Exploiting information geometry to improve the convergence of nonparametric active contours

Exploiting Information Geometry to Improve the Convergence of Nonparametric Active Contours

Marcelo Pereyra, *Member, IEEE*, Hadj Batatia, and Steve McLaughlin, *Fellow, IEEE*

***Abstract*— This paper presents a fast converging Riemannian steepest descent method for nonparametric statistical active contour models, with application to image segmentation. Unlike other fast algorithms, the proposed method is general and can be applied to any statistical active contour model from the exponential family, which comprises most of the models considered in the literature. This is achieved by first identifying the intrinsic statistical manifold associated with this class of active contours, and then constructing a steepest descent on that manifold. A key contribution of this paper is to derive a general and tractable closed-form analytic expression for the manifold's Riemannian metric tensor, which allows computing discrete gradient flows efficiently. The proposed methodology is demonstrated empirically and compared with other state of the art approaches on several standard test images, a phantom positron-emission-tomography scan and a B-mode echography of in-vivo human dermis.**

***Index Terms*— Active contours, level sets, variational methods on Riemannian manifolds, information geometry.**

I. Introduction

THIS paper addresses the problem of solving active contour (AC) models with application to image segmentation. More precisely, we seek to develop Riemannian optimisation algorithms that exploit the underlying geometrical structure of the contour's intrinsic manifold to achieve fast convergence. The motivation for this work arises from observations by Bar *et al.* [1] who highlighted two important open problems regarding the solution of active contours: "*the selection of the most appropriate inner product*" and how "*to incorporate non flat manifolds instead of Euclidean spaces*" as a means of overcoming the slow convergence of the Euclidean methods commonly used in the literature. In a recent paper [2], we addressed these problems for the specific case of the Chan-Vese active contour [3] and showed that the algorithm resulting from performing a steepest descent on the model's intrinsic manifold converges extremely fast. However, the method studied in [2] is not immediately amenable to other active contour models and can only be applied to images with Gaussian statistics. In this paper, we seek to provide a more general solution to these problems by identifying the intrinsic manifold and collection of inner products for the class of active contour models from the exponential family, which we subsequently use to derive a fast converging methodology that can be applied to most of the ACs studied in the literature [4].

Active contour models are a powerful framework for estimating the boundaries of an object within a given image. In this framework, contours are represented as curves that evolve subject to certain constraints to minimize an energy functional. This paper considers nonparametric ACs, a particularly useful class of segmentation methods where curves are represented implicitly as the zero level set of a surface that evolves with a fictitious time t [5]. In particular, we focus on region-based ACs that evolve according to the statistical characteristics of the object of interest and the background, as opposed to evolving according to image gradients or edges. We emphasise at this point that there are many state-of-the-art approaches to image segmentation that do not use AC models (see [6]–[10] for examples based on Potts-Markov random fields, fuzzy region competition and convex models).

Region-based nonparametric ACs were first postulated in the seminal work of Chan and Vese [3], which defined an active contour for images composed by a foreground and background with Gaussian statistics. That work was subsequently generalised to images with other specific statistics, such as Rayleigh [11], gamma [12], Weibull [13] and Laplace [10]. A unified framework for AC models for distribution from the exponential family was finally proposed in [4].

Inherent in the active contour segmentation problem is the solution of the Euler-Lagrange differential equations that guide the contour's evolution (for a detailed analysis of contour evolution equations see [14]). In most applications these are solved using standard first-order Euler methods which are relatively simple to derive and implement. However, it is well known that Euler's method can take a large number of iterations to converge. This drawback has recently motivated numerous papers in the literature that study more advanced algorithms to solve active contour (a detailed survey of the state-of-the-art up to 2012 was presented in [2]). Most state-of-the-art algorithms are steepest descent methods on alternative spaces or manifolds whose inner-products induce favourable properties on gradient flows, such as fast convergence and regularity. However, these algorithms have been designed for Gaussian region-based

Manuscript received December 16, 2013; accepted November 24, 2014. Date of publication December 18, 2014; date of current version January 23, 2015. This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) through the SuSTaIN Program under Grant EP/D063485/1, in part by the Department of Mathematics, University of Bristol, Bristol, U.K., in part by a Post-Doctoral Fellowship through the French Ministry of Defence, and in part by the EPSRC under Grant EP/J015180/1. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Olivier Bernard.

M. Pereyra is with the School of Mathematics, University of Bristol, Bristol BS8 1TH, U.K. (e-mail: marcelopereyra@ieee.org).

H. Batatia is with the University of Toulouse, Toulouse 31071, France (e-mail: hadj.batatia@enseeiht.fr).

S. McLaughlin is with the Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, U.K. (e-mail: s.mclaughlin@hw.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2014.2383318

(i.e. Chan-Vese type) active contours and cannot be directly applied to other contours from the exponential family.

This paper presents a general Riemannian optimisation method for nonparametric ACs from the exponential family and is structured as follows. In the next section we introduce the mathematical concepts underlying region-based active contour segmentation and briefly highlight some of the difficulties that our approach seeks to address. In Section III we develop an information geometry framework for active contours from the exponential family and propose a Riemannian steepest descent method to compute the contour's evolution. More precisely, we derive a smooth natural gradient descent algorithm [15] for nonparametric active contour, with application to image segmentation. Following on from this in Section IV we present some results illustrating the power of the approach and contrast the performance on alternative existing approaches. Finally we draw some conclusions and discuss potential extensions of the approach. A MATLAB implementation of the proposed method is available at http://www.stats.bris.ac.uk/~mp12320/code/SmoothNaturalGradient.zip.

II. Region-Based Level Set Segmentation

Let Ω be a bounded subset of $\mathbb{R}^D$ and $y : \Omega \to \mathbb{R}^p$ a D-dimensional image composed by p channels (i.e., pixels take their values in $\mathbb{R}^p$). Assume that y is constituted by a foreground Ω_1 and a background $\Omega_2 = \Omega/\Omega_1$, each characterised by its own statistical distribution. Precisely, assume that at a point $\boldsymbol{x} \in \mathbb{R}^D$ image values are distributed according to the statistical model

$$\begin{array}{ll} y(\boldsymbol{x}) \sim f(\cdot|\boldsymbol{\theta}_1) & \text{if } \boldsymbol{x} \in \Omega_F \\ y(\boldsymbol{x}) \sim f(\cdot|\boldsymbol{\theta}_2) & \text{if } \boldsymbol{x} \in \Omega_B \end{array} \tag{1}$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the statistical parameters associated with the foreground and background respectively (we assume that $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$), and where $f : \mathbb{R}^p \to \mathbb{R}^+$ denotes the probability density function of a generic distribution from the exponential family, i.e.,

$$f(\boldsymbol{s}|\boldsymbol{\theta}) = h(\boldsymbol{s}) \exp\left[\eta(\boldsymbol{\theta})^T S(\boldsymbol{s}) - A(\boldsymbol{\theta})\right], \tag{2}$$

for some sufficient statistic $S(\cdot)$, canonical parameter vector $\eta(\boldsymbol{\theta})$ and log-normalizer $A(\boldsymbol{\theta})$. Note that (2) defines a very general class of statistical models that comprises most distributions used in signal and image processing, such as the Gaussian, exponential, gamma, Poisson, Rayleigh, binomial, categorical, log-normal and Dirichlet.

Following an active contour approach, the segmentation of y is addressed by finding a curve $\hat{C} \subset \Omega$ that minimises the following energy functional [4]:

$$\begin{aligned} \hat{C} = \underset{C}{\operatorname{argmin}} & -\int_{\text{inside}(C)} \log f\left(y(\mathbf{x})|\boldsymbol{\theta}_1\right) \mathrm{d}\mathbf{x} \\ & -\int_{\text{outside}(C)} \log f\left(y(\mathbf{x})|\boldsymbol{\theta}_2\right) \mathrm{d}\mathbf{x} \end{aligned} \tag{3}$$

derived from the anti-loglikelihood of (1). In a manner akin to [1], we present our analysis in the case where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are known a-priori. However, the proposed approach is general and can be applied in the more general setup where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are unknown by alternatively minimising (3) with respect to C, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. The joint minimisation of (3) with respect to $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ for fixed C is generally straightforward via the use of maximum likelihood estimators.

This paper considers nonparametric contours where the curve $C \in \Omega$ is defined implicitly as the zero level set of a Lipschitz function $\phi : \Omega \to \mathbb{R}$, such that

$$C = \{\boldsymbol{x} : \phi(\boldsymbol{x}) = 0\} \tag{4}$$

$$\text{inside}(C) = \{\boldsymbol{x} : \phi(\boldsymbol{x}) > 0\} \tag{5}$$

$$\text{outside}(C) = \{\boldsymbol{x} : \phi(\boldsymbol{x}) < 0\} \tag{6}$$

and the energy minimization problem (3) is restated as follows

$$\hat{\phi} = \underset{\phi}{\operatorname{argmin}}\, E(y;\phi) \tag{7}$$

where

$$\begin{aligned} E(y;\phi) \triangleq & -\int_{\Omega} \log f\left[y(\mathbf{x})|\boldsymbol{\theta}_1\right] H[\phi(\boldsymbol{x})] d\boldsymbol{x} \\ & -\int_{\Omega} \log f\left[y(\mathbf{x})|\boldsymbol{\theta}_2\right] H[-\phi(\boldsymbol{x})] d\boldsymbol{x} \end{aligned} \tag{8}$$

and where $H(\cdot)$ denotes the Heaviside function. The functional optimisation problem (7) can be solved by introducing a fictitious time t and solving the associated Euler-Lagrange differential equations $\partial_t \phi = -\partial_\phi E$, which lead to the following flow for ϕ

$$\partial_t \phi(\boldsymbol{x}) = \delta\left[\phi(\boldsymbol{x})\right] \left\{\log f\left[y(\mathbf{x})|\boldsymbol{\theta}_1\right] - \log f\left[y(\mathbf{x})|\boldsymbol{\theta}_2\right]\right\} \tag{9}$$

where $\delta(\cdot)$ is the Dirac delta function and $\partial_\phi E$ the 1st variation of E with respect to ϕ.

In practice, equations (8) and (9) must be computed over a discrete space-time grid and using sampled functions $\boldsymbol{y} = \left(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\right)$ and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_N)$. Equation (8) becomes

$$E(\boldsymbol{y};\boldsymbol{\phi}) = \sum_{i=1}^{N} -H_\epsilon(\phi_i) \log f\left(\boldsymbol{y}_i|\boldsymbol{\theta}_1\right) - H_\epsilon(-\phi_i) \log f(\boldsymbol{y}_i|\boldsymbol{\theta}_2). \tag{10}$$

Similarly, equation (9) is now a discrete flow

$$\boldsymbol{\phi}^{t+1} = \boldsymbol{\phi}^t - \gamma_t \nabla_\phi E(\boldsymbol{y};\boldsymbol{\phi}^t) \tag{11}$$

with

$$\left(\nabla_\phi E(\boldsymbol{y};\boldsymbol{\phi}^t)\right)_i = -\delta_\varepsilon(\phi_i^t)\left(\log f\left(y_i|\boldsymbol{\theta}_1\right) - \log f\left(y_i|\boldsymbol{\theta}_2\right)\right) \tag{12}$$

and where γ_t controls the length of the step. For numerical stability, AC methods use smooth approximation of the Heaviside function $H(\cdot)$ and of Dirac's delta function $\delta(\cdot)$ [3]

$$\begin{aligned} H_\varepsilon(u) &= \frac{1}{2}\left(1 + \frac{2}{\pi}\arctan\frac{u}{\varepsilon}\right), \\ \delta_\varepsilon(u) &= \frac{1}{\pi}\left(1 + \frac{\varepsilon}{\varepsilon^2 + u^2}\right). \end{aligned} \tag{13}$$

Finally, note that there are several approaches to selecting the step size γ_t in (11). In most algorithms γ_t is set to some fixed value $\gamma_t = \gamma$ or $\gamma_t = \gamma / ||\nabla_\phi E(\boldsymbol{y};\boldsymbol{\phi}^t)||$. Alternatively, in many state-of-the-art algorithms γ_t is set by means of a line search [1], [2].

As explained previously, the iterative algorithm (11) can suffer from very long transient regimes and take a large number of iterations to converge. The reason for this is that the gradient $\nabla_\phi E(\boldsymbol{y};\boldsymbol{\phi})$ used in (11) is often strongly anisotropic; that is, there is a significant difference in the order of magnitude of the partial derivatives $\frac{\partial}{\partial\phi_i}E(\boldsymbol{y};\boldsymbol{\phi})$. In such cases, using a preconditioned gradient $\boldsymbol{A}^{-1}(\boldsymbol{\phi})\nabla_\phi E(\boldsymbol{y};\boldsymbol{\phi})$, for some appropriately specified positive definite matrix $\boldsymbol{A}(\boldsymbol{\phi})$, can improve the convergence dramatically [1], [15].

Gradient preconditioning arises naturally by considering that $\boldsymbol{\phi}$ belongs to a non-Euclidean space (note that $\boldsymbol{\phi}$ may take values in $\mathbb{R}^n$ yet belong to spaces with non-Euclidean geometry). From this perspective, the poor convergence properties of (11) and the anisotropy of $\nabla_\phi E(\boldsymbol{y};\boldsymbol{\phi})$ are a result of a mismatch between the underlying geometry of the space of $\boldsymbol{\phi}$ and the canonical Euclidean geometry that is implicitly assumed by using the gradient $\nabla_\phi E(\boldsymbol{y};\boldsymbol{\phi})$. This should be overcome by formulating (11) directly on the space of $\boldsymbol{\phi}$, with inner products given by $\langle\boldsymbol{\phi}', \boldsymbol{A}(\boldsymbol{\phi})\boldsymbol{\phi}\rangle$, thus we obtain

$$\boldsymbol{\phi}^{t+1} = \boldsymbol{\phi}^t - \gamma_t \boldsymbol{A}^{-1}(\boldsymbol{\phi}^t)\nabla_\phi E(\boldsymbol{y};\boldsymbol{\phi}^t). \tag{14}$$

Therefore the problem of specifying the preconditioning matrix $\boldsymbol{A}(\boldsymbol{\phi})$ to cure slow convergence of (11) is equivalent to identifying the geometry of the space of $\boldsymbol{\phi}$, which allows (14) to be solved with the collection of inner products of that space. This has motivated numerous papers in the literature that investigate specific instances of (14) for alternative spaces and manifolds with geometries and inner products that are more appropriate for $\boldsymbol{\phi}$ than $\mathbb{R}^n$. For instance, some papers use shape sensitivity calculus to solve (14) in a space of geometrical variables or shapes [16]–[18]. Many papers have proposed solving (14) in a Sobolev space because its inner product acts as a smoothing operator inducing favourable regularity properties on the contour [19]–[22]. Alternatively, other papers consider Newton-type methods that solve (14) on manifolds whose inner products are related the Hessian matrix of $E(\boldsymbol{y};\boldsymbol{\phi})$ [1], [16]–[18], [23], [24]. The empirical results reported as described above confirm that solving (14) in an appropriate space, with inner products that induce favourable properties to the gradient flow, can improve the convergence speed dramatically.

As explained previously, Bar *et al.* recently highlighted the important open problems of "*the selection of the most appropriate inner product associated with a particular functional*" and "*to incorporate non flat manifolds instead of Euclidean spaces*". In our recent paper [2] we addressed these two important questions for the specific case of the Chan-Vese AC [3]. Our main contributions in [2] were the observations that $\boldsymbol{y}$ and $\boldsymbol{\phi}$ are related by a parametric statistical model which induces a specific geometry to $\boldsymbol{\phi}$'s space, and to use information geometry to determine the collection of inner-products of that space, which we found to be a Riemannian statistical manifold. We then proposed a Riemannian steepest descent algorithm for the Chan-Vese model that solves (14) on $\boldsymbol{\phi}$'s manifold and conducted a series of experiments that confirm that this algorithm converges extremely fast.

However, the fast methods presented in the papers described above were designed for specific AC models and are not directly amenable to other active contours. As a result they can only be applied to images with certain specific statistics (e.g. Gaussian) and are unsuitable for many important types of data that are better analysed with other statistical models and active contours (e.g., ultrasound and SAR images [11], [12]). In this paper, we provide a more general solution to the problems of finding the most appropriate inner products and non flat manifolds by considering the information geometry of the entire class of active contour models from the exponential family [4], which includes the popular Chan-Vese contour as well as many other important AC models [10]–[13]. Based on this, we subsequently derive a general Riemannian steepest descent method for this class of active contours.

III. Proposed Optimisation Method

This section presents a general Riemannian steepest descent method that can be applied to any AC model of the exponential family. In a manner akin to [2], the method is derived by first using information geometry to obtain the statistical manifold associated with the active contour and then proposing a Riemannian steepest descent method to solve (14) on that manifold. Note that the method presented here includes the algorithm proposed in [2] for the Gaussian AC (i.e., Chan-Vese AC) as a particular case.

A. Proposed Riemannian Steepest Descent

We begin by considering the geometry of the space $\mathcal{M}$ whose points are the probability distributions $F(\boldsymbol{y}|\boldsymbol{\phi})$ with density

$$f(\boldsymbol{y}|\boldsymbol{\phi}) = \prod_{\{i:\phi_i>0\}} f(\boldsymbol{y}_i|\boldsymbol{\theta}_1) \prod_{\{i:\phi_i<0\}} f(\boldsymbol{y}_i|\boldsymbol{\theta}_2) \tag{15}$$

where the marginal densities $f(\boldsymbol{y}_i|\boldsymbol{\theta})$ belong to the exponential family defined in (2). We observe that the energy $E(\boldsymbol{y};\boldsymbol{\phi}) = -\log f(\boldsymbol{y}|\boldsymbol{\phi})$ and that algorithms to minimise it are effectively computing maximum likelihood estimators of $\boldsymbol{\phi}$ given $\boldsymbol{y}$. In other words, (14) is selecting the statistical model $f(\boldsymbol{y}|\boldsymbol{\phi})$ that fits the data best.

Space $\mathcal{M}$ is a statistical manifold parametrised by $\boldsymbol{\phi}$ [25]. More precisely, a point $\boldsymbol{\phi}$ on $\mathcal{M}$ represents the probability distribution of a random variable $\boldsymbol{Y}$ with distribution $F(\boldsymbol{y}|\boldsymbol{\phi}) = P[\boldsymbol{Y} \leq \boldsymbol{y}]$. Similarly, the distance between $\boldsymbol{\phi}$ and a second point $\boldsymbol{\phi}'$ is understood as the distance between the probability distributions $F(\boldsymbol{y}|\boldsymbol{\phi})$ and $F(\boldsymbol{y}|\boldsymbol{\phi}')$, measured by an appropriate metric on $\mathcal{M}$. According to information geometry $\mathcal{M}$ is endowed with a *natural* or *intrinsic* metric that allows measuring distances in a way that is independent of the parametrisation (i.e., to changes of variables $\boldsymbol{\psi} = h(\boldsymbol{\phi})$). $\mathcal{M}$ is endowed with a natural Riemannian metric tensor, or collection of inner products, specified by the Fisher information matrix (FIM) [25]

$$(\boldsymbol{G}(\boldsymbol{\phi}))_{(i,j)} \triangleq -\mathrm{E}_{\boldsymbol{Y}|\boldsymbol{\phi}}\left\{\frac{\partial^2}{\partial\phi_i\partial\phi_j}\log[f(\mathbf{Y}|\boldsymbol{\phi})]\right\} \tag{16}$$

where $\mathrm{E}_{\boldsymbol{Y}|\boldsymbol{\phi}}$ denotes the expectation operator with respect to $f(\boldsymbol{y}|\boldsymbol{\phi})$ and should not be confused with the energy term (10). Moreover, at each point $\boldsymbol{\phi}$, $\mathcal{M}$ is locally equivalent to an

Euclidean tangent space $\mathcal{T}_\phi$ with inner product $\langle \boldsymbol{\phi}', \boldsymbol{G}(\boldsymbol{\phi})\boldsymbol{\phi}\rangle$ varying smoothly with $\boldsymbol{\phi}$. Finally, a steepest descent on $\mathcal{M}$ to optimise E is given by

$$\boldsymbol{\phi}^{t+1} = \boldsymbol{\phi}^t - \gamma_t \boldsymbol{G}^{-1}\left(\boldsymbol{\phi}^t\right) \nabla_\phi E(\boldsymbol{y}; \boldsymbol{\phi}^t) \tag{17}$$

where $\boldsymbol{G}^{-1}\left(\boldsymbol{\phi}^t\right) \nabla_\phi E(\boldsymbol{y}; \boldsymbol{\phi}^t)$ is the gradient of E on $\mathcal{T}_{\boldsymbol{\phi}^t}$. This gradient represents the change of E with respect to the probability distribution $F(\boldsymbol{y}|\boldsymbol{\phi})$ indexed by $\boldsymbol{\phi}$, rather than with respect to the vector $\boldsymbol{\phi}$ itself.

Iteration (17) results in a so-called "natural gradient descent algorithm" in which preconditioning the Euclidean gradient with $\mathbf{G}^{-1}\left(\boldsymbol{\phi}^t\right)$ "*provides isotropic convergence properties about any local minima independent of the model parametrisation and of any dependencies within the signal being processed by the algorithm*" [15]. These algorithms generally exhibit very good convergence properties and have been been extensively applied to maximum likelihood estimation problems [15] and to the Chan-Vese AC in [2]. Finally, note that (17) is closely related to a Newton iteration where the preconditioning matrix is the Hessian matrix of $E(\boldsymbol{y}; \boldsymbol{\phi})$; indeed $\boldsymbol{G}(\boldsymbol{\phi})$ is the expectation of this Hessian matrix with respect to the random variable $\boldsymbol{Y}$, whose realisation is $\boldsymbol{y}$. We will see that this expectation introduces a key difference between (17) and a Newton iteration because it guarantees the positive definiteness of $\boldsymbol{G}(\boldsymbol{\phi})$ (the Hessian matrices of $E(\boldsymbol{y}; \boldsymbol{\phi})$ do not have this fundamental property).

B. Computation of $\boldsymbol{G}(\boldsymbol{\phi})$

We consider that a main contribution of this paper is to propose an efficient strategy to evaluate $\boldsymbol{G}(\boldsymbol{\phi})^{-1}$, which will allow using the natural gradient descent (17) to develop fast image segmentation algorithms. We demonstrate that $\boldsymbol{G}(\boldsymbol{\phi})$ is diagonal, thus trivial to invert, and that its elements admit closed-form expressions that are easy to evaluate.

Theorem 3.1: Suppose that (2) *holds. Then the Fisher information matrix* $\boldsymbol{G}(\boldsymbol{\phi})$ *defined in* (16) *is diagonal with elements*

$$(\boldsymbol{G}(\boldsymbol{\phi}))_{(i,j)} = \begin{cases} |\delta'(\phi_i)|\, B_f(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1) & \text{if } \phi_i \geq 0 \\ |\delta'(\phi_i)|\, B_f(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) & \text{if } \phi_i < 0 \end{cases} \tag{18}$$

if $i = j$ *and* $(\boldsymbol{G}(\boldsymbol{\phi}))_{(i,j)} = 0$ *otherwise, where* $B_f(\boldsymbol{\theta}^*||\boldsymbol{\theta})$ *is the* f*-Bregman divergence*

$$B_f(\boldsymbol{\theta}^*||\boldsymbol{\theta}) \triangleq A(\boldsymbol{\theta}^*) - A(\boldsymbol{\theta}) - \langle \eta(\boldsymbol{\theta}^*) - \eta(\boldsymbol{\theta}), \nabla_\eta A(\boldsymbol{\theta})\rangle \tag{19}$$

where $A(\cdot)$ *is the logarithm of the normalising constant of* f *and* $\nabla_\eta A(\cdot)$ *its gradient with respect to the canonical parameter vector* $\eta = \eta(\boldsymbol{\theta})$*. Moreover, if* $\delta'(u)$ *is evaluated using a regularised approximation* $\delta'_\varepsilon(u)$ *bounded away from zero, then* $\boldsymbol{G}(\boldsymbol{\phi})$ *is positive definite.*

Proof: To prove these results we develop the derivatives in (16) and obtain that $\mathbf{G}(\boldsymbol{\phi})$ is diagonal with elements

$$\begin{aligned}(\boldsymbol{G}(\boldsymbol{\phi}))_{(i,j)} = &-\delta'(\phi_i) E\left[\log f(\boldsymbol{y}_i; \boldsymbol{\theta}_1)\middle|\phi_i\right] \\ &-\delta'(-\phi_i) E\left[\log f(\boldsymbol{y}_i; \boldsymbol{\theta}_2)\middle|\phi_i\right]\end{aligned} \tag{20}$$

if $i = j$ and $(\boldsymbol{G}(\boldsymbol{\phi}))_{(i,j)} = 0$ otherwise, and where $E(\ldots|\phi_i)$ denotes the expectation with respect to the marginal likelihood

$$f(\boldsymbol{y}_i; \phi) = \begin{cases} f(\boldsymbol{y}_i; \boldsymbol{\theta}_1) & \text{if } \phi_i \geq 0 \\ f(\boldsymbol{y}_i; \boldsymbol{\theta}_2) & \text{if } \phi_i < 0. \end{cases}$$

TABLE I
BREGMAN DIVERGENCES I

	$f(s\|\theta)$	$\mathrm{B}_f(\theta_2\|\|\theta_1)$
Gamma	$\frac{s^{\alpha-1}}{\Gamma(\alpha)\theta^\alpha}\exp\left(-\frac{s}{\theta}\right)\mathbf{1}_{\mathcal{R}^+}(s)$	$\log\left(\frac{\Gamma(\theta_2)}{\theta_1}\right) + \alpha\frac{\theta_2-\theta_1}{\theta_1}$
Gaussian	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(s-\theta)^2}{2\sigma^2}\right)$	$\frac{(\theta_2-\theta_1)^2}{2\sigma^2}$
Laplace	$\frac{1}{2\theta}\exp\left(-\frac{\lvert s\rvert}{\theta}\right)$	$\log\left(\frac{\theta_2}{\theta_1}\right) + \frac{\theta_1-\theta_2}{\theta_2}$
Poisson	$\frac{\theta^s}{s!}\exp(-\theta)\mathbf{1}_{\mathcal{Z}^*}(s)$	$\theta_2 - \theta_1 - \theta_1\log\left(\frac{\theta_2}{\theta_1}\right)$
Rayleigh	$\frac{s}{2\theta^2}\exp\left(-\frac{s^2}{2\theta^2}\right)\mathbf{1}_{\mathcal{R}^+}(s)$	$2\log\left(\frac{\theta_2}{\theta_1}\right) + \frac{\theta_1^2-\theta_2^2}{\theta_2^2}$

TABLE II
BREGMAN DIVERGENCES II

	Multivariate Gaussian
$f(\boldsymbol{s}\|\boldsymbol{\theta})$	$\frac{1}{(2\pi)^{d/2}\lvert\Sigma\rvert^{d/2}}\exp\left(-\frac{(\boldsymbol{\theta}_2-\boldsymbol{\theta}_1)^T\Sigma^{-1}(\boldsymbol{\theta}_2-\boldsymbol{\theta}_1)}{2}\right)$
$\mathrm{B}_f(\boldsymbol{\theta}_2\|\|\boldsymbol{\theta}_1)$	$(\boldsymbol{\theta}_2-\boldsymbol{\theta}_1)^T\Sigma^{-1}(\boldsymbol{\theta}_2-\boldsymbol{\theta}_1)$

Using the fact that $\delta'(-x) = -\delta'(x)$ and rearranging equation (20) we obtain that for $i = j$

$$(\boldsymbol{G}(\boldsymbol{\phi}))_{(i,j)} = -\delta'(\phi_i) E\left[\log\left(\frac{f(\boldsymbol{y}_i, \boldsymbol{\theta}_1)}{f(\boldsymbol{y}_i, \boldsymbol{\theta}_2)}\right)\middle|\phi_i\right], \tag{21}$$

which we express in terms of Kullback-Leibler divergences [26], i.e.,

$$(\boldsymbol{G}(\boldsymbol{\phi}))_{(i,j)} = \begin{cases} |\delta'(\phi_i)|\, \mathrm{KL}_f(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) & \text{if } \phi_i \geq 0 \\ |\delta'(\phi_i)|\, \mathrm{KL}_f(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1) & \text{if } \phi_i < 0 \end{cases} \tag{22}$$

where

$$\mathrm{KL}_f(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2) \triangleq \int_{\mathbb{R}^p} \log\left(\frac{f(\boldsymbol{s}; \boldsymbol{\theta}_1)}{f(\boldsymbol{s}; \boldsymbol{\theta}_2)}\right) f(\boldsymbol{s}; \boldsymbol{\theta}_1)\mathrm{d}\boldsymbol{s}.$$

Finally, given that f belongs to the exponential family, we can also express (22) in terms of Bregman divergences that admit closed-form expressions [26]

$$(\boldsymbol{G}(\boldsymbol{\phi}))_{(i,j)} = \begin{cases} |\delta'(\phi_i)|\, \mathrm{B}_f(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1) & \text{if } \phi_i \geq 0 \\ |\delta'(\phi_i)|\, \mathrm{B}_f(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2)) & \text{if } \phi_i < 0. \end{cases}$$

The proof is concluded by noting that because $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ the terms $\mathrm{B}_f(\boldsymbol{\theta}_1||\boldsymbol{\theta}_2)$ and $\mathrm{B}_f(\boldsymbol{\theta}_2||\boldsymbol{\theta}_1)$ are strictly positive and therefore $\mathbf{G}(\boldsymbol{\phi})$ is a positive definite matrix for all $\boldsymbol{\phi} \in \mathbb{R}^n$. □

Theorem 3.1 states that for energy functionals of the form of (10), associated with distributions from the exponential family (2), the FIM $\boldsymbol{G}(\boldsymbol{\phi})$ is diagonal and has a closed-form expression that is trivial to compute and invert. This is very important from a computational point of view because it allows evaluating the gradient of E on the tangent space $\mathcal{T}_\phi$, i.e., $\boldsymbol{G}(\boldsymbol{\phi})^{-1}\nabla_\phi E(\boldsymbol{y}; \boldsymbol{\phi})$, making it possible to develop fast image segmentation algorithms based on the natural gradient (17). To evaluate (18) numerically we use the regularised approximation $\delta'_\varepsilon(u) = \frac{-2\varepsilon}{\pi}\operatorname{sign}(u)\max(u, \varepsilon)/(\varepsilon^2 + u^2)^2$. Tables I and II provide the Bregman divergences of the statistical distributions commonly used in active contour literature.

C. Regularisation of $\boldsymbol{\phi}$

Finally, we now need to regularise $\boldsymbol{\phi}$ so as to promote smooth solutions that are robust to noise. This regularisation improves significantly the segmentation results by introducing prior knowledge about image structure, i.e., pixels belonging

to the foreground and background are organised in spatial groups (as opposed to being randomly distributed across the image). The active contour literature discusses several approaches to regularise $\boldsymbol{\phi}$ to promote solutions with specific properties (e.g., smoothness, shapes, topologies, etc.). Conventional active contour algorithms typically enforce spatial regularity by introducing the penalty term $\lambda \int_\Omega |\delta[\boldsymbol{\phi}(\boldsymbol{x})]| d\boldsymbol{x}$, which selects contours with minimal length, or alternatively by using a convex approximation of this penalty that is easier to minimise [10], [27]. The parameter λ controls the amount of regularity that is enforced and is commonly set by cross-validation.

An increasingly popular alternative, that we will explore below, is to promote regularity by using an inner product that applies spatial smoothing to the gradient flow. As noted in [1], this approach is adopted implicitly by all methods that use Sobolev gradient flows [22], since Sobolev inner products act as smoothing operators, and explicitly by most other state-of-the-art AC methods that use Gaussian smoothing operators such as [1] and [2]. In this case the amount of regularity enforced depends on the bandwidth of the smoothing kernel. Here we follow this approach and regularise $\boldsymbol{\phi}$ by combining the inner product associated with $\mathbf{G}(\boldsymbol{\phi})$, which reflects the information geometry of the statistical model $f(\boldsymbol{y};\boldsymbol{\phi})$ relating $\boldsymbol{\phi}$ to the observed image $\boldsymbol{y}$, with a Sobolev-type inner product that incorporates prior knowledge by promoting smooth solutions. Precisely, we propose to use a "smooth natural gradient flow" defined as follows:

$$\boldsymbol{\phi}^{t+1} = \boldsymbol{\phi}^t + \gamma_t \mathbf{H}_\sigma \mathbf{G}^{-1}\left(\boldsymbol{\phi}^t\right) \nabla_\phi E_\varepsilon(\boldsymbol{y};\boldsymbol{\phi}^t) \tag{23}$$

where $\mathbf{H}_\sigma$ is a smoothing operator given by

$$\mathbf{H}_\sigma = \text{Toeplitz}(h_\sigma),$$

and where

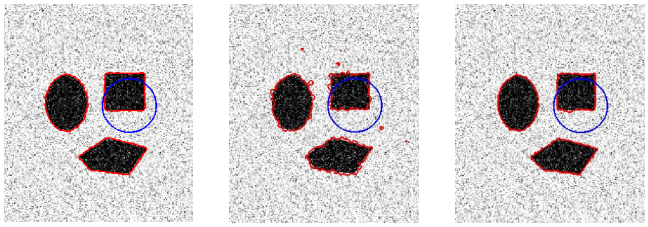
$$h_\sigma(s,u) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{s^2+u^2}{2\sigma^2}\right)$$

is a 2D Gaussian kernel of scale σ. The parameter σ determines the width of the Gaussian kernel and therefore the amount of smoothness enforced by $\mathbf{H}_\sigma$. The choice of this specific smoothing operator follows from the computational consideration that $\mathbf{H}_\sigma$ should be Toeplitz and separable, as this allows for particularly efficient algorithms to compute the product $\mathbf{H}_\sigma \mathbf{G}^{-1}\left(\boldsymbol{\phi}^t\right)$ [28]. In particular, the Gaussian smoothing operator is the only separable operator that is isotropic, i.e., that enforce the same regularization in all directions [29], which is important in many image segmentation applications. Note however that $h_\sigma(s,u)$ can be easily replaced by an anisotropic Gaussian kernels $h_{\sigma_s,\sigma_u}(s,u) = \frac{1}{2\pi\sigma_s\sigma_u}\exp\left(-\frac{s^2}{2\sigma_s^2}-\frac{u^2}{2\sigma_u^2}\right)$ if there is reliable prior knowledge about the shape and orientation of the image patterns. Note that the appropriate amount of regularisation required for a specific image will typically involve a trade-off between robustness to noise and preservation of sharp details in the contour. Images with low noise levels allow for low values of σ that preserve these details, whereas noisy images typically require more aggressive regularisations. More details regarding the choice of the smoothing operator and about the selection of σ can be found in [2].

In addition, we note that in some applications the performance of the proposed methods can be improved by introducing some form of localisation that takes into account the lack of stationarity in the image [30]. In our method this could be achieved by defining position dependent statistical parameters $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1(\boldsymbol{x})$ and $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2(\boldsymbol{x})$, which could then be estimated from the neighbourhood of $y(\boldsymbol{x})$ as in [30]. The resulting FIM would be equal (18) with pixel dependent Bregman divergences evaluated using the values of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ for that pixel; this matrix would still be diagonal and positive definite. Finally, it is also worth mentioning that although in this paper we apply (23) to all the elements of vector $\boldsymbol{\phi}$, one could also consider using a narrow band approach where only the elements of $\boldsymbol{\phi}$ that are close to zero are updated (i.e., the pixels within a neighbourhood of C), similarly to [16].

IV. Experimental Results and Observations

This section presents three experiments conducted to assess and compare the performance of the proposed smooth natural gradient method. In the first experiment we compare our method with four fast methods from the state of the art: two steepest descent algorithms (the generalised Newton method [1], and the Sobolev gradient [21]), and two optimisation algorithms based on convex approximation of $E(\boldsymbol{y};\boldsymbol{\phi})$ (the fast global minimisation algorithm (FGMA) [10], and the recent two-stage algorithm [27]). These methods are specifically designed for the Chan-Vese AC and we conducted the comparisons with standard test images that are well segmented by this model. The second and third experiments present applications of our method to two medical problems that motivated this work: in-vivo human dermis segmentation in high-frequency ultrasound images, and high-radioactivity region segmentation in Positron-Emitted-Tomography (PET) images. For those experiments we implemented our method using Rayleigh AC and Poisson AC models, and compared its performance with the Eulerian algorithm presented in [11] for Rayleigh ACs and with the supervised thresholding method proposed by Schaefer *et al.* in [31] for PET images. An observation that we would make prior to presenting results is that most of the real images used in the experiments do not have a ground truth and therefore it is not always completely clear what is the best segmentation result nor how a precise segmentation result could be objectively specified (though for the PET phantom experiment the sizes of the radioactive regions is known and is used as the ground truth). We present what we consider are fair and reasonable comparisons, where in most cases the segmentation achieved with our method is very similar to that obtained with state-of-the-art approaches but at a significantly faster computational speed. Finally, all experiments have been computed on an Intel i7 quad-core workstation running MATLAB 2013a and by using the values of the algorithm parameters that yield the best segmentation results.



(a) (b) (c)

Fig. 1. Comparison with the Sobolev gradient [21] and Generalized Newton [1] methods using the synthetic shape image from [1] (216×187 pixels, additive Gaussian noise, SNR 5.36 dB). (a) Proposed (23). (b) Sobolev [21]. (c) Newton [1].

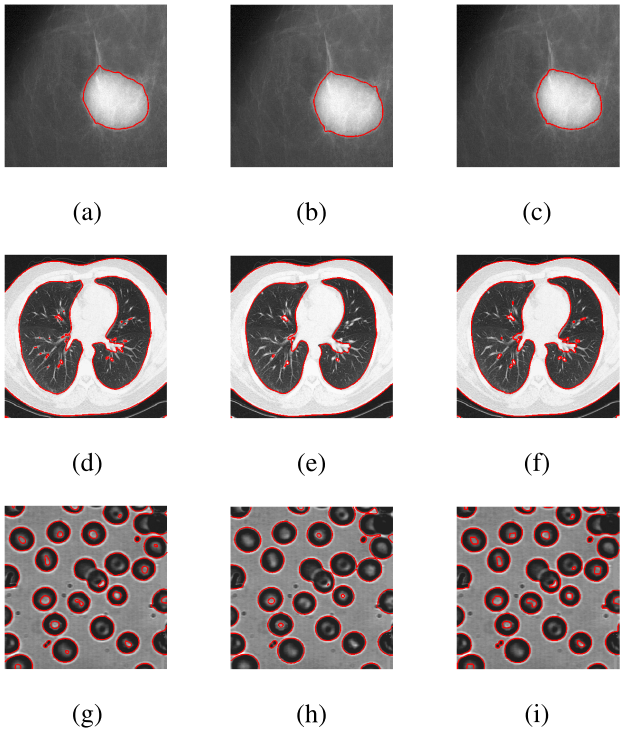
TABLE III
CONVERGENCE AND COMPUTING TIMES FOR EXPERIMENT 1

	Iterations	Computing time (seconds)
Proposed method (23)	2	0.13
Sobolev [21]	10	1.11
Generalised Newton [1]	30	5.77

A. Comparison With State-of-the-Art Algorithms for the Chan-Vese AC

1) Comparison With Other Steepest Descent Algorithms: In the first part of this experiment we compare our method with two other fast steepest descent algorithms: the generalised Newton method [1], and the Sobolev gradient [21]. These methods have been specifically designed for the Chan-Vese AC, which represents the image foreground and background using Gaussian distribution. For this experiment we also implemented our method using the Chan-Vese AC and the corresponding Bregman divergence, which is provided in the second row of Table I. To guarantee that the comparisons are fair we have used a synthetic shape image from [1] and the accompanying MATLAB codes. This image contains 216×187 pixels and is contaminated with white additive Gaussian noise of 5.36dB SNR. Similarly, all methods were initialised by setting $\boldsymbol{\phi}^0$ to a right circular cone as in [1]. We implemented our method with $\gamma_t = 5$ and $\sigma = 0.75$ as these values produced the best results. For the other methods we used the values recommended in [1] for this image, which also produce best results.

Fig. 1 depicts the contours estimated with each method (for each experiment the initialisation and the final contour are depicted in blue and red respectively). Fig. 1(a) shows the segmentation obtained with the proposed natural gradient descent method. The results obtained with the Sobolev gradient [21] and the generalised Newton method [1] are presented in Figures 1(b) and 1(c). The segmentation results obtained with the three methods are very similar but Table III shows that the proposed method converged in only 2 iterations and was between 5 and 40 times faster than the state-of-the-art methods, which required 10 and 30 iterations to converge. The convergence speed of the proposed method results from the fact that active contours form a statistical manifold that is not Euclidean and our algorithm is taking steps along the steepest direction on that manifold. The Sobolev gradient does not take into account this geometrical information and a result requires more iterations to converge. The generalised Newton method is taking steps along the steepest direction on the manifolds associated with the Hessian matrices of $E(\boldsymbol{y}; \phi)$. However, because $E(\boldsymbol{y}; \phi)$ is not convex the method has to use a trust-region constraint to limit the length of the steps, and as a result it requires a large number of iterations to converge. Finally, it is worth mentioning that segmentation results obtained by using our algorithm with line search for γ_t can be found in [2]. In that case our algorithm also converges in 2 iterations, which require 0.24 seconds instead of 0.13 because of the additional computational cost of the line search. The contours obtained with an Euclidean steepest descent and with an Euclidean Newton method are provided in [1]. Additional experiments with other standard test images are also provided in [2].



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 2. Comparison with the state of the methods [10] and [27] using the breast lesion (400×400 pixels), lung (336×336 pixels) and bacteria (380×380 pixels) images from the supplementary material of [10]. (a) Proposed (23). (b) FGMA [10]. (c) Two-stage [27]. (d) Proposed (23). (e) FGMA [10]. (f) Two-stage [27]. (g) Proposed (23). (h) FGMA [10]. (i) Two-stage [27].

2) Comparison With Convex Optimisation Algorithms: In the second part of this experiment we compare our method with two state-of-the-art segmentation algorithms based on convex approximations of the Chan-Vese AC model: the fast global minimisation algorithm (FGMA) [10], and the recent two-stage algorithm [27]. To provide a fair comparison we have used three standard test images from the supplementary material of [10] that have very different compositions and structures. The segmentation results and computing times for this experiment are reported in Figure 2 and Table IV. These results show that our methods performs similarly to the

TABLE IV
CONVERGENCE AND COMPUTING TIMES FOR THE "LESION", "LUNG" AND "BACTERIA" IMAGES DEPICTED IN FIG. 2

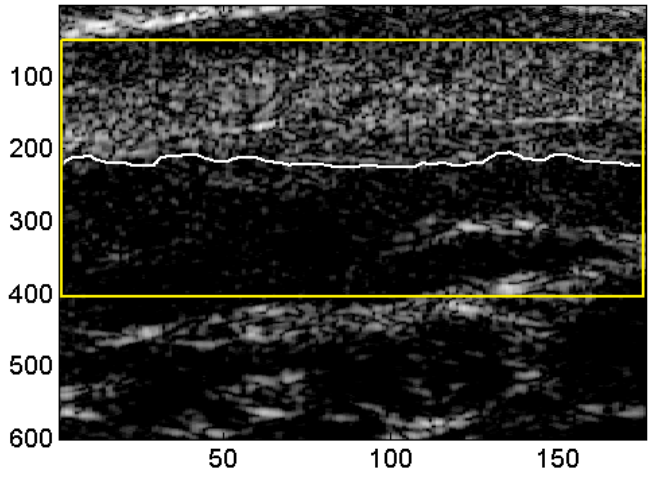
	Iterations	Computing time (seconds)
Proposed method (23)	2	0.20
FGMA [10]	7	0.66
Two-stage algorithm [27]	17	2.78
Proposed method (23)	2	0.18
FGMA [10]	3	0.32
Two-stage algorithm [27]	16	1.96
Proposed method (23)	2	0.21
FGMA [10]	2	0.47
Two-stage algorithm [27]	19	3.02

algorithms of the state-of-the-art in terms of segmentation, but is significantly faster as illustrated in Table IV. Because it is not entirely clear what is the best segmentation for each image we do not claim that one specific method outperforms the others in terms of segmentation performance. For example, the second image is a CT scan of the interior of a lung and a segmentation result could either ignore the alveoli branches within the lung and only segment the boundary around the interior, or equally it could segment the boundary and also the alveoli branches, (note that this will also depend on the amount of regularisation specified by the user). Similarly, for the bacteria image a segmentation could only identify the external boundaries of the individual bacteria, or it could also seek to highlight the structures in the interior of the bacteria. In all cases we have chosen, for comparison purposes, to produce similar segmentation results.

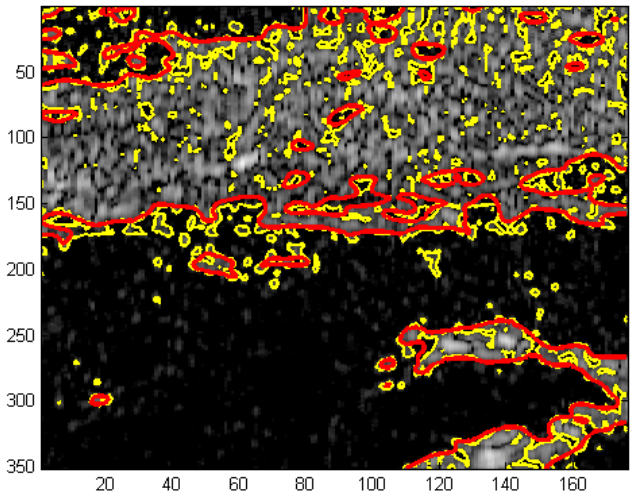
B. In-Vivo Ultrasound Image

Here we consider the Rayleigh AC model proposed in [11] for ultrasound image segmentation, which was applied in that work to echocardiography. This model represents the image foreground and background using Rayleigh likelihood functions and to the best of our knowledge the Euler method presented in [11] is the only algorithm that has been proposed for this AC. Here we compare that method with our Riemannian descent algorithm (23), which we have implemented using the Bregman divergence for the Rayleigh distribution provided in the last row of Table I. To guarantee that the comparisons are fair both algorithms use the same initialization and step size (we used $\gamma_t = 0.1$ as recommended in [11] and initialised $\boldsymbol{\phi}^0$ to a right circular cone as in [1]).

Fig. 3(a) shows a B-mode ultrasound image of in-vivo human dermis, which is a particularly challenging type of image for image segmentation problems [6]. The region of interest used in the experiment is depicted in yellow (the size of this region is 350×175 pixels). This ultrasound image was acquired at the dermatology service of the Hospital of Toulouse with a dermocup system (Atys Medical, France), equipped with a single-element focalized 25MHz wide-band (40-percent) probe sampled at 100MHz with a $53\mu m$ mechanic lateral step. In this experiment we use the Rayleigh



(a)



(b)

Fig. 3. Comparison of our method (red) with Euler's method (yellow) on a Rayleigh AC with application to ultrasound image segmentation. (a) Ultrasound image. (b) Segmentation results.

AC model to identify automatically the dermis-hypodermis junction, which has been annotated approximately by an expert (coarse white line). Identifying this skin structure is important in dermatology, cosmetology and pharmaceutical science. It allows evaluating the thickness of the dermis, which is important for *"general aesthetics, the use of cosmetics and drugs, optimally positioning skin grafts, and effective massage. Since you can also detect early signs of pathological skin thickening, it is possible to use this knowledge to offer preventative treatment."* [32].

Fig. 3(b) shows in coarse red the segmentation results obtained with our method and in yellow those obtained with [11], using the values for the algorithm parameters that produced the best results. We observe that the proposed method converged to a solution that is in better agreement with the expert's annotation of the dermis-hypodermis junction

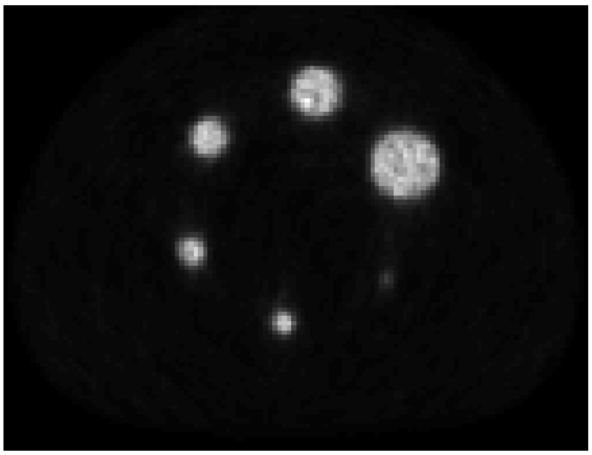
TABLE V
SPHERE DIAMETER ESTIMATION (IN MILLIMETERS)

	True	proposed (23)	Schaefer [31]	Otsu [34]
Sphere 1	37.0	37.5	35.1	35.1
Sphere 2	28.0	28.1	26.9	25.7
Sphere 3	22.0	23.4	21.1	21.1
Sphere 4	17.0	16.4	15.2	14.1
Sphere 5	13.0	14.0	12.1	11.7
Sphere 6	10.0	n/a	5.9	n/a

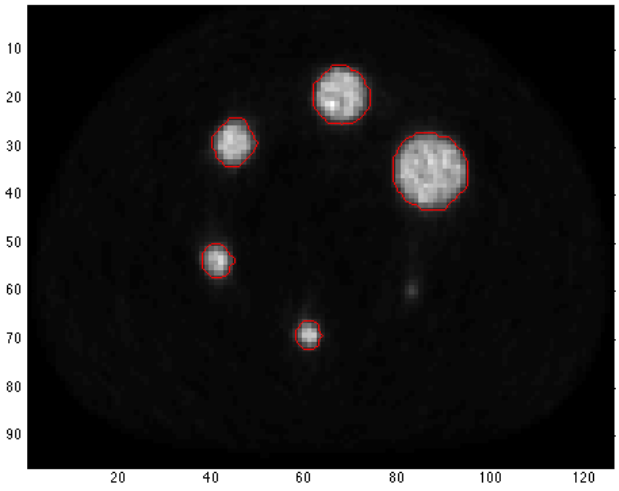
than the solution obtained with [11], which is more corrupted by speckle noise (note that further increasing the amount of regularisation for this algorithm produced a degenerate segmentation in which all pixels belonged to a single class). More importantly, the proposed method converged in only 18 iterations, which took 0.34 seconds, whereas Euler's method required 9860 iterations and 269 seconds to produce a stable solution. Again, this dramatic difference in speed results from the fact that proposed method takes steps along the steepest direction on the parameter's intrinsic manifold, as opposed to using the default Euclidean gradient.

C. PET Phantom Image

In this experiment we consider the problem of segmenting high radioactivity regions in Positron-Emitted-Tomographic (PET) images. Accurate segmentation of high activity regions in PET images is very important in oncology, where clinicians use these images to measure tumour metabolic activity and devise appropriate treatment plans. Here we study an active contour method for PET image segmentation based on a Poisson likelihood, which is a widely used statistical model for these images. We compute this Poisson AC with a Riemannian descent algorithm (23) implemented using the Bregman divergence of the Poisson distribution provided in the fourth row of Table I. In order to assess the performance of this method under controlled conditions we have used the PET phantom image depicted in Fig. 4(a), which has been acquired at the Claudius Regaud Cancer Institute of Toulouse using a GE-DST clinical PET scanner. This image has been produced by imaging a phantom containing 6 radioactive spheres whose size is known and is regularly used to benchmark and calibrate PET segmentation techniques. Table V shows the true diameter of each sphere and the estimates obtained with the proposed method (23). For comparison, Table V also shows the estimates obtained with a state-of-the-art thresholding method that is currently extensively used in clinical applications. More precisely, we compare our estimates with the results obtained with the thresholding method proposed by Schaefer *et al.* in [31], which was recently reported to be among the best PET image segmentation methods in an independent comparative study [33]. We also report the results obtained with the widely used thresholding method proposed by Otsu in [34]. To ease visual interpretation, the best estimate for each sphere has been highlighted in red. The results obtained with the proposed methods are comparable to those obtained with the state-of-the-art method as indicated



(a)



(b)

Fig. 4. Segmentation of a PET radioactive phantom image using a Poisson AC model. (a) PET image (96 × 126). (b) Proposed method (23).

in (Table V). However, note that the method proposed in [31] requires the user provide a coarse pre-segmentation of the interior of each sphere (which is used to compute the threshold), whereas our method detects lesions automatically without any user intervention. For completeness, Fig. 4(b) shows in red the segmentation produced by the proposed method. This segmentation was computed in 4 iterations using $\sigma = 0.2$, $\gamma_t = 0.1$, and required 0.013 seconds.

V. CONCLUSION

This paper has addressed the problem of solving active contour models from the exponential family, (which comprises most of the active contours used in the literature), with application to image segmentation. Through appropriate mathematical developments we have shown that the Fisher Information matrix, which determines the natural metric of the statistical manifold is diagonal and has a closed-form

analytical expression. This insight enabled a fast converging segmentation methodology which we have demonstrated and compared on three images: a synthetic image, a PET image and an ultrasound image segmentation problem where a 2-orders of magnitude increase in segmentation speed is achieved.

Future work will focus on the application of the proposed methodology to the development of segmentation algorithms for applications with real-time constraints. In particular, we plan to develop 2D and 3D ultrasound image segmentation techniques for in-vivo fetal and cardiac echography, as well as 4D MRI segmentation methods based on a Rician AC.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of the Hospital of Toulouse, Pierre Fabre Laboratories and Magellium for the ultrasound image acquisition and interpretation. The authors would particularly like to thank Dr. Nicolas Meyer, Dr. Siham Lourari, and Mr. Jérôme Georges. The authors are also grateful to Dr. Frédéric Courbon and his team at the Claudius Regaud Cancer Institute of Toulouse for the PET phantom image used in Section IV-C. Finally, the authors would like to thank the anonymous reviewers who commented on an earlier version of this paper which have significantly improved it.

REFERENCES

[1] L. Bar and G. Sapiro, "Generalized Newton-type methods for energy formulations in image processing," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 508–531, 2009.

[2] M. Pereyra, H. Batatia, and S. McLaughlin, "Exploiting information geometry to improve the convergence properties of variational active contours," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 4, pp. 700–707, Aug. 2013.

[3] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.

[4] F. Lecellier, J. Fadili, S. Jehan-Besson, G. Aubert, M. Revenu, and E. Saloux, "Region-based active contours with exponential family observations," *J. Math. Imag. Vis.*, vol. 36, no. 1, pp. 28–45, Jan. 2010.

[5] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations," *J. Comput. Phys.*, vol. 79, no. 1, pp. 12–49, 1988.

[6] M. Pereyra, N. Dobigeon, H. Batatia, and J. Tourneret, "Estimating the granularity coefficient of a Potts–Markov random field within a Markov chain Monte Carlo algorithm," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2385–2397, Jun. 2013.

[7] J. Yuan, E. Bae, X.-C. Tai, and Y. Boykov, "A continuous max-flow approach to Potts model," in *Proc. 11th Eur. Conf. Comput. Vis. (ECCV)*, vol. 6316. 2010, pp. 379–392.

[8] F. Li, M. K. Ng, T. Y. Zeng, and C. Shen, "A multiphase image segmentation method based on fuzzy region competition," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 277–299, 2010.

[9] T. F. Chan, S. Esedoglu, and M. Nikolova, "Algorithms for finding global minimizers of image segmentation and denoising models," *SIAM J. Appl. Math.*, vol. 66, no. 5, pp. 1632–1648, 2006.

[10] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, and S. Osher, "Fast global minimization of the active contour/snake model," *J. Math. Imag. Vis.*, vol. 28, no. 2, pp. 151–167, Jun. 2007.

[11] A. Sarti, C. Corsi, E. Mazzini, and C. Lamberti, "Maximum likelihood segmentation of ultrasound images with Rayleigh distribution," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 52, no. 6, pp. 947–960, Jun. 2005.

[12] I. Ben Ayed, A. Mitiche, and Z. Belhadj, "Multiregion level-set partitioning of synthetic aperture radar images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 793–800, May 2005.

[13] I. Ben Ayed, N. Hennane, and A. Mitiche, "Unsupervised variational image segmentation/classification using a Weibull observation model," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3431–3439, Nov. 2006.

[14] G. Aubert, M. Barlaud, O. Faugeras, and S. Jehan-Besson, "Image segmentation using active contours: Calculus of variations or shape gradients?" *SIAM J. Appl. Math.*, vol. 63, no. 6, pp. 2128–2154, 2003.

[15] S. Amari and S. C. Douglas, "Why natural gradient?" in *Proc. IEEE ICASSP*, Seattle, WA, USA, May 1998, pp. 1213–1216.

[16] M. Hintermüller and W. Ring, "A second order shape optimization approach for image segmentation," *SIAM J. Appl. Math.*, vol. 64, no. 2, pp. 442–467, 2004.

[17] M. Hintermüller and W. Ring, "An inexact Newton-CG-type active contour approach for the minimization of the Mumford–Shah functional," *J. Math. Imag. Vis.*, vol. 20, nos. 1–2, pp. 19–42, 2004.

[18] M. Hintermüller, "A combined shape Newton and topology optimization technique in real time image segmentation," in *Real-Time PDE-Constrained Optimization*. Philadelphia, PA, USA: SIAM, 2007, pp. 253–274.

[19] G. Sundaramoorthi, A. Yezzi, and A. C. Mennucci, "Sobolev active contours," *Int. J. Comput. Vis.*, vol. 73, no. 3, pp. 345–366, Jul. 2007.

[20] G. Charpiat, P. Maurel, J.-P. Pons, R. Keriven, and O. Faugeras, "Generalized gradients: Priors on minimization flows," *Int. J. Comput. Vis.*, vol. 73, no. 3, pp. 325–344, 2007.

[21] G. Sundaramoorthi, A. Yezzi, A. C. Mennucci, and G. Sapiro, "New possibilities with Sobolev active contours," *Int. J. Comput. Vis.*, vol. 84, no. 2, pp. 113–129, May 2009.

[22] R. J. Renka, "Image segmentation with a Sobolev gradient method," *Nonlinear Anal., Theory, Methods, Appl.*, vol. 71, no. 12, pp. e774–e780, Dec. 2009.

[23] R. Chartrand and V. Staneva, "A faster-converging algorithm for image segmentation with a modified Chan–Vese model," in *Proc. IPCV*, 2008.

[24] E. Mendi and M. G. Milanova, "Quasi-Newton minimization for active contours with Chan–Vese model," in *Proc. IPCV*, 2009, pp. 575–578.

[25] S.-I. Amari, *Differential-Geometrical Methods in Statistics* (Lecture Notes in Statistics 28). New York, NY, USA: Springer-Verlag, 1990.

[26] F. Nielsen and R. Nock, "Entropies and cross-entropies of exponential families," in *Proc. 17th IEEE ICIP*, Sep. 2010, pp. 3621–3624.

[27] X. Cai, R. Chan, and T. Zeng, "A two-stage image segmentation method using a convex variant of the Mumford–Shah model and thresholding," *SIAM J. Imag. Sci.*, vol. 6, no. 1, pp. 368–390, Aug. 2013.

[28] R. Deriche, "Recursively implementing the Gaussian and its derivatives," in *Proc. 2nd ICIP*, Sep. 1992, vol. 71, no. 12, pp. 263–267.

[29] P. Cachier, E. Bardinet, D. Dormont, X. Pennec, and N. Ayache, "Iconic feature based nonrigid registration: The PASHA algorithm," *Comput. Vis. Image Understand.*, vol. 89, nos. 2–3, pp. 272–298, Feb. 2003.

[30] S. Lankton and A. Tannenbaum, "Localizing region-based active contours," *IEEE Trans. Image Process.*, vol. 17, no. 11, pp. 2029–2039, Nov. 2008.

[31] A. Schaefer, S. Kremp, D. Hellwig, C. Rübe, C.-M. Kirsch, and U. Nestle, "A contrast-oriented algorithm for FDG-PET-based delineation of tumour volumes for the radiotherapy of lung cancer: Derivation from phantom measurements and validation in patient data," *Eur. J. Nucl. Med. Molecular Imag.*, vol. 35, no. 11, pp. 1989–1999, 2008.

[32] Y. Miyamae, M. Kawabata, Y. Yamakawa, J. Tsuchiya, and Y. Ozaki, "Non-invasive estimation of skin thickness by near infrared diffuse reflection spectroscopy—Separate determination of epidermis and dermis thickness," *J. Near Infr. Spectrosc.*, vol. 20, no. 6, pp. 617–622, 2012.

[33] H. Zaidi, M. Abdoli, C. L. Fuentes, and I. M. El Naqa, "Comparative methods for PET image segmentation in pharyngolaryngeal squamous cell carcinoma," *Eur. J. Nucl. Med. Molecular Imag.*, vol. 39, no. 5, pp. 881–891, 2012.

[34] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

Marcelo Pereyra (S'09–M'13) was born in Buenos Aires, Argentina, in 1984. He received the M.Eng. degree in electronics engineering from the Instituto Tecnológico de Buenos Aires, Buenos Aires, and the Institut National des Sciences Appliquées de Toulouse (INSA Toulouse), Toulouse, France, the M.Sc. degree from INSA Toulouse in 2009, and the Ph.D. degree in signal processing from the Institut National Polytechnique de Toulouse, Toulouse, in 2012. He currently holds a Brunel Post-Doctoral Research Fellowship in Statistics with the School of Mathematics, University of Bristol, Bristol, U.K. His research activities are centered around statistical image processing with a particular interest in Bayesian methods, Monte Carlo algorithms, and medical imaging applications.

Hadj Batatia received the M.Sc. and Ph.D. degrees in computer engineering from the University of Toulouse, Toulouse, France, in 1992. He was a Senior Lecturer with the University of Teesside, Middlesbrough, U.K., before joining the University of Toulouse in 1999. He is currently a Senior Lecturer with the Institut National Polytechnique de Toulouse, Toulouse. His research activity focuses on medical image processing. His interests cover statistical medical image modeling, motion compensation, segmentation, and tissue characterization. He supervised several Ph.D. theses in this field.

Steve McLaughlin (F'11) was born in Clydebank, U.K., in 1960. He received the B.Sc. degree in electronics and electrical engineering from the University of Glasgow, Glasgow, U.K., in 1981, and the Ph.D. degree from the University of Edinburgh, Edinburgh, U.K., in 1989.

He was a Development Engineer with the industry from 1981 to 1984, where he was involved in the design and simulation of integrated thermal imaging and fire control systems. From 1984 to 1986, he worked on the design and development of high frequency data communication systems. In 1986, he joined the Department of Electronics and Electrical Engineering, University of Edinburgh, as a Research Fellow, where he studied the performance of linear adaptive algorithms in high noise and nonstationary environments. In 1988, he joined as an Academic Staff with the University of Edinburgh. From 1991 to 2001, he held a Royal Society University Research Fellowship to study nonlinear signal processing techniques. In 2002, he was the Personal Chair of Electronic Communication Systems with the University of Edinburgh.

Prof. McLaughlin research interests lie in the fields of adaptive signal processing and nonlinear dynamical systems theory and their applications to biomedical, geophysical, imaging, and communication systems. He is a fellow of the Royal Academy of Engineering, the Royal Society of Edinburgh, and the Institute of Engineering and Technology.