



ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg  
NeuroImage xx (2008) xxx–xxx

## Bayesian template estimation in computational anatomy

Jun Ma,<sup>a,\*</sup> Michael I. Miller,<sup>b</sup> Alain Trouvé,<sup>c</sup> and Laurent Younes<sup>d</sup>

<sup>a</sup>Center for Imaging Science and Department of Biomedical Engineering, The Johns Hopkins University, 320 Clark Hall, Baltimore, MD 21218, USA

<sup>b</sup>Center for Imaging Science and Department of Biomedical Engineering, The Johns Hopkins University, 301 Clark Hall, Baltimore, MD 21218, USA

<sup>c</sup>CMLA, Ecole Normale Supérieure de Cachan, 61 Avenue du Président Wilson, F-94 235 Cachan CEDEX, France

<sup>d</sup>Center for Imaging Science and Department of Applied Math and Statistics, The Johns Hopkins University, 3245 Clark Hall, Baltimore, MD 21218, USA

Received 7 June 2007; revised 17 March 2008; accepted 27 March 2008

**Templates play a fundamental role in Computational Anatomy. In this paper, we present a Bayesian model for template estimation. It is assumed that observed images  $I_1, I_2, \dots, I_N$  are generated by shooting the template  $J$  through Gaussian distributed random initial momenta  $\theta_1, \theta_2, \dots, \theta_N$ . The template is  $J$  modeled as a deformation from a given hypertemplate  $J_0$  with initial momentum  $\mu$ , which has a Gaussian prior. We apply a mode approximation of the EM (MAEM) procedure, where the conditional expectation is replaced by a Dirac measure at the mode. This leads us to an image matching problem with a Jacobian weight term, and we solve it by deriving the weighted Euler–Lagrange equation. The results of template estimation for hippocampus and cardiac data are presented.**

© 2008 Elsevier Inc. All rights reserved.

**Keywords:** Template estimation; Computational anatomy; Bayesian; Weighted Euler–Lagrange equation

### Introduction

Computational Anatomy (CA) is the mathematical study of variability of anatomical and biological shapes. The framework has been pioneered by Grenander (1994) through the notion of deformable templates. Given a template  $I_{\text{temp}}$ , the group of diffeomorphisms  $\mathcal{G}$  acts on it to generate an orbit  $\mathcal{I} = \mathcal{G} \cdot I_{\text{temp}}$ , a whole family of new objects with similar structure as  $I_{\text{temp}}$ . Hence, one can model elements in the orbit  $\mathcal{I}$  via diffeomorphic transformations.

Templates play an important role in CA. They are usually used to generate digital anatomical atlases and act as a reference when computing shape variability. Often, templates have been chosen to be manually selected “typical” observed images. It is however preferable to build a template based on statistical properties of the

observed population. There has been by now several publications addressing the issue of shape averaging over a dataset. In this context, the average is based on metric properties of the space of shapes; assuming a distance in shape space is given, the average of a set of shapes is a minimizer of the sum of square distances to each element of the set (Fréchet or Karcher mean). When the shape space is modeled as a Riemannian manifold, a local minimum of this sum of squared distances must be such that the sum of initial velocities of the geodesics between the average and each of the elements in the set vanishes. This leads to the following averaging procedure (sometimes called *procrustean averaging*) which consists of (i) starting with an initial guess of the average, (ii) computing all the geodesics between this current average and each element in the set of shapes, (iii) averaging their initial velocities and (iv) displacing the current average to the endpoint of the geodesic starting with the initial velocity, this being iterated until convergence (Guimond et al., 2000; Mio et al., 2007; Fletcher et al., 2004; Le and Kume, 2000; Pennec, 2006). A different variational definition of the average has been provided in (Joshi et al., 2004; Avants and Gee, 2004b,a). In the present work, however, we do not build the template as a metric average, but as the central component of a generative statistical model for the anatomy. This is reminiscent of the construction developed in (Allasonnière et al., 2007) for linear models of deformations, and in (Glasbey and Mardia, 2001) for thin-plate models of points set.

In our application of Grenander's pattern theory, anatomical shapes are modeled as an orbit under the action of the group of diffeomorphisms. Because of this, diffeomorphisms play an important role in our statistical model. The nonlinear space of diffeomorphisms can be studied as an infinite dimensional Riemannian manifold, on which, with a suitable choice of metric, geodesic equations are described by a *momentum conservation law* (Arnold, 1966, 1978; Holm et al., 1998; Marsden and Ratiu, 1999). In our context, the methodology of *geodesic shooting* (Vailliant et al., 2004; Miller et al., 2006) relies on this conservation law to derive statistical models on diffeomorphisms and deformable objects. Through the geodesic equations, the flow at any point along the geodesic is completely

\* Corresponding author.

E-mail addresses: junma@cis.jhu.edu (J. Ma), mim@cis.jhu.edu (M.I. Miller), trouve@cmla.ens-cachan.fr (A. Trouvé), laurent.younes@jhu.edu (L. Younes).

Available online on ScienceDirect (www.sciencedirect.com).

determined (once a template is fixed) by the momentum at the origin. The initial momentum therefore provides a linear representation of the nonlinear diffeomorphic shape space in a local chart around the template, to which linear statistical analysis can be applied. Note that metric averaging can be reconnected to geodesic shooting (Vailliant et al., 2004; Helm et al., 2006). In this case, the algorithm is: first, compute the geodesic from a given template  $I_0$  to several target images and obtain the initial momenta of each transformation; second, compute the mean initial momentum  $\bar{m}$ ; then, shoot  $I_0$  with initial momentum  $\bar{m}$  to get a new image  $\bar{I}$ ; iterate this procedure. This approach was used in landmark matching (Vailliant et al., 2004), 3D average digital atlas construction (Beg and Khan, 2006) and quantifying variability in heart geometry (Helm et al., 2006).

In this paper, we introduce a random statistical model on the initial momentum to represent random deformations of a template. The generative model we use combines this deformation with some observation noise. We will then develop a strategy to estimate the template from observations, based on a mode approximation of the EM algorithm (MAEM), under a Bayesian framework.

The paper is organized as follows. We first provide some background material and notation related to diffeomorphisms and their use in computational anatomy. We then discuss template estimation and detail the statistical model and the implementation of the MAEM procedure. This will require in particular introducing an extension of the LDDMM algorithm (Beg et al., 2005; Miller et al., 2003) to the case where the data attachment term has a nonstationary spatial weight. We finally provide experimental results, with a comparison with a simplified, non-Bayesian, approach.

## Background

Let the background space  $\Omega \subset \mathbb{R}^d$  be a bounded domain on which the images are defined. To a template  $I_{\text{temp}}$  corresponds the orbit  $\mathcal{I} = \{I_{\text{temp}} \circ g^{-1} : g \in \mathcal{G}\}$  under the group of diffeomorphisms  $\mathcal{G}$ . For any two anatomical images  $J, I \in \mathcal{I}$ , there exists a set of diffeomorphisms (denote  $g$  as an arbitrary element in the set) that registers the given images:  $I = J \circ g^{-1}$ . Following (Dupuis et al., 1998; Trouvé, unpublished), when we define the orbit  $\mathcal{I}$ , we restrict to diffeomorphisms that can be generated as flows  $g_t$ ,  $t \in [0, 1]$  controlled by a velocity field  $v_t$ , with the relation

$$\frac{\partial g_t}{\partial t}(x) = v_t(g_t(x)), \quad x \in \Omega, \quad t \in [0, 1] \quad (1)$$

with initial condition  $g_0 = id$ . To ensure that the ODEs generate diffeomorphisms, the vector fields are constrained to be sufficiently smooth (Dupuis et al., 1998; Trouvé, unpublished). More specifically, they are assumed to belong to  $(V, \|\cdot\|_V)$ , a Hilbert space with squared norm defined as  $\|v\|_V = (Av, v)$  through an operator  $A: V \rightarrow V^*$ , where  $V^*$  is the dual space of  $V$ . For  $v \in V$ ,  $Av$  can be considered as a linear form on  $V$  (a mapping from  $V$  to  $\mathbb{R}$ ) through the identification  $(Av, w) = \langle v, w \rangle_V$ , where  $(Av, w)$  is the standard notation for a linear form  $Av$  applied to  $w$ . Interpreting  $\|v\|_V^2 = (Av, v)$  as an energy,  $Av$  will be called the *momentum* associated to the velocity  $v$ . We assume that  $V$  can be embedded in a space of smooth functions, which makes it a reproducing kernel Hilbert space with kernel  $K = A^{-1}: V^* \rightarrow V$ .

The geodesics in the group of diffeomorphisms are time-dependent diffeomorphisms  $t \rightarrow g_t$  defined by Eq. (1) such that the integrated energy

$$\int_0^1 \|v_t\|_V^2 dt \quad (2)$$

is minimal with fixed boundary conditions  $g_0$  and  $g_1$ . The image matching problem between  $J$  and  $I$  is formalized as the search for the optimal geodesic starting at  $g_0 = id$  such that  $I = J \circ g_1^{-1}$ . From this is derived the inexact matching problem, which consists in finding a time-dependent vector field  $v_t$  solution of the problem

$$\hat{v} = \underset{v: \dot{g}_t = v_t(g_t)}{\operatorname{argmin}} \left( \int_0^1 \|v_t\|_V^2 dt + \frac{1}{\sigma^2} \|J \circ g_1^{-1} - I\|_2^2 \right). \quad (3)$$

Geodesics are characterized by the following Euler equation (sometimes called EPDiff; Holm et al., 1998), which can be interpreted as a conservation equation for the momentum  $Av$  (Arnold, 1966, 1978). The equation is

$$\frac{\partial Av_t}{\partial t} + (Dv_t)^* Av_t + \operatorname{div}(v_t) Av_t + D(Av_t)v_t = 0. \quad (4)$$

The conservation of momentum is described as follows. Let  $w_0$  be a vector field, and  $w_t = Dg_t(g_t^{-1})w_0(g_t^{-1})$  be the transported vector field along the geodesic. Then, if  $Av_t$  satisfies Eq. (4), we have:

$$\frac{\partial}{\partial t}(Av_t, w_t) = 0. \quad (5)$$

This implies  $Av_t = (Dg_t^{-1})^* Av_0(g_t^{-1}) |Dg_t^{-1}|$ ,  $t \in [0, 1]$ , meaning that the geodesic evolution in the orbit  $J \circ g_t^{-1}$  depends only on  $J$  and the momentum at time 0. The solution of Eq. (3) satisfies this property in an even simpler form, which explicitly provides the momentum  $Av_t$  in function of the deformed images and the diffeomorphism  $g_t$  (Beg et al., 2005; Miller et al., 2003). Moreover, Eq. (5) can be shown to have singular solutions that propagate over time. In fluid mechanics, EPDiff is used to model the propagation of waves on shallow water. In our context, it provides a very simple form to finitely generate models of deformation (see Maximization via the Euler equation with Jacobian weight).

Hence, the nonlinear diffeomorphic shapes can be represented by the initial momenta which lie on a linear space (the dual of  $V$ ). This provides a powerful vehicle for statistical analysis of shapes. In this paper, we investigate a statistical model of deformable template estimation using this property.

Note that, if  $\hat{v}$ , with associated diffeomorphism  $\hat{g}$ , is a (local) minimum of Eq. (3), it is also a (local) minimum of the problem:

$$\underset{v: \dot{g}_t = v_t(g_t), g_1 = \hat{g}_1}{\operatorname{argmin}} \left( \int_0^1 \|v_t\|_V^2 dt \right). \quad (6)$$

since the data term in Eq. (3) only depends on  $g_1$ . Therefore (since Eq. (6) is equivalent to the geodesic minimization on groups of diffeomorphisms),  $\hat{v}$  satisfies the EPDiff equation, which is therefore also relevant for inexact matching (as noticed in Miller et al., 2006).

## Methodology of template estimation

### Statistical model for the anatomy

In addition to the conservation of momentum discussed in the previous section, solutions of Eq. (3) have also their energy conserved (since they are geodesics):  $\|v_t\|_V^2$  is independent of time. Because of this, problem Eq. (3) is equivalent to

$$\hat{v} = \underset{v: \dot{g}_t = v_t(g_t)}{\operatorname{argmin}} \left( \|v_0\|_V^2 + \frac{1}{\sigma^2} \|J \circ g_1^{-1} - I\|_2^2 \right). \quad (7)$$

where  $v_0$  is the initial velocity. So the minimization is now restricted to time-dependent vector fields  $v$  that satisfy Eq. (4). The minimized expression may formally be interpreted as a  $I$  joint log-likelihood for the initial velocity and observed image in which would be a random field (with  $V$  as a reproducing space) generating, via Eq. (4), a diffeomorphism  $g_1$ , and  $I$  would be obtained from the deformed template by the addition of a white noise.

This is essentially the model we adopt in this paper, under a discrete form that will be more amenable to rigorous computation. Recall that we have defined the *duality operator*  $A$  on  $V$  as associating to  $v$  in  $V$  in the linear form  $Av \in V^*$  defined by  $(Av, w) = \langle v, w \rangle_V$ . By the Riesz representation theorem,  $A$  is invertible with inverse  $A^{-1} = K$ . The discretization will be done on the momentum,  $m = Av$ , instead of the velocity field  $v$ .

Note that the sets  $V$  and  $V^*$  are isometric if  $V^*$  is equipped with the product  $\langle m, \tilde{m} \rangle_{V^*} = (m, K\tilde{m})$  for  $m, \tilde{m} \in V^*$ . Therefore, the norm of initial velocity is equal to the norm of corresponding initial momentum, that is, for  $m_0 = Av_0$ , we have

$$(Av_0, v_0) = \|v_0\|_V^2 = \|m_0\|_{V^*}^2 = (m_0, Km_0). \quad (8)$$

Formally again, this may be interpreted as the log-likelihood of a Gaussian distribution on  $V^*$  with covariance operator  $A = K^{-1}$ , characterized by the property that, for any  $w \in V$ ,  $(m_0, w)$  is a centered Gaussian distribution with variance

$$E\{(m_0, w)^2\} = (Aw, w) = \|w\|_V^2. \quad (9)$$

We now discuss a discrete version of this random field. For  $x, a \in \mathbb{R}^d$ , denote  $a \odot \delta_x$  the linear form  $w \mapsto (a \odot \delta_x, w) := a^T w(x)$ . Noting that  $K(a \odot \delta_x) \in V$  is, by definition, a vector field on  $V$  that depends linearly on  $a$ , we make the abuse of notation

$$K(a \odot \delta_x)(y) = K(y, x)a$$

where  $K(y, x)$  is a  $d$  by  $d$  matrix (the reproducing kernel of  $V$ ). It can be checked that  $K(x, y) = K(y, x)^T$  and

$$\langle K(\cdot, y)a, K(\cdot, z)b \rangle_V = a^T K(y, z)b.$$

We model the random momentum  $\theta$  as a sum of such measures

$$\theta = \sum_{i=1}^S a_i \odot \delta_{x_i} \quad (10)$$

where  $x_1, x_2, \dots, x_S \in \Omega$  form a set of fixed (deterministic) points (for example, the grid supporting the image discretization) and  $a = (a_1, a_2, \dots, a_S)$  are random variables such that  $a \sim \mathcal{N}(0, \Sigma)$  in  $\mathbb{R}^{Sd}$ . We want to choose  $\Sigma$  consistently with our formal interpretation of Eq. (8). For this, we can compute, for  $w \in V$

$$(\theta, w) = \sum_{i=1}^S a_i^T w(x_i) \quad (11)$$

which is a Gaussian with mean 0 and variance

$$\sum_{i,j=1}^S w(x_i)^T \Sigma_{ij} w(x_j)$$

where  $\Sigma_{ij}$  is the  $d$  by  $d$  matrix  $E(a_i a_j^T)$ . We want to compare this to Eq. (9), and in particular ensure that both expressions coincide

when  $w = K(\cdot, x_k)b$  for some  $b \in \mathbb{R}^d$  and  $k \in \{1, \dots, S\}$ . In this case, we have  $\|w\|_V^2 = b^T K(x_k, x_k)b$  yielding the constraint: for all  $k$

$$K(x_k, x_k) = \sum_{i,j=1}^S K(x_k, x_i) \Sigma_{ij} K(x_j, x_k).$$

Define  $\tilde{K}_{ij} = K(x_i, x_j)$  and assume that the block matrix  $\tilde{K} = (\tilde{K}_{ij})$  is invertible. Then the equation above is equivalent to  $\Sigma = \tilde{K}^{-1}$ , which completely describes the distribution of  $(a_1, \dots, a_S)$ .

The probability density function (p.d.f) of  $a$  is given by

$$p(a) = \frac{1}{Z} e^{-\frac{1}{2} a^T \Sigma^{-1} a} = \frac{1}{Z} e^{-\frac{1}{2} a^T \tilde{K} a} \quad (12)$$

where  $A = (2\pi)^{Sd/2} / \sqrt{\det \tilde{K}}$ .

It is interesting to notice that we have, with  $\theta = \sum_{i=1}^S a_i \odot \delta_{x_i}$ ,

$$(\theta, K\theta) = \sum_{i,j=1}^S (a_i \odot \delta_{x_i}, K(\cdot, x_j) a_j) \quad (13)$$

$$\begin{aligned} &= \sum_{i,j=1}^S a_i^T K(x_i, x_j) a_j \\ &= a^T \tilde{K} a. \end{aligned} \quad (14)$$

So the p.d.f of  $\theta$  can be written as

$$p(\theta) = \frac{1}{Z} e^{-\frac{1}{2}(\theta, K\theta)} \mathbf{1}_{V^*(x)}(\theta) \quad (15)$$

where  $\mathbf{1}$  represents the indicator function and  $V^*(x) = \{\theta = \sum_{i=1}^S a_i \odot \delta_{x_i}, a_1, \dots, a_S \in \mathbb{R}^d\}$ . Since in the infinite dimensional space  $V^*$ , our model is a singular Gaussian distribution supported by the finite dimensional space  $V^*(x)$ . When restricted on this space, it is continuous with respect to the Lebesgue measure with a density provided by Eq. (15).

This describes the deformation part of the model, represented as the distribution of the initial momentum  $\theta$ . We can solve Eq. (4) with initial condition  $Av_0 = \theta$  from time  $t=0$  to  $t=1$  and integrate the velocity field  $t \mapsto v_t$  to obtain a diffeomorphism, that we shall denote  $g_\theta$ , at time  $t=1$ . Although we will not use this fact in our numerical methods, it is important to notice that  $g_0$  can be obtained from  $\theta$  (given by Eq. (10)) via the solution of a system of ordinary differential equations to which Eq. (4) reduces. It can be shown that this system has solutions over all times, so our model is theoretically consistent. With this model, the image  $J \circ g_\theta^{-1}$  is therefore a random deformation of the template (since  $\theta$  is random). We assume that the observed image  $I$  obtained from the deformed template after discretization and addition of noise. More precisely, denoting  $[J \circ g_\theta^{-1}] = \sum_{i=1}^S \delta_{x_i} J \circ g_\theta^{-1}$ , the observation  $I$  is a discrete image given by

$$I = [J \circ g_\theta^{-1}] + W, \quad W \sim \mathcal{N}(0, \sigma^2 Id). \quad (16)$$

The complete process is thus via the pair  $(\theta, I)$ . Our goal is, given observations  $I_1, \dots, I_N$  having the same distribution as  $I$  above, to estimate the template  $J$  and the noise variance  $\sigma^2$ .

#### Prior distribution on the template

We want to constrain the virtually infinite dimensional template estimation problem within a Bayesian strategy. We will introduce for this a *hypertemplate*,  $J_0$ , and describe  $J$  as a random deformation of  $J_0$ . The hypertemplate is given, usually provided

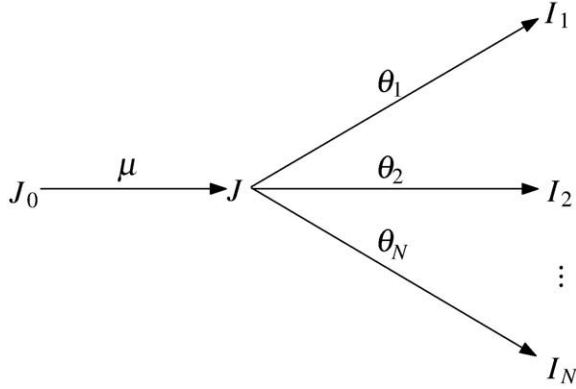


Fig. 1. The template  $J$  is to be estimated given hypertemplate  $J_0$  and observed images  $I_1, I_2, \dots, I_N$ . We model the template as  $J = J_0 \circ g_\mu^{-1}$ , where  $\mu$  is the random initial momentum. In the following EM algorithm,  $\mu$  is computed iteratively, with random initial momenta  $\theta_1, \theta_2, \dots, \theta_N$  being hidden variables.

by an anatomical atlas. The template  $J$  is modeled  $J = J_0 \circ g_\mu^{-1}$  as with  $\mu$  the initial momentum. We model  $\mu$  as a discrete momentum like in the previous section, with distribution

$$\pi(\mu) = \frac{1}{Z_\pi} e^{-\frac{1}{2}(\mu, K_\pi \mu)} 1_{V^*(x)}(\mu) \quad (17)$$

for a reproducing kernel  $K_\pi$ . In our experiments, we made the simple choice  $K_\pi = \lambda K$ , for some regularization parameter  $\lambda > 0$ .

The relations between  $J_0, J$  and  $I_1, I_2, \dots, I_N$  is illustrated in Fig. 1.

*Remark:* In this model,  $J_0$  and  $J$  are continuous.  $J_0$  is a given continuous function defined on  $\Omega \in \mathbb{R}^d$ , although it may be finitely generated (using a finite element representation, for example). Observations  $I_1, \dots, I_N$  are discrete images. From these images, we will estimate the initial momentum  $\mu$ , and further get the continuous template  $J = J_0 \circ g_\mu^{-1}$ . For simplicity, we will still denote, with a little abuse,  $J \circ g_{\theta_n}^{-1}$  to refer to its discretization  $[J \circ g_{\theta_n}^{-1}]$ .

*Remark:* The kernels  $K$  and  $K_\pi$  are important components of the model, and can be considered as infinite dimensional parameters. While not an impossible task, trying to estimate them would significantly complicate our procedure. Consequently, the kernels have been selected a priori, and left fixed during the estimation procedure.

### Template estimation with the MAEM algorithm

#### The complete-data log-likelihood

We here describe the estimation of parameters  $\mu$  (which uniquely describes the template) and  $\sigma^2$  based on the observation of  $\mathbf{I} = \{I_1, I_2, \dots, I_N\}$ . Our goal is to compute  $\arg\max_{\mu, \sigma^2} p_\sigma(\mu | \mathbf{I})$ . We let  $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$  denote the sequence of hidden initial momenta.

To use the EM algorithm we first write down the complete-data log-likelihood. The joint likelihood of the model given the observations is

$$\begin{aligned} p(\mu, \mathbf{I}, \Theta) &= \pi(\mu) p(\mathbf{I}, \Theta | \mu) \\ &= \frac{1}{Z_\pi} e^{-\frac{1}{2}(\mu, K_\pi \mu)} \prod_{n=1}^N \frac{1}{Z} e^{-\frac{1}{2}(\theta_n, K \theta_n) - \frac{1}{2\sigma^2} \|J_0 \circ g_\mu^{-1} \circ g_{\theta_n}^{-1} - I_n\|_2^2 - \frac{1}{2} S \log \sigma^2} \end{aligned} \quad (18)$$

where  $\|I - I'\|_2^2 = \sum_{i=1}^S (I(x_i) - I'(x_i))^2$ .

The EM algorithm generates a sequence  $(\mu^{(k)}, \sigma^{2(k)})$  according to the transition  $(E_{\mu^{(k)}, \sigma^{2(k)}})$  is the expectation under the assumption that the true parameters are  $\mu^{(k)}$  and  $\sigma^{2(k)}$ .

$$(\mu^{(k+1)}, \sigma^{2(k+1)}) = \arg\max_{\mu, \sigma^2} E_{\mu^{(k)}, \sigma^{2(k)}} (\log p_\sigma(\mu, \mathbf{I}, \Theta) | \mathbf{I}).$$

The maximization is decomposed into two steps implying a generalized EM algorithm:

$$\begin{aligned} \sigma^{2(k+1)} &= \arg\max_{\sigma^2} E_{\mu^{(k)}, \sigma^{2(k)}} (\log p_\sigma(\mu, \mathbf{I}, \Theta) | \mathbf{I}) \\ \mu^{(k+1)} &= \arg\max_{\mu} E_{\mu^{(k)}, \sigma^{2(k+1)}} (\log p_\sigma(\mu, \mathbf{I}, \Theta) | \mathbf{I}). \end{aligned}$$

Define the complete-data log-likelihood  $Q(\mu, \sigma^2 | \mu^{(k)}, \sigma^{2(k)}, \mathbf{I})$  as the right hand side of the previous equation which must be maximized alternatively in  $\mu$  and  $\sigma^{2(k)}$ :

$$\begin{aligned} Q(\mu, \sigma^2 | \mu^{(k)}, \sigma^{2(k)}, \mathbf{I}) &= E_{\mu^{(k)}} \left\{ -\frac{1}{2}(\mu, K_\pi \mu) - \sum_{n=1}^N \frac{1}{2}(\theta_n, K \theta_n) \right. \\ &\quad \left. - \frac{1}{2\sigma^2} \sum_{n=1}^N \|J_0 \circ g_\mu^{-1} \circ g_{\theta_n}^{-1} - I_n\|_2^2 - \frac{NS}{2} \log \sigma^2 \right\} + C \quad (19) \\ &= -\frac{1}{2}(\mu, K_\pi \mu) - \frac{1}{2\sigma^2} \sum_{n=1}^N E_{\mu^{(k)}} \left\{ \|J_0 \circ g_\mu^{-1} \circ g_{\theta_n}^{-1} - I_n\|_2^2 | I_n \right\} \\ &\quad - \frac{NS}{2} \log \sigma^2 + \tilde{C} \end{aligned}$$

where  $S$  is the number of grids (or pixels, voxels) and  $C$  and  $\tilde{C}$  are expressions that do not depend on  $\mu$  or  $\sigma^2$ . The maximization M-step at transition step  $k$  yields

$$\sigma^{2(k+1)} = \frac{1}{SN} \sum_{n=1}^N E_{\mu^{(k)}} \left\{ \|J_0 \circ g_{\mu^{(k)}}^{-1} \circ g_{\theta_n}^{-1} - I_n\|_2^2 | I_n \right\} \quad (20)$$

$$\begin{aligned} \mu^{(k+1)} &= \arg\min_{\mu} \left\{ (\mu, K_\pi \mu) \right. \\ &\quad \left. + \frac{1}{\sigma^{2(k+1)}} \sum_{n=1}^N E_{\mu^{(k)}} \left\{ \|J_0 \circ g_\mu^{-1} \circ g_{\theta_n}^{-1} - I_n\|_2^2 | I_n \right\} \right\} \end{aligned} \quad (21)$$

*Remark:* The resulting MAEM algorithm formally coincides with the Maximum a Posterior (MAP) estimation, which maximizes the likelihood with respect to all parameters. We use MAEM here instead of MAP to be reminded that the deformation (with respect to which the mode is computed) is not a parameter, but a hidden variable, and that our procedure should be considered as an approximation of the computation of the maximum likelihood of the observations.

#### Maximization via the Euler equation with Jacobian weight

This is the EM framework for template estimation. Now, the main difficulty of the algorithm is the minimization in Eq. (21). To derive the Euler–Lagrange equation for the minimizer, we use the integral formula for the norm, so we can avoid the interpolation problem associated with discrete sum definition  $\|I - I'\|_2^2 = \sum_{i=1}^S (I(x_i) - I'(x_i))^2$  which corresponds to signal plus additive white noise. This makes the change of variable formula straightforward and links us to the Euler–Lagrange equations on vector fields which have been

previously published. As well, our implementation is a discretization of that continuum equation. Note that one can provide a fully discrete analysis of the problem (relying on a representation of the images using linear interpolations as in Allasonnière et al., 2007, where both images and deformations are linear combinations of the kernels centered at the landmark points).

For the variation, let  $V_\pi$  be the reproducing kernel Hilbert space associated to the prior kernel  $K_\pi$ . Since the energy is conserved along geodesics, we have

$$(\mu, K_\pi \mu) = \|v_0\|_{V_\pi}^2 = \int_0^1 \|v_t\|_{V_\pi}^2 dt \quad (22)$$

where  $v_0 = K_\pi \mu$  is the initial velocity. This connects our optimization in the MAEM algorithm to the original LDDMM Euler–Lagrange equation of Beg et al. (2005).

If  $g_t$  is a time-dependent flow of diffeomorphisms, we let  $g_{s,t}: \Omega \rightarrow \Omega$  denote the composition  $g_{s,t}(y) = g_t \circ (g_s)^{-1}(y)$ , meaning position  $t$  at time of a particle that is at position  $y$  at time  $s$ . Let  $D_{g_{s,t}}$  denote the Jacobian of mapping  $g_{s,t}$ , the matrix composed with the space derivatives. We add a superscript  $v$  to indicate that is  $g_t = g_t^v$  the flow arising from Eq. (1). We state how perturbations of the vector field affect the variation of the mapping in the following lemma.

**Lemma 1.** The variation of mapping  $g_{s,t}^v$  when  $v \in L^2([0,1], V_\pi)$  is perturbed  $h \in L^2([0,1], V_\pi)$  along is given by

$$\partial_h g_{s,t}^v = \lim_{\varepsilon \rightarrow 0} \frac{g_{s,t}^{v+\varepsilon h} - g_{s,t}^v}{\varepsilon} = Dg_{s,t}^v \int_t^s (Dg_{s,u}^v)^{-1} h_u \circ g_{s,u}^v du. \quad (23)$$

Refer to Beg et al., (2005) for proof.

Then the maximization is reduced to what we term the *weighted-LDDMM image matching problem*.

**Proposition 1.** (i) (Formalization to a weighted-LDDMM problem) At stage  $k+1$  define the following ancillary average of the conditional mean:

$$\bar{I}^{(k+1)}(y) = \frac{\sum_{n=1}^N E_{\mu^{(k)}} \{I_n \circ g_{\theta_n}(y) | Dg_{\theta_n}(y) | |I_n\}}{\sum_{n=1}^N E_{\mu^{(k)}} \{|Dg_{\theta_n}(y) | |I_n\}}. \quad (24)$$

Then the M-step of the generalized EM algorithm reduces to the weighted-LDDMM algorithm:

$$\begin{aligned} \mu^{(k+1)} &= \underset{\mu}{\operatorname{argmin}} \left\{ (\mu, K_\pi \mu) \right. \\ &\quad \left. + \sum_{n=1}^N \frac{1}{\sigma^{2(k+1)}} E_{\mu^{(k)}} \left\{ \|J_0 \circ g_\mu^{-1} \circ g_{\theta_n}^{-1} - I_n\|_2^2 | I_n \right\} \right\} \\ &= \underset{\mu}{\operatorname{argmin}} \left\{ (\mu, K_\pi \mu) \right. \\ &\quad \left. + \frac{1}{\sigma^{2(k+1)}} \int_\Omega (J_0 \circ g_\mu^{-1}(y) - \bar{I}^{(k+1)}(y))^2 \alpha^{(k+1)}(y) dy \right\} \end{aligned} \quad (25)$$

with the Jacobian weight

$$\alpha^{(k+1)}(y) = \sum_{n=1}^N E_{\mu^{(k)}} \{|Dg_{\theta_n}(y) | |I_n\}. \quad (26)$$

(ii) (The weighted Euler–Lagrange equation) Given a continuously differentiable template image  $J_0$ , a target image  $\bar{I}$  and

Jacobian weight  $\alpha$ , the optimal velocity field  $\hat{v}_2 [(0,1), V_\pi]$  with  $\hat{v}_0 = K_\pi \hat{\mu}$  for inexact matching of  $J_0$  and  $\bar{I}$  defined as

$$\underset{v: \hat{g}_t = v_t(g_t)}{\operatorname{argmin}} \left\{ \int_0^1 \|v_t\|_{V_\pi}^2 dt + \frac{1}{\sigma^2} \int_\Omega (\bar{I}(y) - J_0 \circ g_\mu^{-1}(y))^2 \alpha(y) dy \right\} \quad (27)$$

satisfies the Euler–Lagrange equation

$$2\hat{v}_t - K_\pi \left( \frac{2}{\sigma^2} |Dg_{t,1}| \nabla H_t^0 (H_t^0 - H_t^1) \alpha \circ g_{t,1} \right) = 0 \quad (28)$$

where  $H_t^0 = J_0 \circ g_{t,0}$ ,  $H_t^1 = \bar{I} \circ g_{t,1}$ .

The crucial idea here is that we are linked back to the basic LDDMM image matching problem of Beg ( $\alpha=0$ ), with the Jacobian playing a role.

**Proof.** (i) Let  $y = g_{\theta_n}^{-1}(x)$ , we have

$$\begin{aligned} &\|J_0 \circ g_\mu^{-1} \circ g_{\theta_n}^{-1} - I_n\|_2^2 \\ &\approx \int_\Omega (J_0 \circ g_\mu^{-1} \circ g_{\theta_n}^{-1}(x) - I_n(x))^2 dx \\ &= \int_\Omega (J_0 \circ g_\mu^{-1}(y) - I_n \circ g_{\theta_n}(y))^2 |Dg_{\theta_n}(y)| dy \end{aligned} \quad (29)$$

With  $\bar{I}^{(k+1)}$  as defined in Eq. (24), we have, for all  $y \in \Omega$

$$\begin{aligned} &\sum_{n=1}^N E_{\mu^{(k)}} \left\{ (J_0 \circ g_\mu^{-1}(y) - I_n \circ g_{\theta_n}(y))^2 |Dg_{\theta_n}(y)| |I_n\} \right\} \\ &= \sum_{n=1}^N E_{\mu^{(k)}} \left\{ (J_0 \circ g_\mu^{-1}(y) \right. \\ &\quad \left. - \bar{I}^{(k+1)}(y) + \bar{I}^{(k+1)}(y) - I_n \circ g_{\theta_n}(y))^2 |Dg_{\theta_n}(y)| |I_n\} \right\} \\ &= \sum_{n=1}^N E_{\mu^{(k)}} \left\{ (J_0 \circ g_\mu^{-1}(y) - \bar{I}^{(k+1)}(y))^2 |Dg_{\theta_n}(y)| \right. \\ &\quad \left. + (\bar{I}^{(k+1)}(y) - I_n \circ g_{\theta_n}(y))^2 |Dg_{\theta_n}(y)| \right. \\ &\quad \left. + 2(J_0 \circ g_\mu^{-1}(y) - \bar{I}^{(k+1)}(y)) (\bar{I}^{(k+1)}(y) \right. \\ &\quad \left. - I_n \circ g_{\theta_n}(y)) |Dg_{\theta_n}(y)| |I_n\} \right\} \\ &\stackrel{(a)}{=} \sum_{n=1}^N E_{\mu^{(k)}} \left\{ (J_0 \circ g_\mu^{-1}(y) - \bar{I}^{(k+1)}(y))^2 |Dg_{\theta_n}(y)| \right. \\ &\quad \left. + (\bar{I}^{(k+1)}(y) - I_n \circ g_{\theta_n}(y))^2 |Dg_{\theta_n}(y)| |I_n\} \right\} \\ &= (J_0 \circ g_\mu^{-1}(y) - \bar{I}^{(k+1)}(y))^2 \sum_{n=1}^N E_{\mu^{(k)}} \{|Dg_{\theta_n}(y) | |I_n\} \\ &\quad + \sum_{n=1}^N E_{\mu^{(k)}} \left\{ (\bar{I}^{(k+1)}(y) - I_n \circ g_{\theta_n}(y))^2 |Dg_{\theta_n}(y) | |I_n\} \right\}. \end{aligned} \quad (30)$$

In (a), the fact that the cross item

$$\begin{aligned} &\sum_{n=1}^N E_{\mu^{(k)}} \left\{ (J_0 \circ g_\mu^{-1}(y) - \bar{I}^{(k+1)}(y)) \right. \\ &\quad \left. \times (\bar{I}^{(k+1)}(y) - I_n \circ g_{\theta_n}(y)) |Dg_{\theta_n}(y) | |I_n\} \right\} \\ &= (J_0 \circ g_\mu^{-1}(y) - \bar{I}^{(k+1)}(y)) \\ &\quad \sum_{n=1}^N E_{\mu^{(k)}} \left\{ (\bar{I}^{(k+1)}(y) - I_n \circ g_{\theta_n}(y)) |Dg_{\theta_n}(y) | |I_n\} \right\} = 0 \end{aligned}$$

comes straightforward from the definition of  $\bar{I}^{(k+1)}(y)$ .

Since the second term of Eq (30) does not depend on  $\mu$ , substituting Eq. (30) into Eq. (21), we see that  $\mu^{(k+1)}$  must minimize

$$(\mu, K_\pi \mu) + \frac{1}{\sigma^{2(k+1)}} \int_{\Omega} (\bar{I}^{(k+1)}(y) - J_0 \circ g_{\mu}^{-1}(y))^2 \alpha^{(k+1)}(y) dy \quad (31)$$

with the Jacobian weight

$$\alpha^{(k+1)}(y) = \sum_{n=1}^N E_{\mu^{(k)}} \{ |Dg_{\theta_n}(y)| |I_n\} \quad (32)$$

With the optimal  $\mu^{(k+1)}$ , we can compute  $g_{\mu^{(k+1)}}$  by geodesic shooting (Eq. (4)), and further obtain the newly estimated template  $J^{(k+1)} = J_0 \circ g_{\mu^{(k+1)}}$ . This gives the first part of the proof.

(ii) The proof of the second half follows the derivation in Beg et al. (2005). Suppose the velocity is  $v \in L^2([0,1], V_\pi)$  perturbed along the direction  $h \in L^2([0,1], V_\pi)$  by an  $\epsilon$  amount. The Gâteaux  $\partial_h E(v)$  variation of the energy function is expressed in term of the Fréchet derivative  $\nabla_v E$

$$\partial_h E(v) = \lim_{\epsilon \rightarrow 0} \frac{E(v + \epsilon h) - E(v)}{\epsilon} = \int_0^1 \langle \nabla_v E_t, h_t \rangle_{V_\pi} dt. \quad (33)$$

The variation of  $E_1(v) = \int_0^1 \|v_t\|_{V_\pi}^2 dt$  is given by

$$\partial_h E_1(v) = 2 \int_0^1 \langle v_t, h_t \rangle_{V_\pi} dt. \quad (34)$$

The second part of the energy is

$$\begin{aligned} E_2(v) &= \frac{1}{\sigma^2} \int_{\Omega} (J_0 \circ g_{\mu}^{-1}(y) - \bar{I}(y))^2 \alpha(y) dy \\ &= \frac{1}{\sigma^2} \langle (J_0 \circ g_{1,0} - \bar{I}) \alpha, J_0 \circ g_{1,0} - \bar{I} \rangle_{L^2}. \end{aligned} \quad (35)$$

The variation of  $E_2(v)$  is

$$\begin{aligned} \partial_h E_2(v) &= \frac{2}{\sigma^2} \langle (J_0 \circ g_{1,0} - \bar{I}) \alpha, DJ_0 \circ g_{1,0} \partial_h g_{1,0} \rangle_{L^2} \\ &\stackrel{(a)}{=} \frac{2}{\sigma^2} \langle (J_0 \circ g_{1,0} - \bar{I}) \alpha, DJ_0 \circ g_{1,0} \left( -Dg_{1,0} \int_0^1 (Dg_{1,t})^{-1} h_t \circ g_{1,t} dt \right) \rangle_{L^2} \\ &\stackrel{(b)}{=} -\frac{2}{\sigma^2} \int_0^1 \langle (J_0 \circ g_{1,0} - \bar{I}) \alpha, D(J_0 \circ g_{1,0})(Dg_{1,t})^{-1} h_t \circ g_{1,t} \rangle_{L^2} dt \end{aligned} \quad (36)$$

with (a) derived straightly from Lemma 1 and (b) from the formulation  $D(J_0 \circ g_{1,0}) = DJ_0 \circ g_{1,0} Dg_{1,0}$ . Changing the variable with  $z = g_{1,t}(y)$  i.e.  $g_{t,1}(z) = y$ , one can obtain  $|dg_{t,1}| dz = dy$ . The chain rule gives  $g_{1,0} \circ g_{t,1} = g_{t,0}$ . In addition,  $D(I \circ g) = (\nabla(I \circ g))^T$ . With these substitutions, we get

$$\begin{aligned} \partial_h E_2(v) &= -\frac{2}{\sigma^2} \int_0^1 \langle |Dg_{t,1}| (J_0 \circ g_{t,0} - \bar{I} \circ g_{t,1}) \alpha \circ g_{t,1}, D(J_0 \circ g_{t,0}) h_t \rangle_{L^2} dt \\ &= -\frac{2}{\sigma^2} \int_0^1 \langle |Dg_{t,1}| \nabla (J_0 \circ g_{t,0}) (J_0 \circ g_{t,0} - \bar{I} \circ g_{t,1}) \alpha \circ g_{t,1}, h_t \rangle_{L^2} dt \\ &= -\int_0^1 \langle K_\pi \left( \frac{2}{\sigma^2} |Dg_{t,1}| \nabla (J_0 \circ g_{t,0}) (J_0 \circ g_{t,0} - \bar{I} \circ g_{t,1}) \alpha \circ g_{t,1} \right), h_t \rangle_{V_\pi} dt \\ &= -\int_0^1 \langle K_\pi \left( \frac{2}{\sigma^2} |Dg_{t,1}| \nabla H_t^0 (H_t^0 - H_t^1) \alpha \circ g_{t,1} \right), h_t \rangle_{V_\pi} dt. \end{aligned}$$

Combining the two parts of energy functional, the gradient is thus

$$(\nabla_v E_t)_{V_\pi} = 2v_t - K_\pi \left( \frac{2}{\sigma^2} |Dg_{t,1}| \nabla H_t^0 (H_t^0 - H_t^1) \alpha \circ g_{t,1} \right) \quad (37)$$

where the subscript  $V_\pi$  in  $(\nabla_v E_t)_{V_\pi}$  is to clarify that the gradient is in the space  $L^2([0,1], V_\pi)$ . The optimizing velocity fields satisfy the Euler–Lagrange equation

$$\nabla_h E(\hat{v}) = \int_0^1 \langle 2\hat{v}_t - K_\pi \left( \frac{2}{\sigma^2} |Dg_{t,1}| \nabla H_t^0 (H_t^0 - H_t^1) \alpha \circ g_{t,1} \right), h_t \rangle_{V_\pi} dt = 0. \quad (38)$$

Since  $h$  arbitrary in  $L^2([0,1], V_p)$  we get Eq. (28).  $\square$

Eq. (37) provides the gradient flow that minimizes Eq. (27). Recall that this problem must be solved to obtain the next deformation of the hypertemplate: given the solution  $\hat{v}$ , compute the initial momentum  $\hat{\mu} = (K_\pi)^{-1} \hat{v}_0$ , the optimal diffeomorphism  $g_{\hat{\mu}}$  and the new template  $J = J_0 \circ g_{\hat{\mu}}^{-1}$ .

Since the Euler–Lagrange equation for the weighted LDDMM only differs from the original equation by the  $\alpha \circ g_{t,1}$  factor, its implementation is a minor modification to the basic one, for which we refer to Beg et al. (2005) for details.

#### Computing the conditional mean via the mode

Another difficulty is to compute the conditional expectations, which cannot be done analytically, given the highly nonlinear relation between  $\theta_n$  and  $I_n \circ g_{\theta_n}$ . The crudest approximation of the conditional distribution is to replace it by a Dirac measure at its mode, and this is the one we will select for the time being. Already having  $J^{(k)}$ , estimation of the template in  $k$ -th iteration, denote  $\theta_n^{(k)}$  to be the minimizer of

$$(\theta_n, K\theta_n) + \frac{1}{\sigma^{2(k)}} \|J^{(k)} \circ g_{\theta_n}^{-1} - I_n\|_2^2. \quad (39)$$

The computation of  $\bar{I}^{(k+1)}$  now becomes:

$$\bar{I}^{(k+1)}(y) = \frac{\sum_{n=1}^N I_n \circ g_{\theta_n^{(k)}}(y) |Dg_{\theta_n^{(k)}}(y)|}{\sum_{n=1}^N |Dg_{\theta_n^{(k)}}(y)|}. \quad (40)$$

This also provides an approximation of the Jacobian weight

$$\alpha^{(k+1)}(y) = \sum_{n=1}^N |Dg_{\theta_n^{(k)}}(y)|. \quad (41)$$

Concerning the implementation, the computation of  $\theta_n^{(k)}$  is done using the LDDMM algorithm between the template  $J^{(k)}$  and the target  $I_n$ . Note that  $\theta_n^{(k)}$  is not needed for Eq. (40), but only the deformed target  $I_n \circ g_{\theta_n^{(k)}}$  and the determinant of the Jacobian  $Dg_{\theta_n^{(k)}}$ .

#### Template estimation algorithm

Now, the template estimation algorithm can be summarized as the following:

**Algorithm 1. (Template estimation).** Having the hypertemplate  $J_0$  and  $N$  observations  $I_1, I_2, \dots, I_N$ , we wish to estimate the template

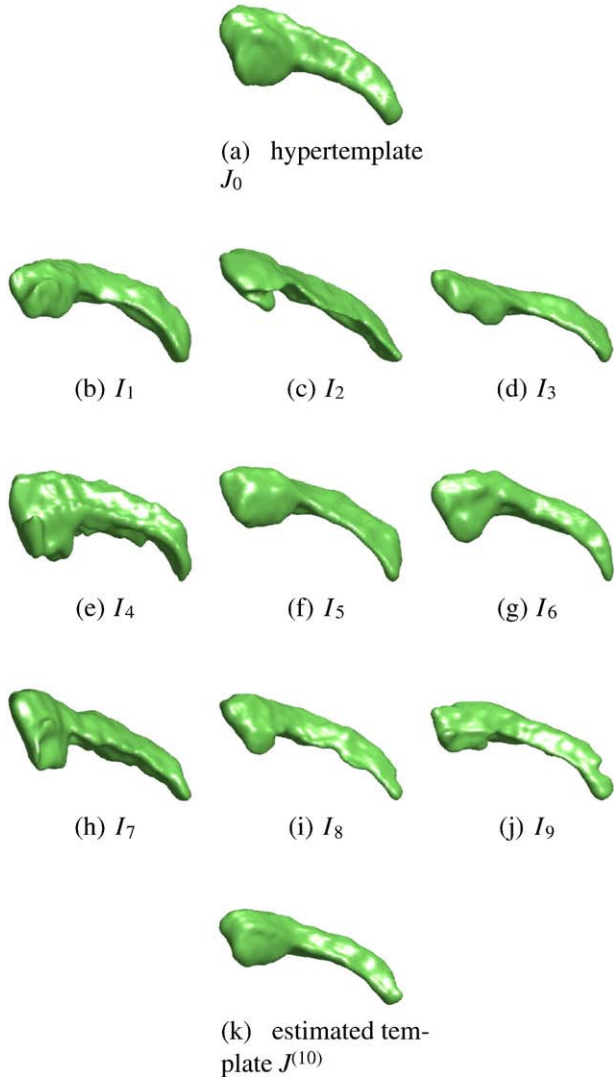


Fig. 2. Estimating the template from 3D hippocampus data. Panel (a) is the hypertemplate. Panels (b)–(j) are observations  $I_n$ ,  $n=1,2,\dots,9$ . Panel (k) is the estimated template after 10 iterations. Data courtesy of Biomedical Informatics Research Network.

$J$  and noise variance  $\sigma^2$ . Let  $J^{(k)}$  denote the estimated template after  $k$  iterations with initial guess  $J^{(0)}=J_0$ . Then, the  $(k+1)$ th step is

- (i) Map the current estimated template  $J^{(k)}$  to  $I_n$ ,  $n=1,2,\dots,N$  using basic LDDMM, and obtain the deformed targets  $I_n \circ g_{\theta_n^{(k)}}$  and Jacobian determinants of the deformations  $|Dg_{\theta_n^{(k)}}|$ .
- (ii) Compute the mean image  $\bar{I}^{(k+1)}$  and the Jacobian weight  $\alpha^{(k+1)}$  defined by

$$\bar{I}^{(k+1)}(y) = \frac{\sum_{n=1}^N I_n \circ g_{\theta_n^{(k)}}(y) |Dg_{\theta_n^{(k)}}(y)|}{\sum_{n=1}^N |Dg_{\theta_n^{(k)}}(y)|}, \quad (42)$$

and

$$\alpha^{(k+1)}(y) = \sum_{n=1}^N |Dg_{\theta_n^{(k)}}(y)| \quad (43)$$

- where  $g_{\theta_n^{(k)}}$  is the optimal diffeomorphic mapping from  $J^{(k)}$  to  $I_n$  and  $|Dg_{\theta_n^{(k)}}(y)|$  is the determinant of its Jacobian matrix.
- (iii) update the noise variance  $\sigma^2$

$$\sigma^{2(k+1)} = \frac{1}{SN} \sum_{n=1}^N \|J_0 \circ g_{\mu^{(k)}}^{-1} \circ g_{\theta_n^{(k)}}^{-1} - I_n\|_2^2 \quad (44)$$

- (iv) find  $\mu^{(k+1)}$  to minimize

$$(\mu, K_\pi \mu) + \frac{1}{\sigma^{2(k+1)}} \left\| \bar{I}^{(k+1)} - J_0 \circ g_{\mu^{(k+1)}}^{-1} \right\|_{\sqrt{\alpha^{(k+1)}}}^2 \quad (45)$$

using the weighted Euler–Lagrange equation described previously.

- (v) the newly estimated template is  $J^{(k+1)} = J_0 \circ g_{\mu^{(k+1)}}^{-1}$ .
- (vi) Stop if  $J^{(k)}$  is stable or the number of iterations is larger than a specific number. Else reiterate (i)–(iv).

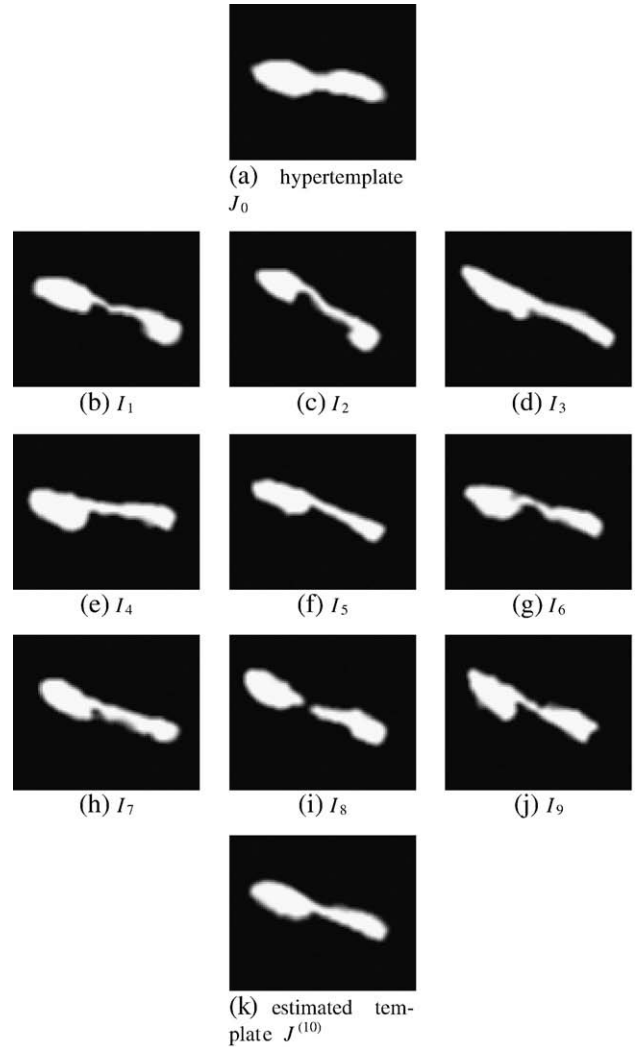


Fig. 3. Section view of 3D hippocampus data. Panel (a) is the hypertemplate. Panels (b)–(j) are observations  $I_n$ ,  $n=1,2,\dots,9$ . Panel (k) is the estimated template after 10 iterations. Data courtesy of Biomedical Informatics Research Network.

**Results and discussions**

Here we present numerical results of template estimation for 3D hippocampus data and 3D cardiac data. All data are binarized segmented images with grayscale 0–255 (the images are not strictly binary because of smoothing and interpolation). For these experiments, we have used  $K_{\pi} = \lambda K$  with a suitable value of  $\lambda$ .

Shown in Fig. 2 is an example of template estimation of 3D hippocampus data. Panel (a) is the hypertemplate and panels (b)–(i) are observations. Panel (j) is the estimated template with  $\lambda = 0.01$ .

We present sections of the 3D data in Fig. 3 to show more clearly that the estimated template adapts to the shapes of observations.

We define the deformation metric,  $\rho_f(J, I_n)$  to be square root of the deformation energy  $\int_0^1 \|\dot{v}\|_v^2$ , for the optimal velocity provided by the LDDMM algorithm. To show that the estimated template is a considerable improvement upon the hypertemplate, we list the metrics  $\rho_f(J_0, I_n)$  and  $\rho_f(J^{(10)}, I_n)$  in Table 1, which are computed with the same parameters. This shows a significant metric reduction from the original hypertemplate to the template estimated after 10 steps.

To assess the convergence of the results, we investigate the differences between the estimated templates in successive iterations

$$\|J^{(k)} - J^{(k+1)}\|_2^2 = \frac{1}{S} \sum_{s=1}^S (J^{(k)}(x_s) - J^{(k+1)}(x_s))^2 \quad (46)$$

where  $S$  is the number of voxels and  $k=0,1,\dots,9$ . The results are shown in Table 2. We see that the differences between  $J^{(k+1)}$  and  $J^{(k)}$  decrease rapidly to a small value (approximately 0) in the first 10 iterations. This indicates that the results converge to a stable shape. In Table 2, we also show the estimated noise levels, which converge too.

Finally we present the result for 3D heart template estimation. Panel (a) of Fig. 4 is the hypertemplate. Panels (b)–(g) are observations  $I_n$ ,  $n=1,2,\dots,6$ . Panel (h) is the estimated template with  $\lambda=0.0001$  at the 10th iteration. Fig. 5 is the section view.

In our model, the hypertemplate is considered as an “ideal” continuous image with fine structure, which can be provided by an atlas obtained from other studies, although we here simply choose a representative image in the population. Actually, as Fig. 6 shows, different hypertemplates yield close results, although they have minor difference.

$\lambda$  controls how strongly the estimated template depends on the hypertemplate. This has been fixed by hands, but our experiments

Table 1  
The metric between observations and  $J_0, J^{(10)}$

$\rho_f(\cdot, \cdot)$	$J_0$	$J^{(10)}$
$I_1$	5.2995	4.2773
$I_2$	7.6836	4.1446
$I_3$	4.7706	3.7492
$I_4$	4.9767	2.4993
$I_5$	4.3480	3.3836
$I_6$	4.2751	2.6810
$I_7$	5.3083	3.2349
$I_8$	5.1540	3.4601
$I_9$	5.8151	3.6120

Table 2

Differences between the estimated templates in successive iterations and estimated noise variances

Iteration $k$	$\ J^{(k+1)} - J^{(k)}\ _2^2$	$\sigma^{(k)}$
0	29.7629	1.0000
1	228.3461	8.5799
2	232.6406	22.1642
3	13.7406	27.1063
4	0.3902	24.5795
5	0.1638	23.8118
6	0.0776	23.6478
7	0.0036	23.5560
8	0.0014	23.5726
9	0.0003	23.5772

show a large range of variation without noticeable difference in the final result. By taking small values of  $\lambda$ , the prior reduces to an almost uniform distribution over the orbit of the hypertemplate. We indeed took values between 0.0001 and 1 and obtained stable results.

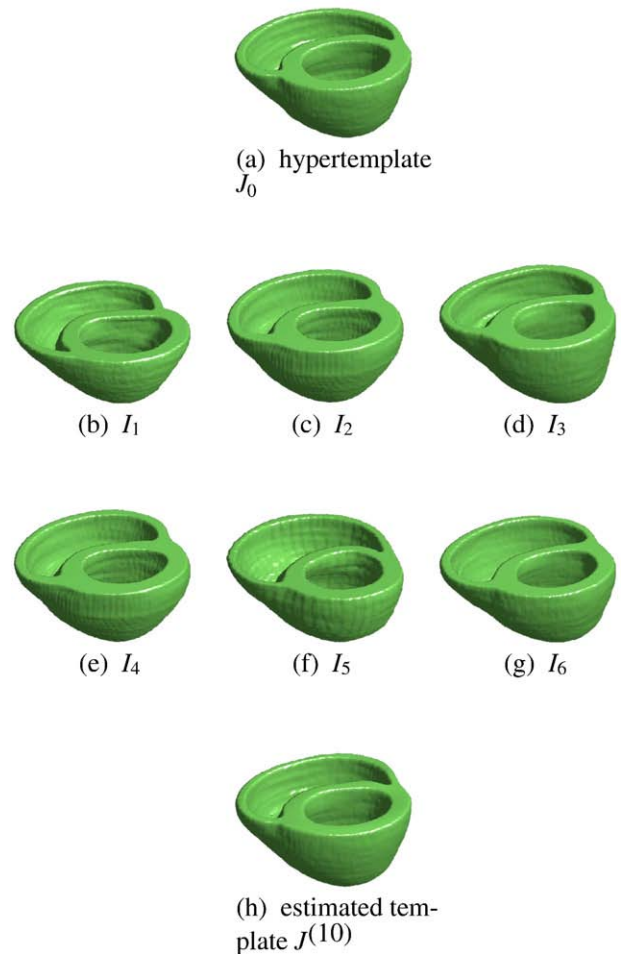


Fig. 4. Estimating the template from 3D heart data. Panel (a) is the hypertemplate. Panels (b)–(g) are observations  $I_n$ ,  $n=1,2,\dots,6$ . Panel (h) is the estimated template after 10 iterations. Data courtesy of Dr. Patrick Helm, previously of Dept. of Biomedical Engineering, Johns Hopkins University.

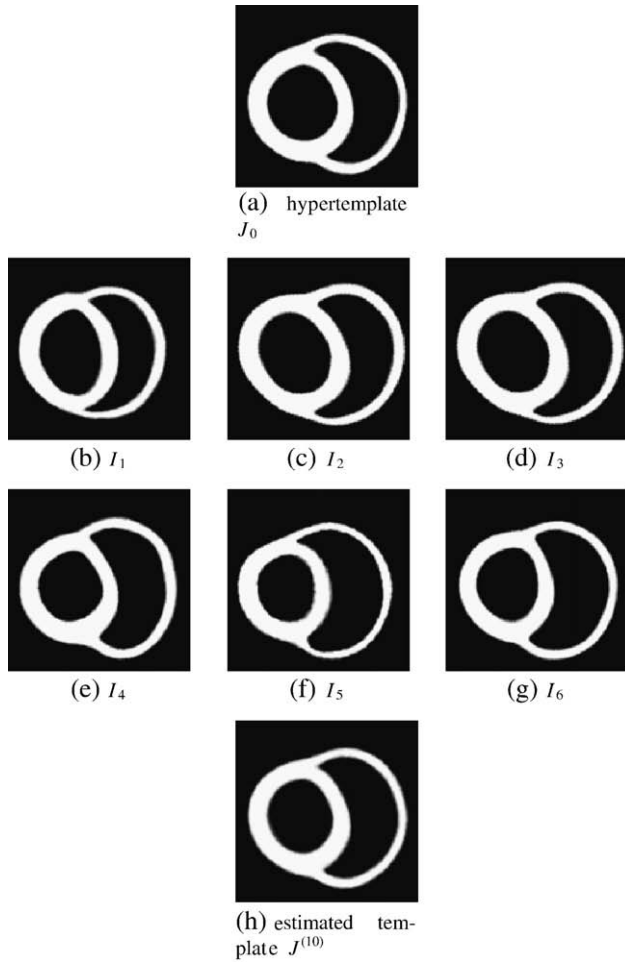


Fig. 5. Section view of 3D heart data. Panel (a) is the hypertemplate. Panels (b)–(g) are observations  $I_n$ ,  $n=1,2,\dots,6$ . Panel (h) is the estimated template after 10 iterations. Data courtesy of Dr. Patrick Helm, previously of Dept. of Biomedical Engineering, Johns Hopkins University.

*Remark:* In the above model, we assume the initial momenta  $\mu$  follows a prior distribution  $p_\pi(\mu)$  and estimate the template given observations. We call this the “full model”. We can simplify this model by neglecting the prior and just estimate the maximum likelihood of the template  $J$ . The MAEM algorithm on this setting will lead us to iterate the procedure of

$$\bar{I}^{(k+1)}(y) = \frac{\sum_{n=1}^N I_n \circ g_{\theta_n^{(k)}}(y) |d_y g_{\theta_n^{(k)}}|}{\sum_{n=1}^N |d_y g_{\theta_n^{(k)}}|}. \quad (47)$$

The resulting algorithm is similar to Joshi et al. (2004). The difference is that the model in Joshi et al. (2004) warps the observations to match the template and places the white noise between the deformed target and the template, which does not induce a Jacobian in the averaging process. However, for a generative model, the logical progression is from template to target. From this point of view, Joshi et al. (2004) implies an observation noise that is proportional to the inverse Jacobian of the deformation, which is hard to justify.

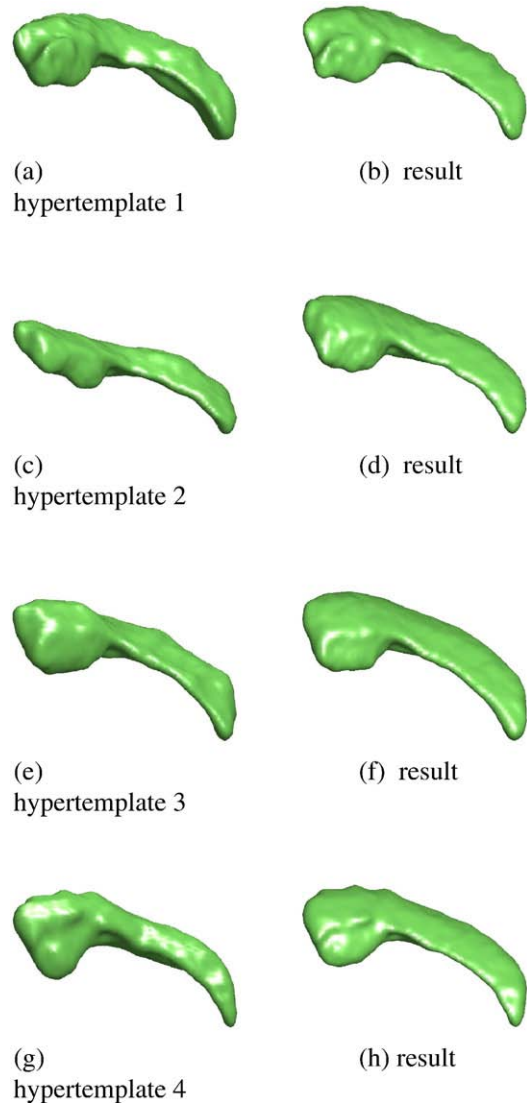


Fig. 6. For the same observed population, we choose different images as hypertemplate,  $\lambda=0.01$ . The results only have minor differences.

Figs. 7 and 8 compare the full model and the simplified model. The simplified model performed relatively poorly compared with the full model that used the prior. This discrepancy comes from the fact that we simultaneously estimate the template and the noise variance. This estimated value of the noise variance allows for

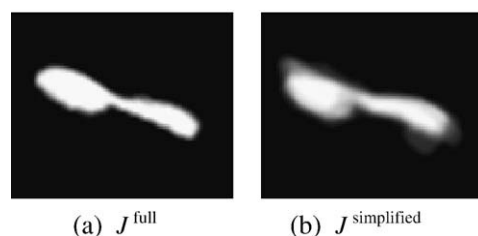


Fig. 7. The template estimation results of full model and simplified model for hippocampus data. Data courtesy of Biomedical Informatics Research Network.

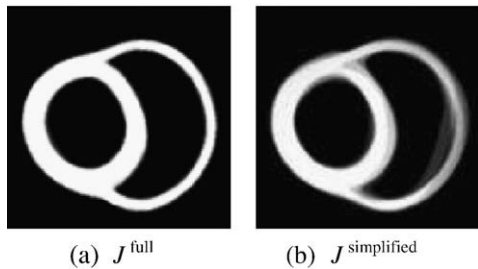


Fig. 8. The template estimation results of full model and simplified model for cardiac data. Data courtesy of Dr. Patrick Helm, previously of Dept. of Biomedical Engineering, Johns Hopkins University.

some difference between the deformed template and the targets, yielding fuzzy boundaries in the simplified model. If we set the noise variance to a small number, we may obtain sharper boundaries using the simplified model, but this would estimate a template essentially as a metric average (a Fréchet mean) of the targets and would not be consistent with our generative model.

## Conclusion

In conclusion, we have presented in this paper a Bayesian model for template estimation in CA. By the momentum conservation law, the space of initial momenta is a linear space where statistical analysis can be applied. It is assumed that observed images  $I_1, I_2, \dots, I_N$  are generated by shooting the template  $J_0$  through Gaussian-distributed random initial momenta  $\theta_1, \theta_2, \dots, \theta_N$ . The template is  $J$  modeled as a deformation from a given hypertemplate  $J_0$  with initial momentum  $\mu$ , which has a Gaussian prior. This allows us to apply a generalized EM algorithm MAEM to computing the Bayesian estimation of the initial momentum  $\mu$ , where the conditional expectation of the EM is approached by a Dirac measure, so that one can take the advantage of the LDDMM algorithm. The MAEM procedure finally leads to an image mapping problem from  $J_0$  to  $\bar{J}$  with Jacobian weight  $\alpha$  in the energy term, which is solved by the weighted Euler–Lagrange equation. In particular, we apply this method to template estimation for hippocampus and cardiac images. We show that the estimated template is “closer” to observations compared to the hypertemplate, and the differences between the estimated templates in successive iterations decrease to almost 0, which indicates the convergence of the algorithm. We also show that the results are stable with different hypertemplates.

## Acknowledgments

This work is partially supported by NSF DMS-0456253, NIH R01-EB000975, NIH P41-RR15241, NIH R01-MH064838, NIH 5 U24 RR021382-04, NIH 2 P01 AG003991-24, NIH 1 P02 AG02627601, and NIH 5 P50 MH071616-04.

## References

- Allasonnière, S., Amit, Y., Trouvé, A., 2007. Toward a coherent statistical framework for dense deformable template estimation. *J. R. Stat. Soc. Ser. B* 69 (1), 3–29.
- Arnold, V.I., 1966. Sur un principe variationnel pour les écoulements stationnaires des liquides parfaits et ses applications aux problèmes de stabilité non linéaires. *J. de Mécanique* 5, 29–43.
- Arnold, V.I., 1978. *Mathematical Methods of Classical Mechanics*, Second Edition. Springer, 1989.
- Avants, B.B., Gee, J.C., 2004a. Geodesic estimation for large deformation anatomical shape averaging and interpolation. *NeuroImage* 23 (supplement 1), S139–S150.
- Avants, B.B., Gee, J.C., 2004b. Symmetric geodesic shape averaging and shape interpolation. *ECCV Workshops CVAMIA and MMBIA*, pp. 99–110.
- Beg, M.F., Khan, A., 2006. Computation of average atlas using lddmm and geodesic shooting. *IEEE Int. Symp. Biomed. Imag.*
- Beg, M.F., Miller, M.I., Trouvé, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vis.* 61 (2), 139–157.
- Dupuis, P., Grenander, U., Miller, M.I., 1998. Variational problems on flows of diffeomorphisms for image matching. *Q. Appl. Math.* 56, 587–600.
- Fletcher, P.T., Joshi, S., Lu, C., Pizer, S.M., 2004. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imag.* 23 (8), 995–1005.
- Glasbey, C.A., Mardia, K.V., 2001. A penalized likelihood approach to image warping. *J. R. Stat. Soc.* 63 (3), 465–492.
- Grenander, U., 1994. *General Pattern Theory*. Oxford Univ. Press, 27.
- Guimond, A., Meunier, J., Thirion, J.P., 2000. Average brain models: a convergence study. *Comput. Vis. Image Underst.* 77 (2), 192–210.
- Helm, P.A., Younes, L., Beg, M.F., Ennis, D.B., Leclercq, C., Faris, O.P., McVeigh, E., Kass, D., Miller, M.I., Winslow, R.L., 2006. Evidence of structural remodeling in the dyssynchronous failing heart. *Circ. Res.* 98, 125–132.
- Holm, D.D., Marsden, J.E., Ratiu, T.S., 1998. The Euler–Poincaré equations and semidirect products with applications to continuum theories. *Adv. Math.* 137, 1–81.
- Joshi, S., Davis, B., Jomier, M., Gerig, G., 2004. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage; Suppl. Issue Math. Brain Imaging* 23 (Supplement 1), S151–S160.
- Le, H., Kume, A., 2000. The Fréchet mean shape and the shape of means. *Adv. Appl. Probab.* 32, 101–113.
- Marsden, J.E., Ratiu, T.S., 1999. *Introduction to Mechanics and Symmetry*. Springer.
- Miller, M.I., Trouvé, A., Younes, L., 2003. On the metrics, Euler equations and normal geodesic image motions of computational anatomy. *Proceedings of the 2003 International Conference on Image Processing, IEEE*, 2, pp. 635–638.
- Miller, M.I., Trouvé, A., Younes, L., 2006. Geodesic shooting for computational anatomy. *J. Math. Imaging Vis.* 24 (2), 209–222 January.
- Mio, W., Srivastava, A., Joshi, S., 2007. On shape of plane elastic curves. *Int. J. Comput. Vis.* 73 (3), 307–324.
- Pennec, X., 2006. Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *J. Math. Imaging Vis.* 25 (1), 127–154.
- Trouvé, A. An infinite dimensional group approach for physics based model. Technical report (electronically available at <http://www.cis.jhu.edu>). unpublished.28.
- Vailliant, M., Miller, M.I., Younes, L., Trouvé, A., 2004. Statistics on diffeomorphisms via tangent space representations. *NeuroImage* 23, S161–S169.