither a concave nor a convex function of the label difference. In Sec. 4, we illustrate the advantage of the proposed optimization scheme on the image noise reduction problem.

## 2. MULTI-LABEL MINIMIZATION

We consider the following energy $E(l)$ to be minimized with respect to the label image $l$:

$$E = \sum_{u \in I} g(l_u) + \sum_{(u,v) \in N} f(l_u, l_v), \quad (1)$$

where $I$ is the set of sites in the image, $N$ is the set of neighborhood links, $u$ and $v$ are sites, $f$ is the pairwise regularization function and $g$ is the unary data cost function. We assume that function $f$ is non-negative. $l_u$ denotes the label at the site $u$. Labels are assumed in a set $L$ of ordered labels.

Let us recall that the global minimum of such discrete energy can be obtained, when the function $f$ is convex, as shown in [4], using graph-cuts. However, this algorithm requires many computations. In practice, approximate solutions, such as [2], are usually preferred. In this latter approach, the initial multi-label problem is decomposed into successive binary sub-problems, which are solved using graph-cuts. The binary sub-problem consists in choosing or not a new label for each site, given a rule. This is called a *move*. A set of moves which spans the label entire set is named a *move space*. Different rules and thus different move spaces were proposed such as: $\alpha$-expansion, jump, swap, relabeling [3]. The set of regularization terms which can be used in the energy is different for each type of move.

In [1], it was proposed to use the Quadratic Pseudo Boolean Optimization (QPBO) theory to rewrite binary sub-problems using a pseudo-Boolean function. When the obtained quadratic pseudo-Boolean function is sub-modular, i. e. coefficients before quadratic terms are all negative, a graph can be built and the sub-problem can be minimized as a maximum flow optimization on the associated graph [5]. As a consequence, the QPBO theory gives us a way to derive the condition the pairwise regularization term must fulfill to be used with a given type of move spaces.

Let $l_u$ be the current label at site $u$, and $\hat{l_u}$ the newly proposed label at this site. The sub-modularity of the binary problem implies, for the all pairs of sites $(u, v) \in N$:

$$f(l_u, l_v) + f(\hat{l_u}, \hat{l_v}) \leq f(l_u, \hat{l_v}) + f(\hat{l_u}, l_v), \quad (2)$$

Notice that this inequality is fulfilled when $l_u = \hat{l_u}$ or when $l_v = \hat{l_v}$. We now focus on two kinds of moves: $\alpha$-expansion and $\beta$-jump moves.

### 2.1. $\alpha$-expansion

An $\alpha$-expansion move consists in proposing a label $\alpha$ for every site $u$, where $\alpha$ is given. This move can be formally written as $\hat{l_u} = \alpha$, $\forall u \in I$. In such a case, the condition of sub-modularity (2) becomes:

$$f(l_u, l_v) + f(\alpha, \alpha) \leq f(l_u, \alpha) + f(\alpha, l_v) \quad (3)$$

If $f(x, x)$ is a constant $k$, for all $u$, the previous condition is fulfilled when the function $f(x, y) - k$ is a metric. We now consider a useful particular case, where $f(l_u, l_v)$ is assumed to be a function $f^*$ of the label difference, i.e, $f(l_u, l_v) = f^*(l_u - l_v)$, then the condition is :

$$f^*(x + y) + f^*(0) \leq f^*(x) + f^*(y), \quad (4)$$

for all $x, y \in \mathbb{R}^2$. This means that $h(x) = f^*(x) - f^*(0)$ is sub-additive on $\mathbb{R}$. A sufficient condition for $h(x)$ to be sub-additive is to be even and concave on $\mathbb{R}^+$. Such functions can not be smooth in zero. As a consequence, $f^*$ is not smooth on $\mathbb{R}$. This is an important limitation in the usage of $\alpha$-expansion, since we are interested in using a smooth function for the pairwise regularization term.

In summary, a first sufficient condition of use of the $\alpha$-expansion is that the pair-wise regularization term is a metric. A second sufficient condition is that it is an increasing concave function of the label absolute difference. Following [1], it is not difficult to derive the graph where max flow should be applied to obtain the solution of each $\alpha$-expansion binary sub-problem.

### 2.2. $\beta$-jump move

A $\beta$-jump move consists in proposing an increment or decrements of the current labels by value $\beta$, for every site $u$. Formally, $\hat{l_u} = l_u + \beta$, $\forall u \in I$. For a $\beta$-jump, the condition of sub-modularity (2) becomes:

$$f(l_u, l_v) + f(l_u + \beta, l_v + \beta) \leq f(l_u, l_v + \beta) + f(l_u + \beta, l_v). \quad (5)$$

Like in the previous section, we now consider the particular case, where $f(l_u, l_v)$ is assumed to be a function $f^*$ of the label difference. The previous inequality can be rewritten as:

$$2f^*(x) \leq f^*(x + \beta) + f^*(x - \beta), \quad (6)$$

for all $x, \beta$. After substitution of $x = (y + z)/2$ and $\beta = (y - z)/2$, we deduce that $f^*$ is mid-convex:

$$f^*(\frac{y + z}{2}) \leq \frac{f^*(y) + f^*(z)}{2}, \quad (7)$$

From Bernstein-Doetsch theorem, $f^*$ being also upper bounded, $f^*(x)$ is a convex function on the label interval.

In summary, considering pair-wise regularization with the label difference, a sufficient and necessary condition of use of the $\beta$-jump is that the pair-wise regularization term is convex. Following [1], it is also not difficult to derive the graph where max flow should be applied to obtain the solution of each $\beta$-jump binary sub-problem.

### 2.3. Extended Smooth Exponential Family

Given a problem, the choice of the energy can be interpreted, in the Bayesian approach, as the implicit choice of a statistical model of the problem. As explained in [6, 7], a useful family of probability distribution functions (pdf) named Smooth Exponential Family (SEF) can be used to help in the statistical modeling. The advantage of the SEF family is that it is parameterized by only two parameters: the shape parameter $a$ and the scale parameter $s$. To better fit observed data, we proposed an Extended SEF family (ESEF) where an extra parameter $k$ is added. The ESEF pdf is defined, up to a normalization factor, as $exp(-ESEF(b))$ where:

$$ESEF(b) = \frac{k}{a}((1 + \frac{b^2}{2s^2})^a - 1). \tag{8}$$

All functions in the ESEF family are smooth and several well known distributions can be found for particular values of $a$, for instance: Gaussian pdf for $a = 1$, smooth Laplace pdf for $a = 0.5$, Cauchy and T pdfs as a limit case when $a$ goes towards 0, Geman&Maclure pdf for $a = -1$. Using the Bayesian approach, it is the minus of the log of the pdf which appears in the energy, i.e $ESEF(b)$ defined in (8).

The behavior of the $ESEF$ function around zero is always quadratic. Therefore, $\alpha$-expansion can not be used. When the $ESEF$ function is convex, i.e when $a \geq 0.5$, the $\beta$-jump can be used, but not in the other cases. Another example of useful regularization function which can not be used, neither with $\alpha$-expansion nor with $\beta$-jump, is the truncated quadratic function.

## 3. PROPOSED APPROACH

To be able to cope with a large set of pairwise regularization terms, we propose an alternated scheme between an $\alpha$-expansion move space and a $\beta$-jump move space. Since the condition of use of each decreasing energy algorithm is not always fulfilled, the optimization is limited to sites where the problem is sub-modular.

For a site where the sub-modularity condition (2) is verified for all its links, a new label is proposed. But, when this condition is not fulfilled, the same label as the current label is proposed. As a consequence, during an $\alpha$-expansion move or a $\beta$-jump move, only the sites where the sub-modularity condition is verified can be modified. We call this variant as partial. This means that, if the energy is decreased at each step, the energy will not necessarily converge to a local minimum due to the possibility of sites with a fixed label. The advantage of alternating between partial $\alpha$-expansion and partial $\beta$-jump moves is to try to minimize the number of sites with a fixed label due to the difference in their conditions of use. The alternated optimization is stopped when neither $\alpha$-expansion nor $\beta$-jump move spaces strictly decrease the energy.

The proposed scheme is shown in Algo. 1, with the partial $\alpha$-expansion in Algo. 2 and the partial $\beta$-jump in Algo. 3. Notice that since $\alpha$-expansion and $\beta$-jump moves are performed on the same graph structure, every graph-cut steps only need to update the graph values and not to rebuild the graph.

**Algorithm 1** $AlternatedOptimization$

```
Initialize labels
E_0 ← ComputeEnergy
while EnergyIsStrictlyDecreasing do
  for α ∈ L do
    PartialAlphaExpansion(α)
  for β ∈ L do
    PartialBetaJump(β)
    PartialBetaJump(−β)
```

**Algorithm 2** $PartialAlphaExpansion(\alpha)$

```
for i ∈ I do
  flag_i ← AllLinksSubmodular
for i ∈ I do
  if flag_i then
    l̂_i ← α
  else
    l̂_i ← l_i
  DoGraphCut
```

**Algorithm 3** $PartialBetaJump(\beta)$

```
for i ∈ I do
  flag_i ← AllLinksSubmodular
for i ∈ I do
  if flag_i & (l_i + β) ∈ L then
    l̂_i ← l_i + β
  else
    l̂_i ← l_i
  DoGraphCut
```

## 4. IMAGE NOISE REDUCTION

We test the proposed schemes on the image noise reduction problem. On each image, we compute the intensity difference histogram over all neighbors. A ESEF model is fitted on the observed distribution to estimate $f$, using Maximum Likelihood (ML) criterion. Then a Gaussian noise (with std 10) is added to the images, so $g$ is set as -log of this pdf, i.e a square function.

### 4.1. Fixed sites

Fig. 1(a) shows the "Lena" image. The used regularization function is a ESEF function with $a = 0.4$, $s = 10$ and $k = 1$.

(a) (b) (c)

**Fig. 1**: (a) Lena image, (b) in gray, pixels where intensity value may be fixed during $\alpha$-expansion (multiplied by 4), (c) fixed pixels during $\beta$-jump (multiplied by 2).

This function is locally convex around zero and concave for higher values. In Fig. 1(b), the number of fixed labels during the $\alpha$-expansion move is shown, with $\alpha \in L$. Fig. 1(c) shows the $\beta$-jump case. We remark that fixed pixels are mainly along objects edges in both cases. This can be explained by the concave part of the ESEF function when $\beta$-jump is used. When $\alpha$-expansion is used, fixed pixels only occur when the value of $\alpha$ is between the two neighbor label values. This case is thus more likely to happen at edges. Notice that on the edges, the numbers of fixed labels during $\alpha$-expansion are lower than for $\beta$-jump in our experiments.

### 4.2. Optimization scheme comparison

A pairwise $L_1$ regularization is used in order to compare performances of $\alpha$-expansion, $\beta$-jump and alternated scheme on the same energy. Indeed, $L_1$ regularization fulfills both conditions (3) and (5), and thus there is no pixel with a fixed label. Over 7 tested images, the 3 optimization schemes give the same final energy. This illustrates the consistancy between the different schemes.

### 4.3. Image noise reduction comparison

The comparison of the previous different methods with bilateral filter, in term of Peak Signal-to-Noise Ratios (PSNR), after parameter optimization, leads to quite similar PSNR. Nevertheless, obtained results differ in terms of smoothness. For instance, Fig. 2 presents details obtained from "Cameraman" and "Lena" images using $\alpha$-expansion (first column) and alternated scheme (second column). In the two first lines, the regularization term for the alternated scheme comes from the ESEF pdf fitting. To apply $\alpha$-expansion, the previous ESEF pdf is approximated by the closest concave function on $\mathbb{R}^+$. In the third line, a $L_1$ regularization is used with $\alpha$-expansion (left), and a smoothed $L_1$ (ESEF $a = 0.5$, $s = 3$, $k = 1$) function with the alternated scheme. One can notice the differences in areas of slow intensity variation, such as the sky in "Cameraman" or Lena's cheek. $\alpha$-expansion leads to a staircase effect due to the concave regularization. Results are smoother with alternated scheme, thanks to a regularization function which is locally convex around zero, even if the noise is a little less decreased in flat areas. On seven images, alternated scheme is slower by a factor between 1.1 to 2.5, compared to $\alpha$-expansion.

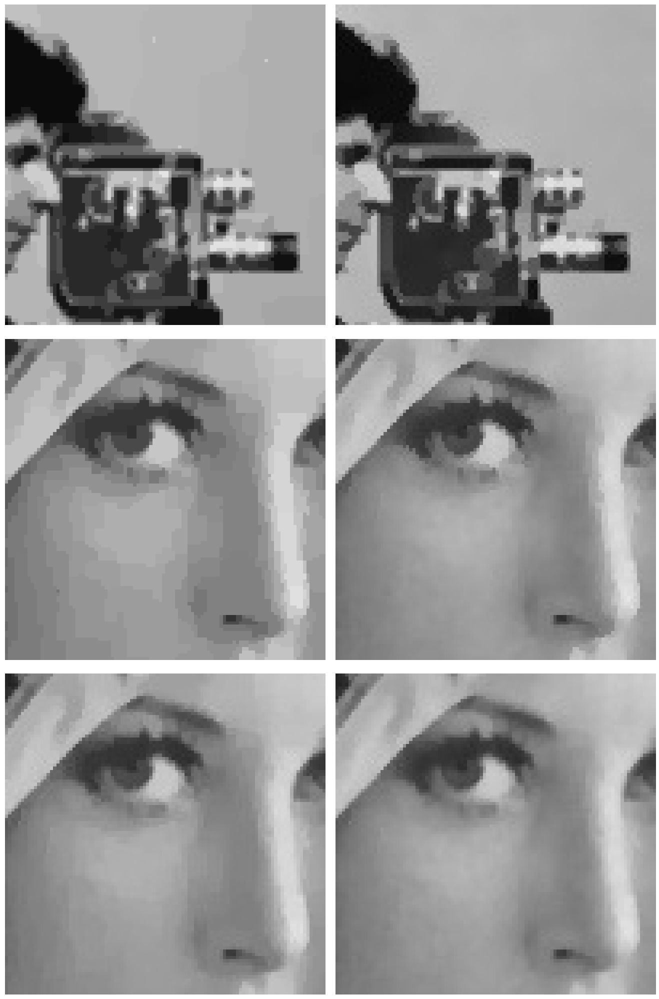

**Fig. 2**: Noise reduction obtained with $\alpha$-expansion (fist column) and Alternated scheme (second column), on line 1 and 2 for approximated ESEF and ESEF model, on line 3 for $L_1$ and Smoothed-$L_1$ model.

## 5. CONCLUSION

From [1], we derive again that concave functions of the absolute labels difference can be used with $\alpha$-expansion and we obtain that convex functions can be used with $\beta$-jump. This leads us to the idea to alternated between $\alpha$-expansion and $\beta$-jump to minimize an energy where regularization function is neither concave nor convex. The proposed optimization scheme can be applied to a set of energies which is much larger that the usable set for $\alpha$-expansion. We illustrated the advantages of the proposed scheme for image noise reduction. This opens many possibilities of improvements on diverse image processing problems.

## 6. REFERENCES


[1] Endre Boros and Peter L Hammer, "Pseudo-boolean optimization," *Discrete applied mathematics*, vol. 123, no. 1, pp. 155–225, 2002.

[2] Yuri Boykov, Olga Veksler, and Ramin Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222–1239, 2001.

[3] Olga Veksler, *Efficient graph-based energy minimization methods in computer vision*, Ph.D. thesis, Cornell University, 1999.

[4] Hiroshi Ishikawa, "Exact optimization for markov random fields with convex priors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 10, pp. 1333–1336, 2003.

[5] Vladimir Kolmogorov and Ramin Zabin, "What energy functions can be minimized via graph cuts?," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 2, pp. 147–159, 2004.

[6] Sio-Song Ieng, Jean-Philippe Tarel, and Pierre Charbonnier, "Modeling non-gaussian noise for robust image analysis," in *Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP'07)*, Barcelona, Spain, 2007, pp. 183–190.

[7] Laurent Caraffa, Jean-Philippe Tarel, and Pierre Charbonnier, "The guided bilateral filter: When the joint/cross bilateral filter becomes robust," *IEEE Transactions on Image Processing*, vol. 24, no. 4, pp. 1199–1208, Apr. 2015.