Total Bregman Divergence and its Applications to DTI Analysis

Baba C. Vemuri, *Fellow, IEEE,* Meizhu Liu, *Member, IEEE,*
Shun-Ichi Amari, *Fellow, IEEE,* and Frank Nielsen, *Senior Member, IEEE,*

Abstract—**Divergence measures provide a means to measure the pairwise dissimilarity between "objects", e.g., vectors and probability density functions (pdfs). Kullback-Leibler (KL) divergence and the square loss (SL) function are two examples of commonly used dissimilarity measures which along with others belong to the family of Bregman divergences (BD). In this paper, we present a novel divergence dubbed the Total Bregman divergence (TBD), which is intrinsically robust to outliers, a very desirable property in many applications. Further, we derive the TBD center, called the *t-center* (using the ℓ_1-norm), for a population of positive definite matrices in closed form and show that it is invariant to transformation from the special linear group. This *t-center*, which is also robust to outliers, is then used in tensor interpolation as well as in an active contour based piecewise constant segmentation of a diffusion tensor magnetic resonance image (DTMRI). Additionally, we derive the piecewise smooth active contour model for segmentation of DTMRI using the TBD and present several comparative results on real data.**

I. INTRODUCTION

IN applications that involve measuring the dissimilarity between two objects (vectors, matrices, functions, images and so on) the definition of a divergence/distance becomes necessary. A good divergence measure should be precise and statistically robust. The state of the art has many widely used divergences. The square loss function (SL) has been used widely for regression analysis; KL divergence [1], has been applied to compare the difference between two pdfs; and the Mahalanobis distance [2], is used to measure the dissimilarity between two random vectors of the same distribution. All the aforementioned divergences are special cases of the Bregman divergence which of late has been widely researched by many both from a theoretical and practical viewpoint, see [3]–[5] and references therein. At this juncture, it would be worth inquiring, why does one need yet another divergence? The answer would be that none of the existing divergences are statistically robust and one would need to use M-estimators from robust statistics literature to achieve robustness. This robustness however comes at a price, which is, computational cost and accuracy. Moreover, some of the divergences lack invariance to transformations such as similarity, affine etc.

Such invariance becomes important when dealing with for example, segmentation, it is desirable to achieve invariance to similarity or affine transformations that two different scans of the same patient might be related by.

In this work, we propose a new class of divergences which measure the orthogonal distance (in contrast to the ordinal distance used in Bregman divergence) between the value of a differentiable convex function at the first argument and its tangent at the second argument. We dub this divergence the total Bregman divergence (TBD). A geometrical illustration of the difference between TBD and BD is given in Fig. 1. $d_f(x, y)$ is the Bregman divergence between x and y based on a differentiable convex function f and $\delta_f(x, y)$ is the TBD between x and y based on the same differentiable convex function f. We can observe that $d_f(x, y)$ will change if we apply a rotation to the graph $(x, f(x))$, while $\delta_f(x, y)$ does not.
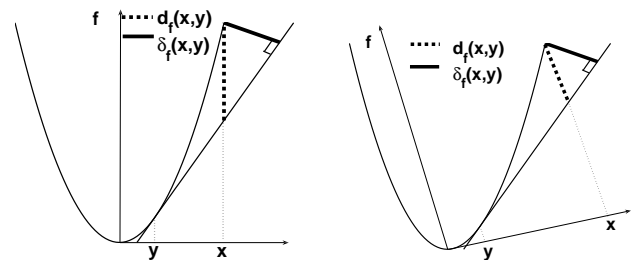

Fig. 1. In each figure, $d_f(x, y)$ (dotted line) is BD, $\delta_f(x, y)$ (bold line) is TBD. Left figure shows $d_f(x, y)$ and $\delta_f(x, y)$ before rotating the coordinate system. Right figure shows $d_f(x, y)$ and $\delta_f(x, y)$ after rotating the coordinate system. $d_f(x, y)$ changes with rotation while $\delta_f(x, y)$ is invariant with rotation.

Bregman divergence has been widely used in clustering, a ubiquitous task in machine learning applications. In clustering, cluster centers are defined using either a distance or a divergence. In this paper, we will define a cluster center using the TBD in conjunction with the ℓ_1-norm that we dub the *t*-center. The t-center can be viewed as the cluster representative that minimizes the ℓ_1-norm TBD between itself and the members of a given population. We derive an analytic expression for the *t*-center which gives it a major advantage over its rivals (for example, the Riemannian median for SPD tensors). Since, the TBD can be viewed as a weighted BD with the weight being inversely proportional to the magnitude gradient of the convex function, the *t-center* can thus be viewed as a weighted median of sorts. This weighting scheme makes *t-center* robust to noise and outliers, since it's inversely dependent on the gradient

of the convex function. We prove the robustness property theoretically in section III and experimentally demonstrate it in the experiments section. Another salient feature of the t-center is that it can be computed very efficiently due to its closed form expression leading to efficient clustering algorithms, a topic of our continuing research.

To illustrate the robustness and computational efficiency properties of TBD, we applied our theory to the space of order-2 symmetric positive definite (SPD) tensors which are commonly encountered in applications such as diffusion tensor magnetic resonance imaging (DT-MRI) [6], [7], Computer Vision (structure tensors), Elastography etc. We prove that TBD between SPD tensors is invariant to transformations from the special linear group ($SL(n)$). Recall that, if $A \in \mathbb{R}^{n \times n}$, then $A \in SL(n)$ implies $\det A = 1$. Given random variables X and Y, where $X \sim \mathcal{N}(x; P)$ and $Y \sim \mathcal{N}(y; Q)$, $\mathcal{N}(.; P)$ and $\mathcal{N}(.; Q)$ are Gaussian distributions with zero mean and covariances P and Q respectively, we can define $BD(X, Y)$ and generalize it to define $TBD(X, Y)$. We can show that $TBD(X, Y)$ is only related to P and Q, thus, $TBD(P, Q) = TBD(X, Y)$. If $x \mapsto Ax$, then $P \mapsto A'PA$, TBD has the property of $TBD(A'PA, A'QA) = TBD(P, Q)$, $\forall A \in SL(n)$. We prove that the ℓ_1 t-center, P^*, for a population of SPD tensors $\{Q_i\}$ has an analytic form; furthermore, t-centers P^* and $\hat{P}^*$ corresponding to the original and transformed tensors is related by A and $\hat{P}^* = A'P^*A$. This property makes t-center suitable for atlas construction applications in tensor field processing, which is very useful in DT-MRI [8]–[11]. Atlases are usually defined as representatives (most often as an average of sorts) of a population of images, shapes etc. Atlas construction using t-center is efficient and robust to outliers, a property that is derived theoretically and demonstrated experimentally in section IV. Further, we present quantitative comparisons between the proposed and other existing divergences, such as the Frobenius norm [12], Riemannian metric [9], [10], [13]–[15], symmetrized KL divergence [14], [16], [17] and the log-Euclidean distance [18]. Although some of these divergences are invariant to transformations in $SL(n)$, none of them are statistically robust to outliers encountered in tensor field processing. When using the Riemannian metric to compute population statistics, neither the mean nor the median can be computed in closed form, which makes this metric computationally very expensive. Finally, we perform segmentation on synthetic tensor fields and real DTI images using TBD and compare the results with other divergences demonstrating the efficiency and robustness of TBD to outliers over its rivals.

The rest of this paper is organized as follows. In Section II we review the conventional Bregman divergence, followed by the definition of TBD and derivation of its properties. Section III introduces the t-center, which is derived from TBD, and delves into its better accuracy as a representative than centers obtained from other divergence measures or metrics. The detailed description of the experimental design and results with quantitative comparison with other divergence are presented in Section IV. Finally, we draw conclusions in Section VI.

II. TOTAL BREGMAN DIVERGENCE

In this section, we first recall the definition of conventional Bregman divergence [4] and then define the TBD. Both divergences are dependent on the corresponding convex and differentiable function $f : X \mapsto \mathbb{R}$ that induces the divergence.

A. Definition of TBD and Examples

Definition II.1. *The Bregman divergence d associated with a real valued strictly convex and differentiable function f defined on a convex set X between points $x, y \in X$ is given by,*

$$d_f(x, y) = f(x) - f(y) - \langle x - y, \nabla f(y) \rangle, \qquad (1)$$

where $\nabla f(y)$ is the gradient of f at y and $\langle \cdot, \cdot \rangle$ is the inner product determined by the space on which the inner product is being taken.

For example, if $f : \mathbb{R}^n \mapsto \mathbb{R}$, then $\langle \cdot, \cdot \rangle$ is just the inner product of vectors in $\mathbb{R}^n$, and $d_f(\cdot, y)$ can be seen as the distance between the first order Taylor approximation to f at y and the function evaluated at x. Bregman divergence d_f is non-negative definite and does not satisfy the triangular inequality thus making it a divergence. As shown in Fig. 1, Bregman divergence measures the ordinate distance.

Definition II.2. *The total Bregman divergence (TBD) δ_f associated with a real valued strictly convex and differentiable function f defined on a convex set X between points $x, y \in X$ is defined as,*

$$\delta_f(x, y) = \frac{f(x) - f(y) - \langle x - y, \nabla f(y) \rangle}{\sqrt{1 + \|\nabla f(y)\|^2}}, \qquad (2)$$

$\langle \cdot, \cdot \rangle$ is inner product as in definition II.1, and $\|\nabla f(y)\|^2 = \langle \nabla f(y), \nabla f(y) \rangle$ generally.

Consider the Bregman divergence $d_f(\eta, \zeta) = f(\eta) - f(\zeta) - \langle \eta - \zeta, \nabla f(\zeta) \rangle$ obtained for strictly convex and differentiable generator f. Let $\hat{x}$ denote the point $(x, z = f(x))$ of $X \times \mathbb{R}$ lying on the graph $\mathcal{F} = \{\hat{x} = (x, z = f(x)) \,|\, x \in X\}$. We visualize the Bregman divergence as the *z-vertical difference* between the hyperplane H_ζ tangent at $\hat{\zeta}$ and the hyperplane parallel to H_ζ and passing through $\hat{\eta}$:

$$d_f(\eta, \zeta) = f(\eta) - H_\zeta(\eta), \qquad (3)$$

with $H_\zeta(x) = f(\zeta) + \langle x - \zeta, \nabla f(\zeta) \rangle$. If instead of taking the vertical distance, we choose the minimum distance between those two hyperplanes H_η and H_ζ, we obtain the total Bregman divergence (by analogy to total least square fitting where the projection is orthogonal).

Since the distance between two parallel hyperplanes $\langle x, a \rangle + b_1 = 0$ and $\langle x, a \rangle + b_2 = 0$ is $\frac{|b_1 - b_2|}{\|a\|}$, letting $a = (\nabla f(\zeta), -1)$ and $b_1 = f(\eta) - \langle \eta, \nabla f(\zeta) \rangle$, and $b_2 = f(\zeta) - \langle \zeta, \nabla f(\zeta) \rangle$, we deduce that the total Bregman divergence is

$$\delta_f(\eta, \zeta) = \frac{d_f(\eta, \zeta)}{\sqrt{1 + \|\nabla f(\zeta)\|^2}} \qquad (4)$$

As shown in Fig. 1, TBD measures the orthogonal distance. Compared to the BD, TBD contains a weight factor (the denominator) which complicates the computations. However,

this structure brings up many new and interesting properties and makes TBD an "adaptive" divergence measure in many applications. Note that, in practice, X can be an interval, the Euclidean space, a d-simplex, the space of non-singular matrices or the space of functions [19]. For instance, in the application to SPD tensor interpolation, we let p and q be two pdfs, and $f(p) := \int p \log p$, then $\delta_f(p, q)$ becomes what we call the Total Kullback-Leibler divergence (tKL). Note that if q is a pdf or a probability mass function (pmf), then $\|\nabla f(q)\|^2 = \int (1 + \log q)^2 q$. More on tKL later in section IV. Table I lists some TBDs with various associated convex functions f.

III. THE t-CENTER

In many applications of computer vision and machine learning such as image and shape retrieval, clustering, classification etc., it is common to seek a representative for a set of objects having similar features. This representative normally is a cluster center of sorts, thus, it is desirable to seek a center that is robust to outliers and is efficient to compute. In this section, we use the TBD to derive an ℓ_1-norm cluster center that we call the t-center, and explore its properties.

A. Definition

Let $f : X \mapsto \mathbb{R}$ be a convex and differentiable function and $E = \{x_1, x_2, \cdots, x_n\}$ be a set of n points in X, then, the TBD δ_f^p between a point $x \in X$ and E associated with f and ℓ_p norm is defined as

$$\delta_f^p(x, E) = \left(\sum_{i=1}^n (\delta_f(x, x_i))^p \right)^{1/p} \quad (5)$$

The ℓ_p-norm center x_p^* is defined as $x_p^* = \arg\min_x \delta_f^p(x, E)$. These centers bear the name of median ($p = 1$), mean ($p = 2$) and circumcenter ($p \to \infty$). In this paper, we call the median ($p = 1$) the t-center and will derive an analytic form for the t-center and focus on its applications. Before delving into the t-center, we define the circumcenter and the centroid (mean).

B. ℓ_∞-norm circumcenter x_c

In the limiting case, the circumcenter amounts to solving for

$$x_c = \arg\min_x \max_i \delta_f(x, x_i). \quad (6)$$

x_c does not have a closed form, even though it can be solved to an arbitrarily fine approximation in Riemannian cases [5].

C. ℓ_2-norm Centroid x_m

$$x_m = \arg\min_x \sqrt{\sum_{i=1}^N (\delta_f(x, x_i))^2} \quad (7)$$

Like the ℓ_∞-norm circumcenter, ℓ_2-norm mean does not have a closed form in general and is hence expensive to compute.

D. The ℓ_1-norm t-center

Given a set E, we can obtain the ℓ_1-norm t-center x^* of E by solving the minimization problem

$$x^* = \arg\min_x \delta_f(x, E) = \arg\min_x \sum_{i=1}^n \delta_f(x, x_i) \quad (8)$$

Theorem III.1. *The t-center of a population of objects (densities, vectors etc.) with respect to a given divergence exists, and is unique.*

Proof: To find x^*, we take the derivative of $F(x) = \sum_{i=1}^n \delta_f(x, x_i)$ with respect to x, and set it to 0, as in

$$\nabla F(x) = \sum_{i=1}^n \frac{\nabla f(x) - \nabla f(x_i)}{\sqrt{1 + \|\nabla f(x_i)\|^2}} = 0 \quad (9)$$

Solving (9) yields:

$$\nabla f(x^*) = \left(\sum_{i=1}^n w_i \nabla f(x_i) \right) \bigg/ \left(\sum_{i=1}^n w_i \right), \quad (10)$$

where $w_i = \left(1 + \|\nabla f(x_i)\|^2\right)^{-1/2}$ is the weight that depends on $\nabla f(x_i)$. Recall that TBD $\delta_f(\cdot, y)$ is convex for any fixed $y \in X$, we know that $F(x) = \delta_f^1(x, E)$ is also convex as it is the sum of n convex functions. Hence, the solution to (9) exists and is indeed a minimizer of $F(x)$. Furthermore, since f is convex, ∇f is monotonic, so x^* is unique.

To obtain an explicit formula for x^*, we will make use of the dual space. Suppose g is the legendre dual function of f in the space X, and the dual space is denoted as Y. Then $\forall x \in X, \exists y \in Y$:

$$g(y) = \sup_x \{yx - f(x)\}, \text{ and } f(x) = \sup_y \{xy - g(y)\}, \quad (11)$$

where x and y satisfy $y = \nabla f(x)$ and $x = \nabla g(y)$. Thus far, we know that the t-center x^* satisfies (10), meaning that its legendre dual y^* should satisfy $y^* = \left(\sum_{i=1}^n w_i \nabla f(x_i)\right) / \left(\sum_{i=1}^n w_i\right)$, and $x^* = \nabla g(y^*)$. ∎

Note that, in most cases, if f is given, then we can get the explicit form for x^*, but x^* does not have the same expression for all convex functions, instead, x^* is dependent on the convex generator function for the chosen divergence. For a better illustration of theorem III.1, we provide three concrete examples of TBD with their t-centers in analytic form. The first is for the square loss function that we call total square loss (tSL), the second is for the exponential function and the third is the total Kullback-Leibler (tKL) as a measure of divergence between probability densities.

tSL: $f(x) = x^2$, the t-center $x^* = \left(\sum_i w_i x_i\right) / \left(\sum_i w_i\right)$, where $w_i = \left(1 + 4x_i^2\right)^{-1/2}$;

Exponentials: $f(x) = e^x$, the t-center $x^* = \left(\sum_i w_i x_i\right) / \left(\sum_i w_i\right)$, where $w_i = \left(1 + e^{2x_i}\right)^{-1/2}$;

tKL: $f(q) = \int q \log q \, dx$, which is the negative entropy [1], and $E = \{q_1, q_2, \cdots, q_n\}$ is a set of pdfs, the t-center

$$q^* = \arg\min_q \sum_{i=1}^n \frac{\int_X q \log \frac{q}{q_i} dx}{\sqrt{1 + \int_X (1 + \log q_i)^2 q_i dx}}. \quad (12)$$

X	$f(\mathbf{x})$	$\delta_f(\mathbf{x}, \mathbf{y})$	t-center	Remark
$\mathbb{R}$	$\mathbf{x}^2$	$(x-y)^2\left(1+4y^2\right)^{-1/2}$	$\sum_i w_i \mathbf{x_i}$	total square loss (tSL)
$\mathbb{R} - \mathbb{R}_-$	$\mathbf{x}\log\mathbf{x}$	$\left(x\log\frac{x}{y} - (x-y)\right)/\sqrt{1+(1+\log y)^2}$	$\prod_i (\mathbf{x_i})^{w_i}$	
$[0,1]$	$\mathbf{x}\log\mathbf{x} + \dot{\mathbf{x}}\log\dot{\mathbf{x}}$	$\left(x\log\frac{x}{y} + \dot{x}\log\frac{\dot{x}}{\dot{y}}\right)/\sqrt{1+y(1+\log y)^2+\dot{y}(1+\log\dot{y})^2}$	$\frac{\sum_i (\mathbf{x_i}/(1-\mathbf{x_i}))^{w_i}}{1+\sum_i (\mathbf{x_i}/(1-\mathbf{x_i}))^{w_i}}$	total logistic loss
$\mathbb{R}_+$	$-\log\mathbf{x}$	$\left(\frac{x}{y} - \log\frac{x}{y} - 1\right)/\sqrt{1+y^{-2}}$	$1/(\sum_i w_i/\mathbf{x_i})$	total Itakura-Saito distance
$\mathbb{R}$	$e^{\mathbf{x}}$	$(e^x - e^y - (x-y)e^y)/\sqrt{1+e^{2y}}$	$\sum_i w_i \mathbf{x_i}$	
$\mathbb{R}^d$	$\|\mathbf{x}\|^2$	$\|x-y\|^2/\sqrt{1+4\|y\|^2}$	$\sum_i w_i \mathbf{x_i}$	total squared Euclidean distance
$\mathbb{R}^d$	$\mathbf{x}'A\mathbf{x}$	$((x-y)'A(x-y))/\sqrt{1+4\|Ay\|^2}$	$\sum_i w_i \mathbf{x_i}$	total Mahalanobis distance
Δ^d	$\sum_{j=1}^d x_j \log x_j$	$\left(\sum_{j=1}^d x_j \log\frac{x_j}{y_j}\right)/\sqrt{1+\sum_{j=1}^d y_j(1+\log y_j)^2}$	$c\prod_i (\mathbf{x_i})^{w_i}$	total KL divergence (tKL)
$\mathbb{C}^{m\times n}$	$\|\mathbf{x}\|_F^2$	$\|x-y\|_F^2/\sqrt{1+4\|y\|_F^2}$	$\sum_i w_i \mathbf{x_i}$	total squared Frobenius

TABLE I

TBD δ_f CORRESPONDING TO f AND THE t-CENTER FOR A SET OF OBJECTS $\{x_i\}$. $\dot{x}=1-x$, $\dot{y}=1-y$, x' IS THE TRANSPOSE OF x. Δ^d IS d-SIMPLEX. c IS THE NORMALIZATION CONSTANT TO MAKE IT A PDF, $w_i = 1/\sqrt{1+\|\nabla f(x_i)\|^2}/(\sum_j 1/\sqrt{1+\|\nabla f(x_j)\|^2})$.

q^* is obtained by solving the Euler-Lagrange equation of the minimization in (12), to obtain

$$q^* = c\prod_i^n q_i^{w_i}, \quad w_i = \frac{\mu_i}{\sum_j \mu_j}, \quad \mu_i = \frac{1}{1+\int (1+\log q_i)^2 q_i dx},$$

where c is a normalization constant such that $\int_X q^*(x)dx = 1$. Furthermore, the t-centers for the most commonly used divergences can be derived in analytic form as shown in the Table I.

Theorem III.2. *The t-center is statistically robust to outliers.*

Proof: The robustness of t-center is analyzed by the influence function of an outlier y. Let x^* be the t-center of $E = \{x_1, \cdots, x_n\}$. When $\epsilon\%$ (ϵ small) of outlier y is mixed with E, x is influenced by the outliers, and the new center becomes to $\tilde{x}^* = x^* + \epsilon z(y)$. We call $z(y)$ the influence function. The center is robust when $z(y)$ does not grow even when y is very large. The influence curve is given explicitly in the case of ℓ_1-norm. $\tilde{x}^*$ is the minimizer of

$$(1-\epsilon)\frac{1}{n}\sum_{i=1}^n \delta(x, x_i) + \delta(x, y), \quad (13)$$

Hence, by differentiating the above function, setting it equal to zero at $\tilde{x}^* = x^* + \epsilon z$ and using the Taylor expansion, we have

$$z(y) = \frac{G^{-1}}{w(y)}(\nabla f(y) - \nabla f(x^*)), \quad (14)$$

where $G = \frac{1}{n}\sum_{i=1}^n \nabla\nabla f(x^*)w(x_i)$, and $w(y) = \sqrt{1+\|\nabla f(y)\|^2}$. Hence, the t-center is robust when z is bounded, i.e., $\forall y, |z| < c$, c is a constant.

$$|z(y)| < \frac{G^{-1}}{w(y)}(\|\nabla f(y)\| + \|\nabla f(x^*)\|). \quad (15)$$

Since $w(y) \geq \|\nabla f(y)\|$ and $w(y) \geq 1$, therefore,

$$|z(y)| < G^{-1}(1 + \|\nabla f(x^*)\|) = c. \quad (16)$$

∎

To make this proof more understandable, we give a simple example of the Euclidean case when $f(x) = \frac{1}{2}\|x\|^2, x \in \mathbb{R}^d$.

Then,

$$G = \frac{1}{n}\sum_{i=1}^n \frac{1}{\sqrt{1+\|x_i\|^2}}I, \quad (17)$$

and $z(y) = G^{-1}\frac{y}{\sqrt{1+\|y\|^2}}$, when $\|y\|$ is large, this is approximated by $z(y) = G^{-1}\frac{y}{\|y\|}$, which implies z is bounded, and the t-center is robust for large $\|y\|$.

Note that the influence function for the ordinary Bregman divergence is $z = y$, and hence is not robust.

Using the ℓ_1-norm t-center x^* has advantages over other centers resulting from the norms with $p > 1$ in the sense that x^* has an analytic form and is robust to outliers. These advantages are explicitly evident in the experiments presented subsequently.

IV. TBD APPLICATIONS

We now develop the applications of the t-center in interpolating diffusion tensors in DT-MRI data and segmentation of tensor fields specifically DT-MRI.

A. SPD Tensor interpolation applications

We now define tKL, the total Kullback-Leibler divergence between symmetric positive definite (SPD) rank-2 tensors (SPD matrices), and show that it is invariant to transformations belonging to the special linear group $SL(n)$. Further, we compute the t-center using tKL for SPD matrices, which has a closed form expression, as the weighted harmonic mean of the population of the tensors.

tKL between order two SPD tensors/matrices P and Q is derived using the negative entropy of the zero mean Gaussian density functions they correspond to. Note that order two SPD tensors can be seen as covariance matrices of zero mean Gaussian densities. Suppose,

$$p(x; P) = \frac{1}{\sqrt{(2\pi)^n \det P}}\exp\left(-\frac{1}{2}x'P^{-1}x\right), \quad (18)$$

$$q(x; Q) = \frac{1}{\sqrt{(2\pi)^n \det Q}}\exp\left(-\frac{1}{2}x'Q^{-1}x\right), \quad (19)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

5

then,

$$tKL(P,Q) = \frac{\int p \log \frac{p}{q} dx}{\sqrt{1 + \int (1 + \log q)^2 q dx}}$$

$$= \frac{\log(\det(P^{-1}Q)) + tr(Q^{-1}P) - n}{2\sqrt{c + \frac{(\log(\det Q))^2}{4} - \frac{n(1+\log 2\pi)}{2}\log(\det Q)}}, \quad (20)$$

where $c = \frac{3n}{4} + \frac{n^2 \log 2\pi}{2} + \frac{(n \log 2\pi)^2}{4}$. When an $SL(n)$ transformation is applied on x, i.e., $x \mapsto Ax$, then $P \mapsto A'PA$ and $Q \mapsto A'QA$. It is easy to see that

$$tKL(P,Q) = tKL(A'PA, A'QA), \quad \forall A \in SL(n), \quad (21)$$

which means that tKL between SPD tensors is invariant under the group action, when the group member belongs to $SL(n)$. Given an SPD tensor set $\{Q_i\}_{i=1}^m$, its t-center P^* can be obtained from (12) and

$$P^* = (\sum_i w_i Q_i^{-1})^{-1}, w_i = \frac{\mu_i}{\sum_j \mu_j}, \quad (22)$$

where $\mu_i = \left(2\sqrt{c + \frac{(\log(\det Q_i))^2}{4} - \frac{n(1+\log 2\pi)}{2}\log(\det Q_i)}\right)^{-1}$. It can be seen that $w_i(Q_i) = w_i(A'Q_iA), \forall A \in SL(n)$. If a transformation $A \in SL(n)$ is applied, the new t-center will be $\hat{P}^* = (\sum w_i(A'Q_iA)^{-1})^{-1} = A'P^*A$, which means that if $\{Q_i\}_{i=1}^m$ are transformed by some member of $SL(n)$, then the t-center will undergo the same transformation.

Also we can compute the tSL between P and Q from the density functions p, q and $f(p) = \int p^2 dx$, and the result is

$$tSL(P,Q) = \frac{\int (p-q)^2 dx}{\sqrt{1 + \int (2q)^2 q dx}} =$$

$$\frac{1/\sqrt{\det(2P)} + 1/\sqrt{\det(2Q)} - 2/\sqrt{\det(P+Q)}}{(2\pi)^n + 4\sqrt{(2\pi)^n}/\sqrt{\det(3Q)}},$$

also, $tSL(P,Q) = tSL(A'PA, A'QA), \quad \forall A \in SL(n)$, which means that tSL is also invariant under $SL(n)$ transformations. Similarly, we can also prove that the total Itakura-Saito distance and total squared Euclidean distance between SPD matrices are invariant under $SL(n)$ transformations. For the rest of this paper, we will focus on tKL.

There are several ways to define the distance between SPD matrices, e.g. using the Frobenius norm [12], Riemannian metric [10], [13]–[15], symmetrized KL divergence [14], [16], [17] and the log-Euclidean distance [18], respectively defined as

$$d_F(P,Q) = \|P - Q\|_F = \sqrt{tr((P-Q)'(P-Q))}, \quad (23)$$

$$d_R(P,Q) = \sqrt{\sum_{i=1}^n \log^2 \lambda_i}, \quad (24)$$

where $\lambda_i, i = 1, \cdots, n$, are eigenvalues of $P^{-1}Q$,

$$KL_s(P,Q) = \frac{1}{4}tr(Q^{-1}P + P^{-1}Q - 2I). \quad (25)$$

$$LE(P,Q) = \|\log(P) - \log(Q)\|_F. \quad (26)$$

$d_F(.,.)$ is not invariant to transformations in $SL(n)$, $d_R(.,.)$, $KL_s(.,.)$ and $LE(.,.)$ are invariant to transformations in $GL(n)$, but none of them are robust to outliers encountered in the data for e.g., in tensor interpolation. For $d_R(.,.)$, neither its mean nor its median is in closed form, which makes it computationally very expensive as the population size and the dimensionality of the space increases. Here, the Karcher mean denoted by P_M^* is defined as the minimizer of the sum of squared Riemannian distances and the median P_R^* is minimizer of the sum of Riemannian distances [9]

$$P_M^* = \arg\min_P \sum_{i=1}^m d_R^2(P,Q_i) \text{ and } P_R^* = \arg\min_P \sum_{i=1}^m d_R(P,Q_i).$$

For simplicity in notation, we denote $d_R^2(P,Q_i)$ by $d_M(P,Q_i)$, and $P_M^* = \arg\min_P \sum_{i=1}^m d_M(P,Q_i)$. Even though there are many algorithms (for example [8]–[10], [15]) to solve the geodesic mean and median, most of them adopt an iterative (gradient descent) method. Performing gradient descent on matrix manifolds can be tricky and rather complicated in the context of convergence issues (see [20]) and hence is not preferred over a closed form computation.

KL_s in (25) has a closed form mean [14], [16], which is $\mathcal{G}(\mathcal{A}(Q_1, \cdots, Q_m), \mathcal{H}(Q_1, \cdots, Q_m))$, where $\mathcal{A}$ is the arithmetic mean of $\{Q_i\}_{i=1}^m$, $\mathcal{H}$ is the harmonic mean of $\{Q_i\}_{i=1}^m$, and $\mathcal{G}$ is the geometric mean of $\mathcal{A}$ and $\mathcal{H}$. However, it can be shown that the mean computed using sum of squared KL_s divergences is not statistically robust since all tensors are treated equally. This is because, neither KL_s nor the sum of squared KL_ss are robust functions. This is demonstrated in the experimental results to follow. First we will observe the visual difference between the aforementioned divergences/distances. Fig. 2 shows the isosurfaces centered at the identity matrix with radii $r = 0.1$, 0.5 and 1 respectively. From left to right are $d_F(P,I) = r$, $d_R(P,I) = r$, $KL_s(P,I) = r$ and $tKL(P,I) = r$. These figures indicate the degree of anisotropy in the divergences/distances.
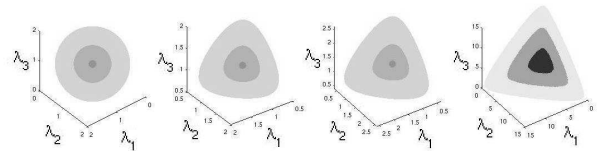

Fig. 2. The isosurfaces of $d_F(P,I) = r$, $d_R(P,I) = r$, $KL_s(P,I) = r$ and $tKL(P,I) = r$ shown from left to right. The three axes are eigenvalues of P.

B. Piecewise constant DTI segmentation

Given a noisy diffusion tensor image (DTI) $T_0(x)$ – a field of positive definite matrices, our model for DTI segmentation is based on the Mumford-Shah functional [21],

$$E(C,T) = \alpha \int_\Omega \delta(T_0(x), T(x))dx + \int_{\Omega-C} |\nabla T(x)|dx + \beta|C|, \quad (27)$$

where α and β are control parameters, δ is the tKL divergence same as (20). Ω is the region of the tensor field, $T(x)$ is an approximation to $T_0(x)$, which can be discontinuous only

along C. However, in a simplified segmentation model, a field $T_0(x)$ can be represented by piecewise constant regions [21], [22]. Therefore, we consider the following binary segmentation model for DTI,

$$E(C,T_1,T_2) = \int_R \delta(T_0(x),T_1)dx + \int_{R^c} \delta(T_0(x),T_2)dx + \beta|C|, \quad (28)$$

T_1 is the t-center of DTI for the region R inside the curve C and T_2 is the t-center of the DTI for the region R^c outside C,

$$T_1 = \left(\sum_{Q_i \in R} w_i Q_i^{-1} \right)^{-1}, w_i = \frac{\mu_i}{\sum_{Q_j \in R} \mu_j}. \quad (29)$$

$$T_2 = \left(\sum_{Q_i \in R^c} w_i Q_i^{-1} \right)^{-1}, w_i = \frac{\mu_i}{\sum_{Q_j \in R^c} \mu_j}. \quad (30)$$

The Euler Lagrange equation of (28) is

$$(\delta(T_0(x),T_1) - \delta(T_0(x),T_2) + \beta\kappa)\mathbf{N} = 0, \quad (31)$$

where κ is the curvature, $\kappa = \nabla \cdot \frac{\nabla C}{|\nabla C|}$, and $\mathbf{N}$ is the normal of C, $\mathbf{N} = \nabla C$. C can be updated iteratively according to the following equation

$$\frac{\partial C}{\partial t} = -(\delta(T_0(x),T_1) - \delta(T_0(x),T_2) + \beta\kappa)\mathbf{N}. \quad (32)$$

At each iteration, we will fix C, update T_1 and T_2 according to (29) and (30), and then freeze T_1 and T_2 to update C.

In the level set formulation of the active contour [23], let ϕ be the signed distance function of C and choose it to be negative inside and positive outside. Then the curve evolution equation (32) can be reformulated using the level set framework

$$\frac{\partial \phi}{\partial t} = (\delta(T_0(x),T_1) - \delta(T_0(x),T_2) + \beta\nabla \cdot \frac{\nabla\phi}{|\nabla\phi|})|\nabla\phi|. \quad (33)$$

C. Piecewise smooth DTI segmentation

For complicated DTI images, the piecewise constant assumption does not hold. Therefore, we have to turn to the more general model, the piecewise smooth segmentation. In this paper, we follow Wang et al's [16] model but replace their divergence (KL_s) with tKL, resulting in the following functional

$$E(C,T) = \alpha \int_R \delta(T_0(x),T_R(x))dx + \int_{R^c} \delta(T_0(x),T_{R^c}(x))dx$$
$$+ \int_{\Omega-C} p(T)(x)dx + \beta|C|,$$

the third term measures the lack of smoothness of the field using the Dirichlet integral [24], giving us the following curve evolution equation:

$$\frac{\partial C}{\partial t} = \left(\alpha\delta(T_0(x),T_{R^c}) + \sum_{y \in N_{R^c}(x)} \delta(T_{R^c},T_{R^c}(y)) \right)\mathbf{N}$$
$$- \left(\alpha\delta(T_0(x),T_R) + \sum_{y \in N_R(x)} \delta(T_R,T_R(y)) - \beta\kappa \right)\mathbf{N}.$$

where, N_R and N_{R^c} are the sets of x's neighboring pixels inside and outside of the region R respectively. In the discrete case, one can use appropriate neighborhoods for 2D and 3D. We apply the two stage piecewise smooth segmentation algorithm in [16] to numerically solve this evolution.

V. EXPERIMENTAL RESULTS

We performed two sets of experiments here, (i) tensor interpolation and (ii) tensor field segmentation. The experimental results are compared with those obtained by using other divergence measures discussed above.

A. Tensor interpolation

SPD tensor interpolation is a crucial component of DT-MRI analysis involving segmentation, registration and atlas construction [25], where a robust distance/divergence measure is very desirable. First, we perform tensor interpolation on a set of tensors with noise and outliers. We fix an SPD tensor $\bar{G}$ as the ground truth tensor and generate noisy tensors G from it by using a the Monte Carlo simulation method as was done in [26], [27]. This entails, given a b-value (we used $1200s/mm^2$) a zero gradient baseline image $\bar{S}_0$, and six vectors $\{q_i\}_{i=1}^6$, the magnitude of the noise free complex-valued diffusion weighted signal is given by, $\bar{S}_i = \bar{S}_0 \exp(-bq_i'\bar{G}q_i)$. We add Gaussian distributed (mean zero and variance $\sigma^2 = 0.1$) noise to the real and imaginary channels of this complex valued diffusion weighted signal and take its magnitude to get $S_i = \bar{S}_0 \exp(-bq_i'Gq_i)$. S_i then has a Rician distribution with parameters $(\sigma, \bar{S}_i)$. G is obtained by fitting the $\{S_i\}_{i=1}^6$ using Log-Euclidean tensor fitting. To generate outliers, we first generate a 3×3 matrix Z, with each entry drawn from the normal distribution, and the outlier tensor G is given by $G = \exp(ZZ')$. We compute the eigen-vectors U and eigen-values V of G, i.e. $G = Udiag(V)U'$, and then rotate U by a rotation matrix $r(\alpha,\beta,\gamma)$ in 3D, to get U_o, where α, β and γ are uniformly distributed in $[0,\pi/2]$. Then the outlier is given by $G_o = U_o diag(randperm(V))U_o'$. To generate an image, the same process is repeated at every voxel.

The tKL t-center, the geometric median, the geometric mean, the KL_s mean and the log-Euclidean mean M for 21 SPD tensors along with 0, 5, 10, 15, and 20 outliers are then computed. The difference between the various means/medians, and the ground truth $\bar{G}$ are measured in four ways, using the Frobenius norm $\|\bar{G} - M\|_F$, the ℓ_∞ distance $\|\bar{G} - M\|_\infty$, angle between the principal eigenvectors of M and $\bar{G}$, and the difference of fractional anisotropy index (FA) ([28], [29]) for M and $\bar{G}$, i.e., $|FA(M) - FA(\bar{G})|$. The results are shown in Table II from left to right, top to bottom.

From the tables, we can see that the t-center yields the best approximation to the ground truth, and faithfully represents the directionality and anisotropy. The robustness of t-center over others such as the Riemannian mean and median, KL_s mean, and log-Euclidean mean is quite evident from this table. Even though geometric median seems to be competitive to the t-center obtained using tKL in the case of lower percentage of outliers, *the geometric median computation however is much slower than that of t-center computation*. This is because

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

7

h

TABLE II

FROBENIUS NORM DISTANCE d_F, ℓ_∞ DISTANCE, ANGLE BETWEEN THE PRINCIPLE EIGENVECTORS OF THE GROUND TRUTH AND THE EIGENVECTOR OF THE MEAN/MEDIAN FOUND BY THE DIVERGENCES, AND THE DIFFERENCE BETWEEN THE FA OF THE GROUND TRUTH AND THE COMPUTED MEAN/MEDIAN ARE SHOWN FROM LEFT TO RIGHT, TOP TO BOTTOM. THERE ARE 21 ORIGINAL TENSORS AND THE NUMBER OF OUTLIERS IS SHOWN IN THE TOP ROW OF EACH TABLE.

Errors in computed mean measured by Frobenius norm d_F					
♯ of outliers	0	5	10	15	20
tKL	0.8857	1.0969	1.1962	1.3112	1.4455
d_R	1.3657	1.4387	1.5348	1.6271	1.7277
d_M	1.4291	1.5671	1.8131	2.1560	2.5402
KLs	3.4518	3.5744	3.7122	4.1040	4.3235
LE	1.0638	1.5721	1.6249	1.6701	1.8227

Errors in computed mean measured by ℓ_∞					
♯ of outliers	0	5	10	15	20
tKL	0.6669	0.8687	1.0751	1.1994	1.5997
d_R	0.8816	1.3498	1.5647	1.8826	2.1904
d_M	1.3056	1.4160	1.6755	1.9390	2.5153
KLs	4.1738	4.3352	4.3463	4.3984	4.4897
LE	0.8768	1.2843	1.6947	1.9418	2.2574

Errors in computed angle					
♯ of outliers	0	5	10	15	20
tKL	0.1270	0.3875	0.6324	1.8575	2.9572
d_R	0.4218	0.9058	8.9649	17.4854	44.9157
d_M	0.8147	1.1985	10.1576	21.8003	43.7922
KLs	0.8069	0.9134	14.9706	26.1419	44.9595
LE	0.6456	0.9373	8.9937	17.0636	44.7635

Errors in computed FA					
♯ of outliers	0	5	10	15	20
tKL	0.0635	0.0738	0.0753	0.0792	0.0857
d_R	0.0983	0.1529	0.1857	0.1978	0.2147
d_M	0.1059	0.1976	0.2008	0.2193	0.2201
KL_s	0.0855	0.1675	0.2005	0.2313	0.2434
LE	0.0972	0.1513	0.1855	0.1968	0.2117

the t-center for tKL has a closed form while the geometric median does not. Table III shows the CPU time to find the tensor mean/median using different divergences. We use 1000 SPD tensors, along with 0, 10, 100, 500 and 800 outliers respectively. All tensors are generated in the same way as described in the first experiment. The time is averaged by repeating the experiment 10 times on a PC, with Intel(R) Core(TM) 2 Duo CPU P7370, 2GHz, 4GB RAM, on 32-bit Windows Vista OS. Table II and III depict the superior robustness to outliers and the computational efficiency in estimating the tKL t-center for SPD tensor interpolation in comparison to its rivals.

TABLE III
TIME (SECONDS) SPENT IN FINDING THE MEAN/MEDIAN USING DIFFERENT DIVERGENCES. THERE ARE 1000 ORIGINAL TENSORS AND THE NUMBER OF OUTLIERS IS SHOWN IN THE TOP ROW.

time					
Divergences\♯ of outliers	0	10	100	500	800
tKL	0.02	0.02	0.02	0.03	0.03
KL_s	0.03	0.03	0.03	0.04	0.04
d_R	0.58	1.67	4.56	93.11	132.21
d_M	0.46	1.42	3.13	72.74	118.05
LE	0.02	0.02	0.03	0.03	0.03

Fig. 3. From left to right are initialization, intermediate step and final segmentation.

Fig. 4. Dice coefficient comparison for tKL, KL_s, d_R, d_M and LE segmentation of synthetic tensor field with increasing level (x-axis) of noise.

B. Tensor field segmentation

We now describe experiments on segmentation of synthetic and real DTI images.

1) Segmentation of synthetic tensor fields: The first synthetic tensor field is composed of two types of homogeneous tensors. Fig. 3 depicts the synthetic data and the segmentation results. We added different levels of noise to the tensor field using the method described in section V-A, segmented it using the aforementioned divergences and compared the results using the dice coefficient [30]. We also added different percentages of outliers to the tensor fields and segmented the resulting tensor fields. Fig. 4 depicts the comparison of segmentation results from different methods using the dice coefficient, for varying noise levels with σ varying
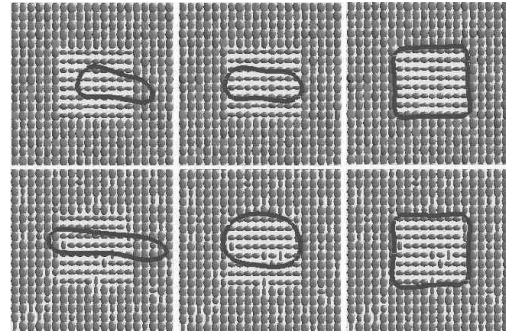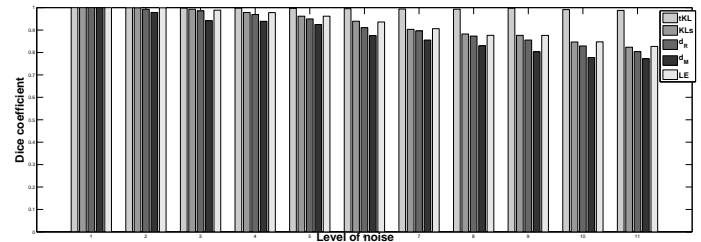
from $0, 0.02, 0.04, \cdots, 0.2$). Fig. 5 displays the comparison of dice coefficient with different percentage $(0, 5, 10, \cdots, 50)$ of outliers. The results show that even in the presence of large amounts of noise and outliers, tKL yields very good segmentation results in comparison to rivals.

2) Segmentation of DTI images: In this section, we present segmentation results on real DTI images from a rat spinal cord, an isolated rat hippocampus and a rat brain. The data were acquired using a PGSE with TR=1.5s, TE=28.3ms, bandwidth=35Khz, 21 diffusion weighted images with a b-value of $1250s/mm^2$ were collected. A 3×3 diffusion tensor in each DTI image is illustrated as an ellipsoid [16], whose axes' directions, and lengths correspond to its eigen-vectors, and
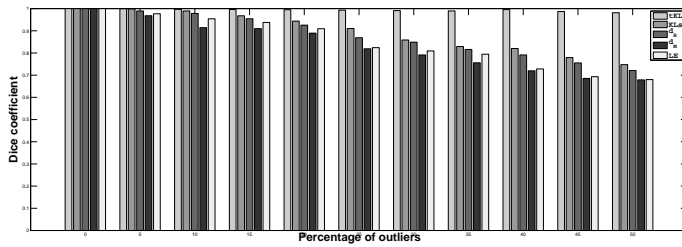
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

8

Fig. 5. Dice coefficient comparison for tKL, KL_s, d_R, d_M and LE segmentation of synthetic tensor field with increasing percentage (x-axis) of outliers.

(a)   (b)   (c)   (d)

(e)

Fig. 7. tKL segmentation of a 3D rat corpus callosum. (a),(b),(c),(d) A 2D slice of the corresponding evolving surface, from left to right are initialization, intermediate steps and final segmentation. (e) A 3D view of the segmentation result.

eigen-values respectively. The same initialization is used for each divergence based segmentation. We let all the divergence based methods run until convergence, record their time and compare the results.

We apply the piecewise constant segmentation model on a single slice (108×108) of the rat spinal cord, and apply the piecewise smooth segmentation model on the molecular layer from single slices of size (114×108) for rat corpus callosum (CC) and (90×90) for the rat hippocampus respectively. Fig. 6(a)(b)(c) show the initialization, Fig. 6(d) shows the segmentation results, and Table IV records their execution time. The results confirm that when compared to other divergences, tKL yields a more accurate segmentation in a significantly shorter amount of CPU time.

TABLE IV
TIME (SECONDS) COMPARISON FOR SEGMENTING THE RAT SPINAL CORD, CORPUS CALLOSUM AND HIPPOCAMPUS USING DIFFERENT DIVERGENCES.

Divergences	tKL	KLs	d_R	d_M	LE
Time for Cord	33	68	72	81	67
Time for CC	87	183	218	252	190
Time for hippocampus	159	358	545	563	324

Apart from segmentation in 2D slices, we also demonstrate 3D DTI image segmentation using the proposed divergence. Fig. 7 depicts the process of segmenting rat corpus callosum ($114 \times 108 \times 11$) using the piecewise constant segmentation model. The result demonstrates that tKL can segment this white matter bundle quite well.

VI. CONCLUSIONS

In this paper, we developed a novel divergence dubbed the *total Bregman divergence (TBD)*, that is intrinsically robust to noise and outliers. The goal here was not simply to develop yet another divergence measure but to develop an intrinsically robust divergence. This was achieved via alteration of the basic notion of divergence which measures the ordinate distance between its convex generator and its tangent approximation to the orthogonal distance between the same. This basic idea parallels the relationship between least-squares and total least squares but the implications are far more significant in the field of information geometry and its applications since the entire class of the well known Bregman divergences have been redefined and some of their theoretical properties studied. Specifically, we derived an explicit formula for the t-center which is the TBD-based median, that is robust to outliers. In
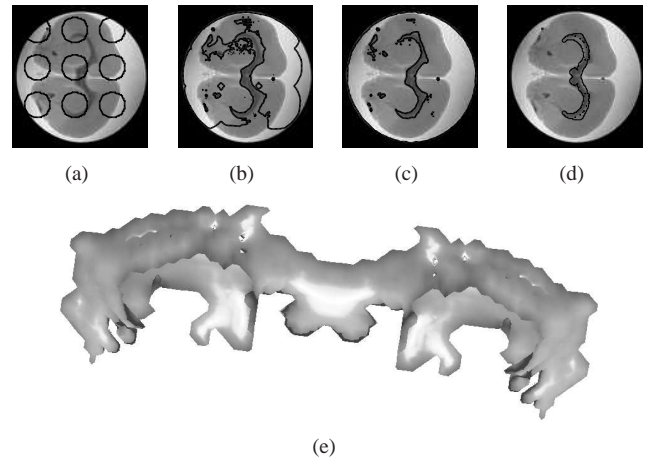
the case of SPD tensors, the t-center was shown to be $SL(n)$ invariant. However, the story is not yet complete and further investigations are currently underway.

The robustness (to outliers) property of TBD was demonstrated here via applications to SPD tensor field interpolation and segmentation. The results favorably demonstrate the competitiveness of our newly defined divergence in comparison to existing methods not only in terms of robustness, but also in terms of computational efficiency and accuracy as well. Our future research will focus on applications of TBD to atlas construction, registration and other such tasks applied to DWMRI acquired from human brains.

REFERENCES

[1] J. A. Thomas and T. M. Cover, *Elements of Information Theory*. Wiley Interscience, 1991.
[2] P. Mahalanobis, "On the generalised distance in statistics," *Proceedings National Institute of Science, India*, 1936.
[3] S. Amari, *Differential-Geometrical Methods in Statistics*. Springer Berlin Heidelberg, 1985.
[4] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, "Clustering with Bregman divergences," *J. Mach. Learn. Res.*, vol. 6, pp. 1705–1749, 2005.
[5] F. Nielsen and R. Nock, "On the smallest enclosing information disk," *Inf. Process. Lett.*, vol. 105, pp. 93–97, 2008.
[6] P. J. Basser, J. Mattiello, and D. Lebihan, "Estimation of the effective self-diffusion tensor from NMR spin echo," *J. Magn. Reson.*, vol. 103, 1994.
[7] B. Jian and B. C. Vemuri, "A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI," *IEEE Trans. Med. Imaging*, vol. 26, no. 11, 2007.
[8] P. T. Fletcher and S. Joshi, "Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors," *CVAMIA*, 2004.
[9] P. T. Fletcher, S. Venkatasubramanian, and S. Joshi, "The geometric median on Riemannian manifolds with application to robust atlas estimation," *NeuroImage*, vol. 45, 2008.
[10] C. Lenglet, M. Rousson, and R. Deriche, "Segmentation of 3D probability density fields by surface evolution: Application to diffusion MRI," *MICCAI*, 2004.
[11] Z. Wang, B. C. Vemuri, Y. Chen, and T. H. Mareci, "A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI," *IEEE Trans. Med. Imag.*, vol. 23, 2004.
[12] Z. Wang and B. C. Vemuri, "Tensor-field segmentation using a region-based active contour model," *ECCV*, 2004.
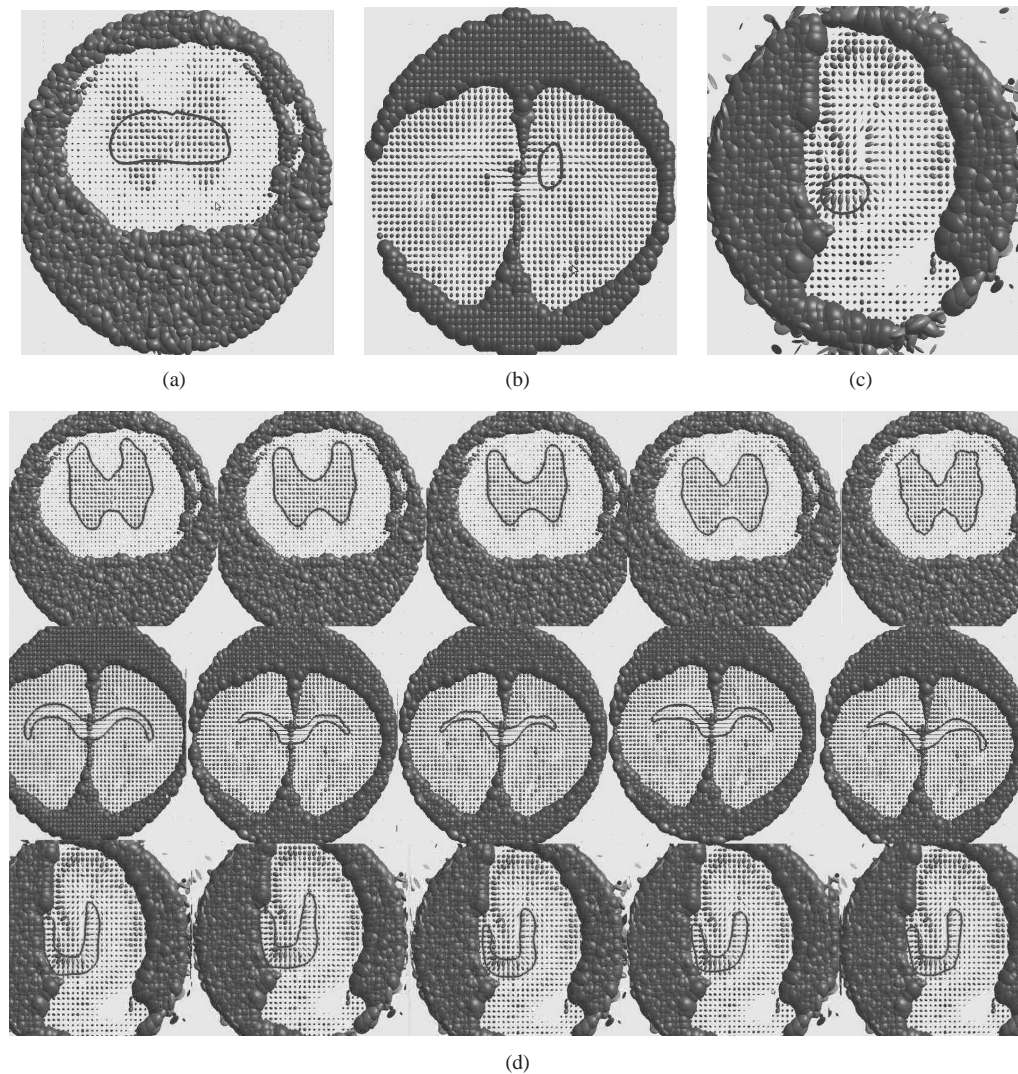[13] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Commun. Pure Appl. Math.*, vol. 30, 1977.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

9

(a) (b) (c)

(d)

Fig. 6. (a)(b)(c) show the initialization. (d) shows the segmentation results, from left to right, using tKL, KL_s, d_R, d_M and LE.

[14] M. Moakher and P. G. Batchelor, *Symmetric Positive-Definite Matrices: From Geometry to Applications and Visualization*. Visualization and Processing of Tensor Fields, Springer, 2006.

[15] X. Pennec, "Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements," *J. Math. Imag. Vis.*, vol. 25, 2006.

[16] Z. Wang and B. Vemuri, "DTI segmentation using an information theoretic tensor dissimilarity measure," *IEEE Trans. Med. Imag.*, vol. 24, 2005.

[17] F. Nielsen and R. Nock, "Sided and symmetrized Bregman centroids," *IEEE Trans. Inf. Theory*, 2009.

[18] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Log-Euclidean metrics for fast and simple calculus on diffusion tensors," *Magn. Reson. Med.*, vol. 56, pp. 411–421, 2006.

[19] B. A. Frigyik, S. Srivastava, and M. R. Gupta, "Functional Bregman divergence," *IEEE Int. Symp. Inf. Theory*, 2008.

[20] P. A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

[21] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational-problems," *Comm. Pure Appl. Math.*, vol. 42, 1989.

[22] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Processing*, vol. 10, 2001.

[23] R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modeling with front propagation: A level set approach," *IEEE Trans. on PAMI*, vol. 17, no. 2, 1995.

[24] F. Helein, "Harmonic maps, conservation laws and moving frames," 2002.

[25] A. Barmpoutis, B. C. Vemuri, T. M. Shepherd, and J. R. Forder, "Tensor splines for interpolation and approximation of DT-MRI with applications to segmentation of isolated rat hippocampi," *IEEE Trans. Med. Imag.*, vol. 26, 2007.

[26] O. Pasternak, N. Sochen, and P. J. Basser, "The effect of metric selection on the analysis of diffusion tensor MRI data," *NeuroImage*, 2010.

[27] S. Pajevic and P. Basser, "Parametric and non-parametric statistical analysis of DT-MRi data," *J. Magn. Reson.*, vol. 161, 2003.

[28] N. G. Papadakis, D. Xing, G. C. Houston, J. M. Smith, M. I. Smith, M. F. James, A. A. Parsons, C. L. Huang, L. D. Hall, and T. A. Carpenter, "A study of rotationally invariant and symmetric indices of diffusion anisotropy," *Magn. Reson. Imag.*, 1999.

[29] C. Pierpaoli and P. Basser, "Toward a quantitative assessment of diffusion anisotropy," *Magn. Reson. Med.*, vol. 36, 1996.

[30] J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, A. W. Toga, and P. M. Thompson, "Comparison of adaboost and support vector machines for detecting alzheimers disease through automated hippocampal segmentation," *IEEE Trans. Med. Imag.*, 2007.