

Optimization of Divergences Within the Exponential Family for Image Segmentation

François Lecellier, Stéphanie Jehan-Besson, Jalal M. Fadili, Gilles Aubert, Marinette Revenu

▶ **To cite this version:**

François Lecellier, Stéphanie Jehan-Besson, Jalal M. Fadili, Gilles Aubert, Marinette Revenu. Optimization of Divergences Within the Exponential Family for Image Segmentation. 2nd International Conference on Scale Space and Variational Methods in Computer Vision, SSVM, 2009, Voss, Norway. pp.137-149, 10.1007/978-3-642-02256-2_12 . hal-00812626

HAL Id: hal-00812626
https://hal.science/hal-00812626

Submitted on 12 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimization of divergences within the exponential family for image segmentation

F. Lecellier[1], S. Jehan-Besson[2], J. Fadili[1], G. Aubert[3], M. Revenu[1]

[1] Laboratoire GREYC, University of Caen, France
[2] Laboratoire LIMOS, University of Clermont-Ferrand, France
[3] Laboratoire J.A. Dieudonné, University of Nice Sophia-Antipolis, France

Abstract

In this work, we propose novel results for the optimization of divergences within the framework of region-based active contours. We focus on parametric statistical models where the region descriptor is chosen as the probability density function (pdf) of an image feature (e.g. intensity) inside the region and the pdf belongs to the exponential family. The optimization of divergences appears as a flexible tool for segmentation with and without intensity prior. As far as segmentation without reference is concerned, we aim at maximizing the discrepancy between the pdf of the inside region and the pdf of the outside region. Moreover, since the optimization framework is performed within the exponential family, we can cope with difficult segmentation problems including various noise models (Gaussian, Rayleigh, Poisson, Bernoulli ...). We also experimentally show that the maximisation of the KL divergence offers interesting properties compare to some other data terms (e.g. minimization of the anti-log-likelihood). Experimental results on medical images (brain MRI, contrast echocardiography) confirm the applicability of this general setting.

1 Introduction

We propose here to focus on the segmentation of homogeneous regions in noisy images using statistical region-based active contour models (RBAC). In RBAC, region-based terms can be advantageously combined with boundary-based ones [1, 2]. The evolution equation is generally deduced from a general criterion to minimize that includes both region integrals and boundary integrals. The combination of those two terms in the energy functional allows the use of photometric image properties, such as texture [3] and noise [4], as well as geometric properties such as the shape prior of the object to be segmented. In statistical region-based active contours, see [5] for a review, image features (e.g. intensity) are considered as random variables whose distribution may be parametric (e.g. Gaussian) or non parametric [6]. Classically, the authors consider the minimization of the anti-log-likelihood for segmentation [7, 8, 4]. In this paper, we rather focus on the optimization of distance between pdfs. Such distances or more generally divergences can be used in two different manner. On the one hand, they can be used for segmentation with distribution intensity prior and in this case, we aim at minimizing the distance between the pdf of the evolving region and a reference one. On the other hand, they can be used for segmentation without reference and in this second

case, we aim at maximizing the distance between the pdf of the inside region and the pdf of the outside region. In the literature, the minimization of divergences between non parametric pdfs has first been proposed in [6] for video sequences. It has then been developed for cardiac structures tracking in perfusion MRI (p-MRI) sequences in [9]. As far as segmentation using the maximization of divergences is concerned, some authors [10] have also proposed to take benefit of the maximization of the Bhattacharya distance of non parametric pdfs for segmentation. On the other hand, divergences between Gaussian distributions have been developed for DTI segmentation in [11].

In this paper, we propose to set a general framework for the optimization of divergences between parametric pdfs within the exponential family. To the best of our knowledge, such a framework has never been studied for region-based active contour segmentation. The rationale behind using the exponential family is that it includes, among others, Gaussian, Rayleigh, Poisson and Bernoulli distributions that have proven to be useful to model the noise structure [4] in many real image acquisition devices (e.g. Poisson for photon counting devices such as X-ray or CCD cameras, Rayleigh for ultrasound images, etc). Using shape derivative tools as in [12, 6], our effort focuses on constructing a general expression for the derivative of the energy (with respect to a domain), and on deriving the corresponding evolution speed. Our general framework is also specialized to some particular cases, such as the optimization of the Kullback-Leibler (KL) divergence [13], which gives a simple expression of the derivative.

This theoretical framework is then more explicitly detailed and illustrated for the case of the segmentation without reference. In this case, we aim at maximizing the dissimilarity between the pdf of the intensity within the region inside the evolving contour and the pdf of the intensity within the region outside the contour. In other words, we perform a competition between the pdfs of these two regions through the maximisation of divergences. Experimental results are given for the particular case of the KL divergence. We experimentally compare this data term to the classical minimization of the anti-log-likelihood [7, 14] for the segmentation of the White Matter in brain MRI and we show that KL maximisation is able to extract a single Gaussian from a mixture of Gaussian. We also show the applicability of our data term for the segmentation of the left ventricle in contrast echocardiography where the noise is modelled using Rayleigh.

In this paper, we first set our general setting and introduce shape gradients in section 2. In section 3, we propose to give some general results for the exponential family and then for the shape derivative of divergences between pdfs. These results are then specialized for the KL divergence using the Maximum Likelihood Estimation (MLE) for the parameters. Experimental results for the maximization of KL divergence are given in section 4.

2 Optimization of divergences between pdfs : general setting

In this section, we set our general setting for segmentation through the optimization of distances between pdfs or more generally divergences.

2.1 General setting

Consider a function $y : \mathbb{R}^n \to \chi \subset \mathbb{R}$ which describes the feature of interest. The term $y(\mathbf{x})$ then represents the value of the feature y at location $\mathbf{x}$ where $\mathbf{x} \in \mathbb{R}^n$. Let $q(y, \Omega)$

be the probability density function (pdf) of the feature y within the image region of interest. We now assume that we have a function $\Psi : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$ which allows us to compare two pdfs. This function is small if the pdfs are similar and large otherwise. It allows us to introduce the following functional which represents the distance or more generally the divergence between the current pdf estimate $q(y,\Omega)$ and another one $p(y)$ which may also depend on another domain:

$$D(\Omega) = \int_{\chi} \Psi(q(y,\Omega), p(y))\, dy. \tag{1}$$

The distance can be for example the symmetrized Kullback-Leibler divergence when $\Psi(q,p) = \frac{1}{2}(p(y)\log\frac{p(y)}{q(y,\Omega)} + q(y,\Omega)\log\frac{q(y,\Omega)}{p(y)})$.

Such divergences represent a general setting for both segmentation with and without reference. Indeed, in segmentation problems, we generally search for homogeneous regions regarding with a given feature. We may then modelize the segmentation problem as the maximization of the distance between the pdf of the feature within the inside region and the pdf of the feature within the outside region. In order to fix ideas, let us consider a partition of an image in two regions where Ω is the inside region and Ω^c the complementary outside region. The segmentation may then be formulated as the maximization of the following criterion:

$$D(\Omega,\Omega^c) = \int_{\chi} \Psi(q(y,\Omega), p(y,\Omega^c))\, dy. \tag{2}$$

On the other hand, we can also consider that we have a reference histogram p_{ref} and that we search for the domain that minimizes the divergence between q and p_{ref}. This last framework may be applied to tracking or to supervised segmentation where a reference pdf is learned on the region of interest. The theoretical results given in this paper can be used for both applications.

2.2 Shape gradient descent

In order to find an optimum, we perform a shape gradient descent using region-based active contours. We then have to compute the derivative of the criterion according to the domain using shape derivation tools [15]. Shape derivative tools applied to region-based active contours are described in [12, 6] and we won't remind all the definitions in this paper. Let us just remind that, from the shape derivative, we can derive the evolution equation that will drive the active contour towards a (local) minimum of the criterion. Let us suppose that the shape derivative of the criterion $D(\Omega)$ in the direction $\mathbf{V}$ may be written as follows:

$$< D'(\Omega), \mathbf{V} >= -\int_{\partial\Omega} \mathbf{speed}(\mathbf{x},\Omega)(\mathbf{V}(\mathbf{x})\cdot\mathbf{N}(\mathbf{x}))d\mathbf{a}(\mathbf{x})\ , \tag{3}$$

where $\mathbf{N}$ is the unit inward normal to $\partial\Omega$ and da its area element.
When minimizing the distance $D(\Omega)$, interpreting equation (3) as the L^2 inner product on the space of velocities, the straightforward choice is to take $\mathbf{V} = \mathbf{speed}(\mathbf{x},\Omega)\mathbf{N}$. When minimizing the distance $D(\Omega)$, we can then deduce the following evolution equation:

$$\frac{\partial\Gamma}{\partial\tau} = \mathbf{speed}(\mathbf{x},\Omega)\,\mathbf{N}(\mathbf{x})\ , \tag{4}$$

On the contrary, when maximizing the criterion, we take the opposite sign for the velocity.

3 General results for shape derivative of divergences within the exponential family

In this paper, we consider that pdfs belong to the exponential family. In this case, the current pdf estimate $q(y,\Omega)$ is now indexed by a set of parameters $\theta \in \Theta \subset \mathbb{R}^{\kappa}$ (e.g. we have $\kappa = 2$ and $\theta = (\mu,\sigma)^T$ where μ is the mean and σ the variance for the Normal family). When using the exponential family, we rather index the pdf by η which is the natural parameter as explained below. In order to derive the criterion, we must take into account the dependence of the natural parameter with the domain. We then restrict our study to the full rank κ-parameter canonical exponential family [16]. For this family, we can establish a 1-1 correspondence between η and Ω and so compute directly the shape derivative of $D(\Omega)$.

In the sequel, let us first introduce the exponential family and some properties and then explain the computation of the shape derivative. We then specialize our result when parameters are estimated using the Maximum Likelihood Estimation (MLE) method. We also give some results for the optimization of the Kullback-Leibler (KL) divergence. In this case, the shape derivative reduces to a very simple general expression.

3.1 The exponential family: Definition and properties

The multi-parameter exponential family [17] is naturally indexed by a κ-dimensional real parameter vector η and a κ-dimensional natural statistic vector $\mathbf{T}(Y)$. We draw the reader's attention to the fact that η is a function of $\theta \in \Theta$ which is the parameter of interest in most applications (for the Gaussian distribution, we have $\theta = (\mu,\sigma)^T$).

Definition 1. *The family of distributions of a Random Variable (RV) Y $\{q_\theta : \theta \in \Theta \subseteq \mathbb{R}^{\kappa}\}$, is said a κ-parameter* **canonical** *exponential family, if there exists real-valued functions:*

- $\eta(\theta) = [\eta_1,...,\eta_\kappa]^T$ *with* $\eta_i : \Theta \subseteq \mathbb{R}^{\kappa} \to \mathbb{R}$
- $h : \mathbb{R} \to \mathbb{R}$
- $B : \Theta \to \mathbb{R}$
- $\mathbf{T} = [T_1,...,T_\kappa]^T : \mathbb{R}^{\kappa} \to \mathbb{R}$

such that the pdf $q_\theta(y)$ may be written as:

$$q_\theta(y) = h(y)\exp[\langle \eta(\theta), \mathbf{T}(y)\rangle - B(\theta)] \quad with \quad y \in \chi \subset \mathbb{R}\,. \tag{5}$$

The term $\mathbf{T}$ is called the natural sufficient statistic and η the natural parameter vector. The term $\langle \eta, \mathbf{T}\rangle$ denotes the scalar product.

Letting the model be indexed by the natural parameter η rather that θ, the canonical $\kappa-$parameter exponential family generated by $\mathbf{T}$ and h is defined as follows:

$$q_\eta(y) = h(y)\exp[\langle \eta(\theta), \mathbf{T}(y)\rangle - A(\eta)]\,, \tag{6}$$

with $A(\eta) = log\int_{-\infty}^{+\infty} h(y)\exp[\langle \eta(\theta), \mathbf{T}(y)\rangle]dy$. The natural parameter space is defined as $E = \{\eta \in \mathbb{R}^{\kappa}; -\infty < A(\eta) < +\infty\}$.

Some common distributions Table 1 provides a synthetic description of some common distributions of the exponential family:

Distribution	θ^T	$\eta(\theta)^T$	$\mathbf{T}(y)^T$	$A(\eta)$
Normal	(μ, σ^2)	$(\frac{\mu}{\sigma^2}, \frac{-1}{2\sigma^2})$	(y, y^2)	$\frac{1}{2}\left(-\frac{\eta_1^2}{2\eta_2} - \log\frac{-\eta_2}{\pi}\right)$
Gamma	(λ, p)	$(-\lambda, p-1)$	$(y, \log y)$	$-(\eta_2+1)\log -\eta_1 + \log\Gamma(\eta_2+1)$
Beta	(r, s)	$(r-1, s-1)$	$(\log y, \log(1-y))$	$-\log B(\eta_1+1, \eta_2+1)$
Poisson	μ	$\log\mu$	y	e^{η}
Exponential	λ	$-\lambda$	y	$-\log -\eta$
Rayleigh	θ^2	$-1/2\theta^2$	y^2	$-\log -2\eta$

Table 1. Some common canonical exponential families. $B(\alpha, \beta)$ is the Euler Beta function.

Properties The following results will be useful for our RBAC scheme based on the exponential family. Their proofs may be found in [16]. These properties give us a relation between the parameters η and the domain Ω through the use of the expectation of the natural statistics $\mathbf{T}(Y)$. The first theorem provides general relations between the gradient of A and the expectation of $\mathbf{T}(Y)$ while the second theorem allows us to establish a $1-1$ correspondence between η and $\mathbb{E}[\mathbf{T}(Y)]$ (for the full rank exponential family). Such a relation may then be used to express the parameter η and derive it according to the domain.

Theorem 1. *Let $\{q_\eta : \eta \in E\}$ a κ-parameter canonical exponential family with natural sufficient statistic $\mathbf{T}(Y)$ and open natural parameter space E, we then have the following properties:*

1. *E is convex.*
2. *$A : E \mapsto S \subseteq \mathbb{R}$ is convex.*
3. *$\mathbb{E}[\mathbf{T}(Y)] = \nabla A(\eta)$.*
4. *$\mathrm{Cov}[T(Y)] = \ddot{A}(\eta)$.*

where $\nabla A = (\frac{\partial A}{\partial \eta_1}, \frac{\partial A}{\partial \eta_2}, .., \frac{\partial A}{\partial \eta_\kappa})^T$ represents the gradient of A, and $\ddot{A}$ is the Hessian matrix of A with $\ddot{A}_{ij} = \frac{\partial^2 A}{\partial \eta_i \partial \eta_j}$.

The following theorem establishes the conditions of strict convexity of A, and then those for ∇A to be 1-1 on E. This is a very useful result for optimization (derivation) purposes:

Theorem 2. *Let $\{q_\eta : \eta \in E\}$ a full rank (i.e. $\mathrm{Cov}[T(Y)]$ is a positive-definite matrix) κ-parameter canonical exponential family with natural sufficient statistic $\mathbf{T}(Y)$ and open natural parameter space E, we have [16]:*

1. *$\eta \mapsto \nabla A(\eta)$ is 1-1 on E.*
2. *The family may be uniquely parameterized by $\mu(\eta) \equiv \mathbb{E}[\mathbf{T}(Y)] = \nabla A(\eta)$.*
3. *The anti-log-likelihood function is a strictly convex function of η on E.*

These results establish a 1-1 correspondence between η and $\mathbb{E}[\mathbf{T}(Y)]$ such that:

$$\mu = \nabla A(\eta) = \mathbb{E}[(\mathbf{T}(Y)] \Leftrightarrow E \ni \eta = \phi\left(\mathbb{E}[\mathbf{T}(Y)]\right), \tag{7}$$

holds uniquely with ∇A and ϕ continuous.

Estimation of the hyperparameters The relation 7 allows us to express the parameter η as a function of $\mathbb{E}[(\mathbf{T}(Y)]$. In order to estimate the parameters, we replace $E[\mathbf{T}(Y)]$ by the empirical estimate of the mean $\overline{\mathbf{T}(Y)}$. This corresponds to the MLE of the parameter. Indeed, the MLE of η corresponds to minimizing the anti-log-likelihood score (for independent and identically distributed (iid) data). By differentiation of the anti-log-likelihood according to η, we find $\nabla A(\eta_{MLE}) = \overline{T(Y)}$. Note however that in this case, this is the discrete sample mean. The following example illustrates this stating:

Example 1. When dealing with the Rayleigh distribution, we have $\eta = \frac{-1}{2\theta^2}$, $A(\eta) = -\log(-2\eta)$ and $T(y) = y^2$. By computing $A'(\eta) = \overline{\mathbf{T}(Y)}$, we find that $-\frac{1}{\eta} = \frac{1}{|\Omega|}\int_{\Omega} y(\mathbf{x})^2 d\mathbf{x}$, which corresponds to the MLE of the parameter θ^2 given by $\widehat{\theta}^2_{\mathrm{ML}} = \frac{1}{2|\Omega|}\int_{\Omega} y(\mathbf{x})^2 d\mathbf{x}$.

3.2 Shape derivative of the criterion

In this section, we propose to derive according to the domain the functional (1). The dependence of the functional with the domain is due to the estimation of the parameter η detailed above. In the sequel, for the sake of simplicity, we will invariably denote η for the natural parameter and its finite sample estimate over the domain (without a slight abuse of notation, this should be $\hat{\eta}$). We are now ready to state our main result:

Theorem 3. *The Gâteaux derivative, in the direction of* $\mathbf{V}$*, of the functional* (1)*, is:*

$$< D'(\Omega), \mathbf{V} >= \langle \nabla_{\mathbf{V}}\eta, \mathbf{C} \rangle \,, \tag{8}$$

where $\nabla_{\mathbf{V}}\eta = [< \nabla\eta'_1(\Omega), \mathbf{V} >, ..., < \nabla\eta'_\kappa(\Omega), \mathbf{V} >]$ *is the Gâteaux derivative of* η *in the direction of* $\mathbf{V}$*,* $\langle ., . \rangle$ *is the usual scalar product of two vectors and:*

$$\mathbf{C} = \mathbb{E}[\partial_1\Psi(q(Y, \eta(\Omega)), p(Y))(\mathbf{T}(Y) - \mathbb{E}[\mathbf{T}(Y)])].$$

The term $\partial_1\Psi$ *denotes the partial derivative of* Ψ *according to the first variable.*

The proof is detailed in Appendix 5.2.

We then have to compute the shape derivative $\nabla_{\mathbf{V}}\eta$. Such a computation requires an estimation of the expectation $\mathbb{E}[\mathbf{T}(Y)]$ as explained in the next section.

3.3 Computing the shape derivative for the MLE estimator

As mentioned in section (3.1.3), the expectation $\mathbb{E}[\mathbf{T}(Y)]$ can be replaced with the empirical estimate of the mean $\overline{\mathbf{T}(Y)}$ which is computed over the considered domain Ω. Using such an estimation for the hyperparameter, we can state the following proposition :

Lemma 1. *Within the full rank exponential family, and using the MLE estimator for the hyperparameters, the shape derivative* $\nabla_{\mathbf{V}}\eta$ *can be expressed as:*

$$\nabla_{\mathbf{V}}\eta = \ddot{A}(\eta)^{-1}\nabla_{\mathbf{V}}(\overline{\mathbf{T}}) \,. \tag{9}$$

where $\ddot{A}(\eta)^{-1} = I(\eta)^{-1}$ *is the inverse of the Hessian matrix of* $\ddot{A}$ *which is also the fisher information matrix I. The derivative* $\nabla_{\mathbf{V}}(\overline{\mathbf{T}})$ *is given by:*

$$\nabla_{\mathbf{V}}(\overline{\mathbf{T}}) = \frac{1}{|\Omega|}\int_{\partial\Omega}\left(\overline{\mathbf{T}(y)} - \mathbf{T}(y(\mathbf{a}))\right)(\mathbf{V}\cdot\mathbf{N})d\mathbf{a}(\mathbf{x}) \,, \tag{10}$$

The proof is given in Appendix 5.3.
We can then replace the shape derivative of the natural parameters given in Lemma 1 in the general Theorem 3. The corollary that gives the shape derivative then follows:

Corollary 1. *The Gâteaux derivative, in the direction of* $\mathbf{V}$*, of the functional* (1)*, is:*

$$< D'(\Omega), \mathbf{V} >= \frac{1}{|\Omega|} \int_{\partial\Omega} \left(\sum_{i=1}^{\kappa} C_i \sum_{j=1}^{\kappa} [\ddot{A}(\eta)]^{-1}_{ij} (\overline{T_j(y)} - T_j(y(\mathbf{a}))) \right) (\mathbf{V} \cdot \mathbf{N}) d\mathbf{a},$$

where the κ *components of the vector* $\mathbf{C}$ *are defined as follows:*

$$C_i = \mathbb{E}[\partial_1 \Psi(q(Y, \eta(\Omega)), p(Y))(T_i(Y) - \overline{T_i(Y)}] \quad i \in [1, \kappa].$$

The term $\partial_1 \Psi$ *denotes the partial derivative of* Ψ *according to the first variable.*

In order to fix ideas, the functional $D(\Omega)$ can be chosen as the Kullback-Leibler divergence, in this case $\partial_1 \Psi(q, p) = \log q + 1 - \log p - \frac{p}{q}$. In order to compute the vector $\mathbf{C}$ in Corollary 1, we can assume that the pdf p belongs to the exponential family and to the same parametric law as the pdf q. Let us denote by η_1 the parameter of the pdf p. This parameter is supposed to be already computed or dependent of another domain and so does not depend on the domain Ω. We then state the following proposition:

Lemma 2. *When* $p(y, \eta_1)$ *and* $q(y, \eta(\Omega))$ *are two members of the exponential family that belong to the same parametric law with respective parameters* η_1 *and* η*, and when the functional* $D(\Omega)$ *is chosen as the KL divergence, we find for the vector* $\mathbf{C}$ *defined in Theorem 1:*

$$\mathbf{C} = \ddot{A}(\eta)(\eta - \eta_1) + \nabla A(\eta) - \nabla A(\eta_1) \ .$$

A proof is given in appendix C.

This expression demonstrates that the derivative can be very simply computed using the natural parameters and the sufficient statistics of the law. Let us give two examples of computation for both the Rayleigh and the Gaussian law.

Example 2. When dealing with the Rayleigh distribution, following example 2, with $\theta^2 = \frac{1}{2}\overline{y^2}$, the term C is equal to $C = 2\theta^2 \left(\frac{\theta^2}{\theta_1^2} - \frac{\theta_1^2}{\theta^2} \right)$. we then find for the derivative of KL divergence:

$$< KL'(\Omega), \mathbf{V} >= \frac{1}{|\Omega|} \int_{\partial\Omega} \frac{C}{2\theta^2} (1 - \frac{y(\mathbf{a})^2}{2\theta^2}) (\mathbf{V} \cdot \mathbf{N}) d\mathbf{a}(\mathbf{x}) \ . \tag{11}$$

Example 3. When dealing with the Gaussian distribution, the term C is equal to

$$C = \begin{bmatrix} (\frac{\sigma^2}{\sigma_r^2} + 1)(\mu - \mu_1) \\ \mu^2 + \sigma^2 - \mu_1^2 - \sigma_1^2 + 2\frac{\sigma^2}{\sigma_1^2} \left(\mu^2 - \mu\mu_1 \right) + \sigma^4 (\frac{1}{\sigma_1^2} - \frac{1}{\sigma^2}) \ . \end{bmatrix} \tag{12}$$

We then find for the derivative of KL divergence:

$$< KL'(\Omega), \mathbf{V} >= \\ \frac{1}{\sigma^2 |\Omega|} \int_{\partial\Omega} -(y - \mu) \left(C_1 (1 + \frac{2\mu}{\sigma^2}) - C_2 \frac{\mu}{\sigma^2} \right) + (y^2 - \sigma^2 - \mu^2) \left(C_1 \frac{\mu}{\sigma^2} - \frac{C_2}{2\sigma^2} \right) (\mathbf{V} \cdot \mathbf{N}) d\mathbf{a}(\mathbf{x}) \ .$$

4 Maximisation of divergences

In this section, we propose to concentrate on the segmentation of an image into two regions (namely Ω and its complement Ω^c) by maximizing the criterion 2.

4.1 Evolution equation

When using the MLE estimator for the parameters, and noting that Ω and Ω^c shares the same boundary with opposite normals, we take $\overline{\mathbf{T}(y)} = \frac{1}{|\Omega|}\int_\Omega \mathbf{T}(y(\mathbf{x}))d\mathbf{x}$ and $\overline{\mathbf{T}(y)}^c = \frac{1}{|\Omega^c|}\int_{\Omega^c}\mathbf{T}(y(\mathbf{x}))d\mathbf{x}$. Using Corollary 1 and the fact that $< D'(\Omega,\Omega^c),\mathbf{V} >= \langle \nabla_{\mathbf{V}}\eta,\mathbf{C}\rangle + \langle \nabla_{\mathbf{V}}\eta^c,\mathbf{C}^c\rangle$, we find for the evolution equation:

$$\frac{\partial\Gamma}{\partial\tau} = \Big[\frac{1}{|\Omega|}\sum_{i=1}^{\kappa}C_i(\Omega)\sum_{j=1}^{\kappa}\ddot{A}(\eta)^{-1}_{ij}(\overline{T_j(y)} - T_j(y(\mathbf{x})) - \frac{1}{|\Omega^c|}\sum_{i=1}^{\kappa}C_i(\Omega^c)\sum_{j=1}^{\kappa}\ddot{A}(\eta_c)^{-1}_{ij}(\overline{T_j(y)}^c - T_j(y(\mathbf{x}))\Big]\mathbf{N}.$$

For the KL divergence, the term C is evaluated as explained in section 3.3. A classical regularization term $\lambda\kappa$ is added where λ is a positive constant and κ the curvature. As far as the numerical implementation is concerned, we use the level set method approach first proposed by Osher and Sethian [18].

4.2 Comparison with other methods in the Gaussian case

In this section, we propose to compare the behavior of our data term based on the maximization of the symmetrized Kullback-Leibler divergence between parametric pdfs to two other well-known region-based methods [7, 14]. The first method is the famous Chan & Vese method [14]. Such a criterion implies a Gaussian distribution for the feature y with a fixed variance. The corresponding evolution equation can be found in [14]. The second method has been first proposed by [7] and aims at minimizing the anti-log-likelihood for a Gaussian distribution. The evolution equation can be found in [7].

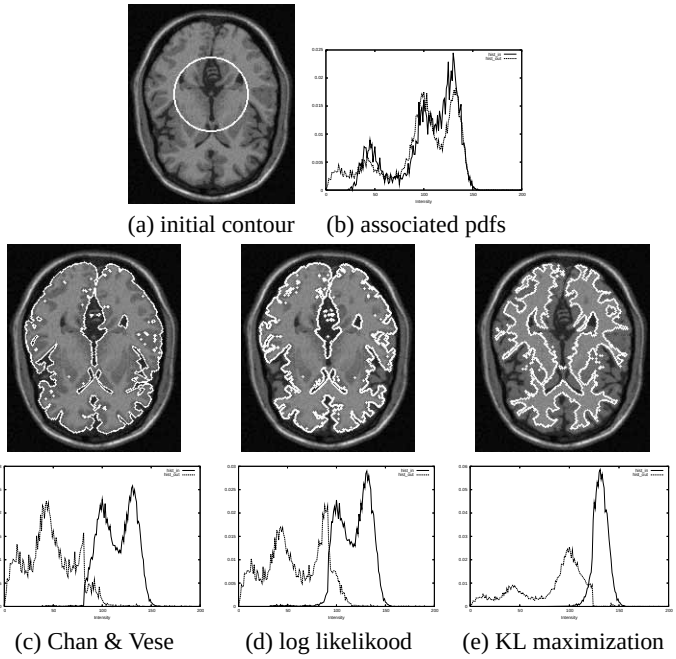
In order to compare these terms, let us express the non symmetrized KL divergence using the expectation under the pdf q, denoted by $\mathbb{E}_q$, as follows:

$$D(q\|p) = \mathbb{E}_q[\log(q(Y,\eta_\Omega))] - \mathbb{E}_q[\log(p(Y,\eta_{\Omega^c}))] \tag{13}$$

To get the gist of using KLD as a criterion in an RBAC functional, consider the data $y_i = \{y(\mathbf{x})|\mathbf{x}\in\Omega\}$ as an iid sequence from the statistical model $q(y,\eta_\Omega)$. Using the weak law of large number for a very large domain Ω, the first term (which corresponds to the entropy) can then be expressed as $\frac{1}{|\Omega|}\int_\Omega \log(q(y(\mathbf{x}),\eta(\Omega))d\mathbf{x}$. Maximizing the first term in KL divergence can then be seen as equivalent to minimizing the anti-log-likelihood score [19] divided by the size of the sample (which corresponds to the entropy under the law of large number). Using the same assumptions, the second term of KL divergence can be seen as the minimization of the plausibility of the data provided by Ω^c in the inside region Ω. When using the symmetrized version, we act both on Ω and Ω^c.

Let us now compare experimentally the behavior of these criterions for the extraction of an homogeneous region corrupted by a Gaussian noise in an image. We propose to take the example of the segmentation of the White Matter (WM) in T1-weighted

brain MRI images. We perform the three evolution equations using the Gaussian assumption for the pdf of the feature y within each region. The feature y is chosen as the Intensity of the image. The initial contour is given in Figure 1.(a) and we also show the two initial pdfs (b), namely $q_\eta(I,\Omega)$ which corresponds to the distribution of the intensity I inside the region Ω and $q_{\eta^c}(I,\Omega^c)$ which corresponds to the distribution of I inside the region Ω^c (i.e. outside the region). In Figure 1, we can observe the final active contour obtained using our criterion (22) and the two other criterions mentioned above. We can remark that our criterion acts as an extractor of the most important Gaussian in the initial mixture of Gaussian (see Figure 1.e). The two other criterions separate the mixture without extracting a single Gaussian. So, with our method, we can directly obtain the White Matter of the brain without a multiphase scheme.



(a) initial contour (b) associated pdfs

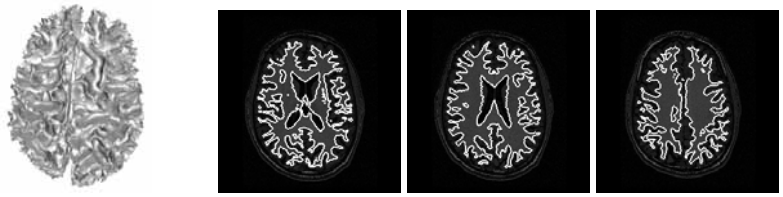
(c) Chan & Vese (d) log likelikood (e) KL maximization

Fig. 1. T1-weighted brain MRI segmentation results (extraction of the White Matter). The pdf of the intensity inside the contour is in solid line, the pdf of the intensity outside the contour is in dotted lines. (a): initial contour and (b) : associated pdfs, column (c): final contour and pdfs for the Chan & Vese method [14], column (d): for the log-likelihood method [7], column (e): for the maximization of the KL divergence.

4.3 Examples of applications

In this part, we consider two examples of application (brain MRI images and contrast echocardiogaphy) using two different noise models (Gaussian and Rayleigh).

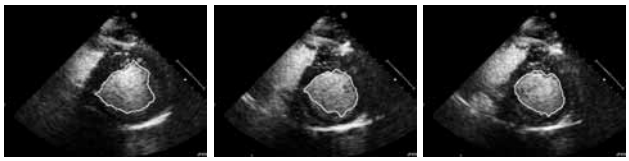
Concerning 3D T1-weighted MRI images of the brain, the noise model is assumed to be represented by a Rician distribution [20]. For large signal intensities the noise distribution can be considered as a Gaussian distribution (this is the case for the White Matter (WM) or the Gray Matter (GM)). We propose in Figure 2 an example of WM segmentation by maximizing the KL divergence between Gaussian distributions. When evaluating quantitatively our results of WM segmentation on the simulated brain T1-weighted MRI images provided by the Montreal Neurological Institute Brain Web URL, we find a dice coefficient of 0.91, a very law False Positive Fraction (FPF) of 0.8% and a True Positive Fraction (TPF) of 84%.



(a) 3D rendering of the WM (b) slice 72 (c) slice 75 (d) slice 84

Fig. 2. 3D Segmentation of WM in a T1 brain MRI using KL maximization

As the Rayleigh distribution is well suited to model noise in echography [20], this noise model was applied for segmentation of the left ventricle in contrast echocardiography. Final contours for several images of the sequence are shown in Figure 3. The segmentation is accurate all along the sequence. Note that experimental results reported in [21, 4] prove that when using the appropriate noise model, segmentation results are more accurate and less sensitive to the choice of the regularization parameters.



frame 1 frame 31 frame 40

Fig. 3. Segmentation of the LV in a contrast echocardiographic sequence

5 Appendix

5.1 Shape derivation tools

Let us remind this useful theorem [15] that will be used in the following proofs.

Theorem 4. *The Gâteaux derivative of the functional* $J(\Omega) = \int_{\Omega} f(\mathbf{x},\Omega)\, d\mathbf{x}$ *in the direction of* $\mathbf{V}$ *is:* $< J'(\Omega), \mathbf{V} >= \int_{\Omega} f_s(\mathbf{x},\Omega,\mathbf{V}) d\mathbf{x} - \int_{\partial\Omega} f(\mathbf{x},\Omega)(\mathbf{V}\cdot N) d\mathbf{a}(\mathbf{x})$ *where* N *is the unit inward normal to* $\partial\Omega$, $d\mathbf{a}$ *its area element and* f_s *the shape derivative of* f *[15].*

5.2 Proof of Theorem 3

To compute $< D'(\Omega), \mathbf{V} >$, we must first get the derivative of $q(y(\mathbf{x}),\eta)$ with respect to the domain, and apply the chain rule to $\Psi(q(y(\mathbf{x}),\eta), p(y))$. To simplify the notation we write the Eulerian derivative of η as $< \eta'(\Omega), \mathbf{V} >= \nabla_{\mathbf{V}}\eta = \left[< \eta'_1(\Omega), \mathbf{V} >, .., < \eta'_\kappa(\Omega), \mathbf{V} >\right]^T$.
Using the definition of $q(y,\eta)$ given in (6) and the chain rule applied to $A(\eta(\Omega))$, we obtain:

$$\begin{aligned} < q'(y,\eta), \mathbf{V} > &= h(y)\left(\langle \nabla_{\mathbf{V}}\eta, \mathbf{T}(y)\rangle - \langle \nabla_{\mathbf{V}}\eta, \nabla A(\eta)\rangle\right) e^{\langle \eta(\Omega), \mathbf{T}(y)\rangle - A(\eta(\Omega))} \\ &= q(y,\eta)\langle \nabla_{\mathbf{V}}\eta, \mathbf{T}(y) - \nabla A(\eta)\rangle \ . \end{aligned} \tag{14}$$

By the chain rule applied to $\Psi(q(y(\mathbf{x}),\eta), p(y))$, we get $< \Psi'(q(y,\eta), p(y)), \mathbf{V} >=< q'(y,\eta), \mathbf{V} > \partial_1\Psi(q,p)$, which gives $< D'(\Omega), \mathbf{V} >= \int_\chi q(y,\eta)\partial_1\Psi(q,p)\langle \nabla_{\mathbf{V}}\eta, \mathbf{T}(y) - \nabla A(\eta)\rangle dy$. We introduce $\mathbf{C} = \int_\chi q(y,\eta)\partial_1\Psi(q,p)\left(\mathbf{T}(y) - \nabla A(\eta)\right) dy = \mathbb{E}[\partial_1\Psi(q,p)\left(\mathbf{T}(Y) - \mathbb{E}[\mathbf{T}(Y)]\right)]$ which completes the proof.

5.3 Proof of Lemma1

When using the MLE, the term $\mathbb{E}[\mathbf{T}(Y)]$ can be empirically estimated with $\overline{\mathbf{T}(Y)}$ and so derived easily with respect to the domain Ω. We propose to directly derive the expression $\nabla A(\eta) = \overline{\mathbf{T}(Y)}$ which gives:

$$\sum_{j=1}^{\kappa} < \eta'_j, \mathbf{V} > \frac{\partial^2 A}{\partial\eta_i\partial\eta_j}(\eta) =< \overline{T_i(Y)}', \mathbf{V} > \quad \forall i \in [1,\kappa] \ , \tag{15}$$

which can be written in the compact form $\nabla_{\mathbf{V}}(\overline{\mathbf{T}}) = \ddot{A}(\eta)\nabla_{\mathbf{V}}\eta$.
Restricting our study to the full rank exponential family, where $\ddot{A}(\eta)$ is a symmetric positive-definite, hence invertible, matrix (Theorem 2), the domain derivative of the parameters η is uniquely determined by $\ddot{A}(\eta)^{-1}\nabla_{\mathbf{V}}(\overline{\mathbf{T}}) = \nabla_{\mathbf{V}}\eta$ where $\nabla_{\mathbf{V}}(\overline{\mathbf{T}})$ is given by:
$\nabla_{\mathbf{V}}(\overline{\mathbf{T}}) = \frac{1}{|\Omega|}\int_{\partial\Omega}\left(\overline{\mathbf{T}(y)} - \mathbf{T}(y(\mathbf{a}))\right)(\mathbf{V}\cdot\mathbf{N}) d\mathbf{a}(\mathbf{x})$ (taking benefit of theorem 4) and the lemma follows.

5.4 Proof of Lemma 2

Since p and q belongs to the same parametric law, they share the same value for $h(y)$, $\mathbf{T}(y)$ and $A(\eta)$ and then $log(q) - log(p) = \langle \eta - \eta_1, \mathbf{T}(y)\rangle - A(\eta) + A(\eta_1)$. The value of $\mathbf{C}$ is then $\mathbf{C} = s_1 - s_2$, with:

$$\begin{aligned} s_1 &= \mathbb{E}[(\langle \eta - \eta_1, \mathbf{T}(y)\rangle - A(\eta) + A(\eta_1) + 1)(T_i(Y) - \mathbb{E}[T_i(Y)]] \\ s_2 &= \mathbb{E}[\frac{p}{q}(T_i(Y) - \mathbb{E}[T_i(Y)]]\mathbb{E}_p[(T_i(Y) - \mathbb{E}[T_i(Y)]] \end{aligned}$$

Developing the expression of the expectation of the second term,we find $s_2 = \mathbb{E}_p[(T_i(Y) - \mathbb{E}[T_i(Y)]] = \nabla A(\eta_1) - \nabla A(\eta)$. Using the linearity of the expectation and the fact that
$E[T_j(Y)(T_i(Y)] - \mathbb{E}[T_i(Y)]\mathbb{E}[T_j(Y)]$ designates the covariance matrix of the sufficient statistics $\mathbf{T}$ and can then be replaced by $\ddot{A}(\eta)_{ij} = Cov[\mathbf{T}(Y)]_{ij} = \ddot{A}(\eta)_{ji}$, we find: $s_1 = \sum_{j=1}^{\kappa}(\eta_j - \eta_{1_j})\ddot{A}(\eta)_{ij}$, and then $\mathbf{C} = \ddot{A}(\eta)(\eta - \eta_1) + \nabla A(\eta) - \nabla A(\eta_1)$.

References

1. Kass, M., Witkin, A., Terzopoulos, D.: Snakes : Active contour models. International Journal of Computer Vision **1** (1988) 321–332
2. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. International Journal of Computer Vision **22**(1) (1997) 61–79
3. Aujol, J.F., Aubert, G., Blanc-Féraud, L.: Wavelet-based level set evolution for classification of textured images. IEEE Transactions on Image Processing **12**(12) (2003) 1634–1641
4. Martin, P., Réfrégier, P., Goudail, F., Guérault, F.: Influence of the noise model on level set active contour segmentation. IEEE PAMI **26** (2004) 799–803
5. Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. International Journal of Computer Vision **72**(2) (April 2007) 195–215
6. Aubert, G., Barlaud, M., Faugeras, O., Jehan-Besson, S.: Image segmentation using active contours: Calculus of variations or shape gradients ? SIAM Applied Mathematics **63**(6) (2003) 2128–2154
7. Zhu, S., Yuille, A.: Region competition: unifying snakes, region growing, and bayes/MDL for multiband image segmentation. IEEE PAMI **18** (September 1996) 884–900
8. Paragios, N., Deriche, R.: Geodesic active regions: A new paradigm to deal with frame partition problems in computer vision. JVCIR **13** (2002) 249–268
9. Rougon, N., Discher, A., Prêteux, F.: Region-based statistical segmentation using informational active contours. In: SPIE Conf.on Mathematics of Data/Image Pattern Recognition, San Diego, CA. (August 2006)
10. Michailovich, O., Rathi, Y., Tannenbaum, A.: Image segmentation using active contours driven by the bhattacharyya gradient flow. IEEE Transactions on Image Processing **16**(November) (2007) 2787 – 2801
11. Wang, Z., Vemuri, B.: DTI segmentation using an information theoretic tensor dissimilarity measure. IEEE Transactions on Medical Imaging **24**(10) (October 2005) 1267–1277
12. Jehan-Besson, S., Barlaud, M., Aubert, G.: DREAM^{2}S: Deformable regions driven by an eulerian accurate minimization method for image and video segmentation. International Journal of Computer Vision (53) (2003) 45–70
13. Kullback, S.: Information Theory and Statistics. New York: Wiley (1959)
14. Chan, T.F., Vese, L.A.: Active contour without edges. IEEE Transactions on Image Processing **10** (2001) 266–277
15. Delfour, M., Zolésio, J.: Shape and geometries. Advances in Design and Control, SIAM (2001)
16. Bickel, P., Docksum, K.: Mathematical statistics: basic ideas and selected topics. 2nd edn. Volume I. Prentice-Hall, London (2001) ISBN 013850363-X(v. 1).
17. Koopman, P.: On distributions admitting a sufficient statistic. Trans. Am. Math. Soc. **39** (1936) 399–409
18. Osher, S., Sethian, J.: Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulation. Journal of Computational Physics **79** (1988) 12–49
19. Cover, T.M., Thomas, J.A.: Elements of Information Theory. New York: Wiley (1991)
20. Goodman, J.: Some fundamental properties of speckle. J. of Optical Society of America **66** (1976) 1145–1150
21. Lecellier, F., Jehan-Besson, S., Fadili, J., Aubert, G., Revenu, M.: Statistical region-based active contours with exponential family observations. In: ICASSP. Volume 2. (2006) 113–116