

MAP Estimation of Statistical Deformable Template Via Nonlinear Mixed Effect Models : Deterministic and Stochastic Approaches

Stéphanie Allassonnière¹, Estelle Kuhn², and Alain Trounev³

¹ Center for Imaging Science, Johns Hopkins University, USA,

² LAGA, University Paris 13, France,

³ CMLA, ENS Cachan, France,

stephanie@jhu.edu, kuhn@math.univ-paris13.fr, alain.trouve@cmla.ens-cachan.fr.

Abstract. In [1], a new coherent statistical framework for estimating statistical deformable template relevant to computational anatomy (CA) as been proposed. It addresses the problem of population average and estimation of the underlying geometrical variability as a MAP computation problem for which deterministic and stochastic approximation scheme have been proposed. We illustrate some of the numerical issues with handwritten digit and 2D medical images and apply the estimated models to classification through maximum likelihood.

1 Introduction

For the last decade, we are witnessing impressive achievements and the emergence of elaborated registration theories [2–4] but the definition of a proper statistical framework for designing and inferring stochastic deformable templates in a principle way is much less mature. Despite a seminal contribution [5] and the fact that deformable templates can be cast into the general Grenander’s Pattern Theory [6], the down-to-earth and fundamental problem of computing *population average* in presence of *unobserved* warping variables has not received so much attention from a more mathematical statistics perspective. More statistically oriented methods are slowly emerging [7–9] based on penalised likelihood or equivalently MDL approaches. Another line of research is to deal with the problem of population average as an estimation issue of proper stochastic (ie generative) models for which *consistency issues* should be addressed. In this direction, nonlinear mixed effects models (NLMM) are common tools in biostatistics and pharmacokinetic [10] to deals with both modelisation and inference of common *population factors* (fixed effects) and *distributions* of *unobserved* individuals factors (random effects). An active realm of research has emerged in the 90’s for designing efficient and consistent estimation algorithms. The importation of such ideas even in the limited context of population average of grey level images in CA is extremely appealing and challenging –both theoretically and practically– because of the very large (virtually infinite) dimensionality of the related factors (common template and individual warpings). This new avenues have started to be explored and theoretically consistent procedures based on recent advances on stochastic approximation algorithms have been proposed in a series of papers [1, 11, 12]. Since these papers are mainly mathematically focussed papers, we would

like in the present paper to address some of the numerical issues of the various “EM-like” algorithms proposed to numerically approximate the Maximum A Posteriori estimator. Some relevant results on the USPS database and 2D medical images are presented, showing the strength of such methods.

The paper is organised as follow. Sections 2, 3 and 4 respectively recall the mixture model and how the estimation is completed and the particular case of the one component. The last section, Section 5, is devoted to the experiments.

2 The observation model: BME-Templates

Consider a population of n gray level images $(y_i(s))_{s \in \Lambda}$ defined on a discrete grid of pixels Λ and assume that each observation y derives from a noisy sampling at the pixels locations $(x_s)_{s \in \Lambda}$ of an *unobserved* deformation field $z : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ of a *common* continuously defined template $I_0 : \mathbb{R}^2 \rightarrow \mathbb{R}^2$. This is what we call the Bayesian Mixed Effect Templates (BME-Templates). To keep things simple, we work within the small deformation framework [5] and assume that $y(s) = I_0(x_s - z(x_s)) + \sigma \epsilon(s) = z I_0(s) + \sigma \epsilon(s)$, where ϵ is a Gaussian normalised white noise and σ^2 is the common noise variance. The template I_0 and the deformation z are restricted to belong to subspaces of reproducing kernel Hilbert spaces V_p (resp V_g) with kernel K_p (resp K_g). Given $(p_k)_{1 \leq k \leq k_p}$ a fixed set of landmarks covering the image domain, the template function I_0 is parametrised by coefficients $\alpha \in \mathbb{R}^{k_p}$ through: $I_\alpha = \mathbf{K}_p \alpha$, where $(\mathbf{K}_p \alpha)(x) = \sum_{k=1}^{k_p} K_p(x, p_k) \alpha(k)$. Similarly we write $z_\beta = \mathbf{K}_g \beta$ with another set of landmarks $(g_k)_{1 \leq k \leq k_g}$ and a vector $\beta \in \mathbb{R}^{2k_g}$ of coefficients. In order to detect a global geometrical behaviour, we consider the parameters β of the deformation field as an unobserved variable which is supposed to be Gaussian centred with covariance matrix Γ_g .

We present a general model based on NLMM defining a Bayesian mixture of m deformable template models (hereafter called components). The model parameters of each component $t \in \{1, \dots, m\}$ are denoted by $\theta_t = (\alpha_t, \sigma_t^2, \Gamma_g^t)$. We assume that θ belongs to the open parameter space $\Theta \doteq \{ \theta = (\alpha_t, \sigma_t^2, \Gamma_g^t)_{1 \leq t \leq m} \mid \forall t \in \{1, \dots, m\}, \alpha_t \in \mathbb{R}^{k_p}, \sigma_t^2 > 0, \Gamma_g^t \in \Sigma_{2k_g, *}^+(\mathbb{R}) \}$ and

For each component t (fixed effects) :

- ρ_t : probability of the component
- α_t : associated template parameter
- Γ_g^t : associated covariance matrix for deformation parameters
- σ_t^2 : associated additive noise variance

For each observation y_i (random effects) :

- τ_i : associated component
- β_i : deformation parameters
- ϵ_i : additive noise

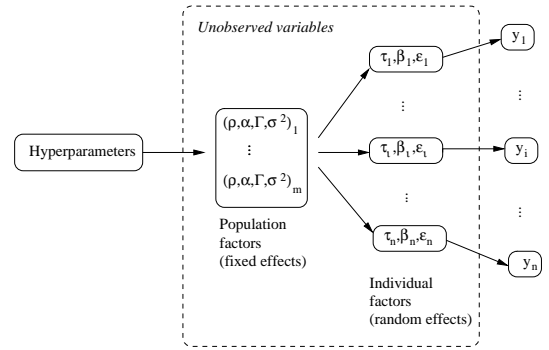


Fig. 1. Mixed effect structure for our BME-template

$\rho = (\rho_t)_{1 \leq t \leq m}$ to the open simplex ϱ . Here $\Sigma_{2k_g, *}^+(\mathbb{R})$ is the set of strictly positive symmetric matrices. Let $\eta = (\theta, \rho)$, the precise hierarchical Bayesian structure of our model is :

$$\left\{ \begin{array}{l} \rho \sim \nu_\rho \\ \theta = (\alpha_t, \sigma_t^2, \Gamma_g^t)_{1 \leq t \leq m} \sim \otimes_{t=1}^m (\nu_p \otimes \nu_g) \mid \rho \\ \tau_1^n \sim \otimes_{i=1}^n \sum_{t=1}^m \rho_t \delta_t \mid \rho, \\ \beta_1^n \sim \otimes_{i=1}^n \mathcal{N}(0, \Gamma_g^{\tau_i}) \mid \tau_1^n, \eta \\ y_1^n \sim \otimes_{i=1}^n \mathcal{N}(z_{\beta_i}, I_{\alpha_i}, \sigma_{\tau_i}^2 Id_\Lambda) \mid \beta_1^n, \tau_1^n, \eta \end{array} \right. \quad \text{with} \quad \left\{ \begin{array}{l} \nu_\rho(\rho) \propto \left(\prod_{t=1}^m \rho_t \right)^{a_\rho}, \\ \nu_p(d\sigma^2, d\alpha) \propto \left(\exp\left(-\frac{\sigma_0^2}{2\sigma^2}\right) \frac{1}{\sqrt{\sigma^2}} \right)^{a_p} \cdot \\ \quad \exp\left(-\frac{1}{2}\alpha^t (\Sigma_p)^{-1} \alpha\right) d\sigma^2 d\alpha, \\ \nu_g(d\Gamma_g) \propto \left(\exp(-\langle \Gamma_g^{-1}, \Sigma_g \rangle / 2) \frac{1}{\sqrt{|\Sigma_g|}} \right)^{a_g} d\Gamma_g, \end{array} \right.$$

where the hyper-parameters are fixed. All priors are the natural conjugate priors and assumed independent. A natural choice for the a priori covariance matrices Σ_p and Σ_g is to consider the matrices induced by the metric of the spaces V_p and V_g . Define the square matrices $M_p(k, k') = K_p(p_k, p_{k'}) \forall 1 \leq k, k' \leq k_p$ and $M_g(k, k') = K_g(g_k, g_{k'}) \forall 1 \leq k, k' \leq k_g$, and then set $\Sigma_p = M_p^{-1}$ and $\Sigma_g = M_g^{-1}$, which are typical prior matrices used in many matching algorithms.

3 Estimation of the parameters

The parameter estimates are obtained by maximising the posterior density on η conditional on y_1^n : $\hat{\eta}_n = \operatorname{argmax}_\eta q(\eta | y_1^n)$. Since the deformation coefficients β_1^n and component labels τ_1^n are unobserved, the natural approach is to use iterative algorithms such as EM [13] to maximise the posterior given the observations y_1^n . In the present context, we initialise the algorithm with the prior model η_0 .

3.1 Fast approximation with modes (FAM)

The expressions in the M step require the computation of expectations with respect to the posterior distribution of $\beta_1^n, \tau_1^n | y_1^n$ which is know up to the re-normalisation constant. To overcome this obstacle, given an observation y_i and a label t , the posterior distribution of the random deformation field is approximated at iteration l by a Dirac law on its mode $\beta_{l,i,t}^*$. This yields the following

$$\text{computation : } \beta_{l,i,t}^* = \operatorname{argmax}_\beta \log q(\beta | \alpha_{t,l}, \sigma_{t,l}^2, \Gamma_{g,l}^t, y_i) = \operatorname{argmin}_\beta \left\{ \frac{1}{2} \beta^t (\Gamma_{g,l}^t)^{-1} \beta + \frac{1}{2\sigma_{t,l}^2} |y_i - K_p^\beta \alpha_{t,l}|^2 \right\},$$

which is a standard template matching problem with the current parameters. We then approximate the joint posterior on (β_i, τ_i) as a discrete distribution concentrated at the m points $(\beta_{l,i,t}^*)_{1 \leq t \leq m}$ with weights given by: $w_{l,i}(t) \propto q(y_i | \beta_{l,i,t}^*, \alpha_{t,l}) q(\beta_{l,i,t}^* | \Gamma_{g,l}^t) \rho_{t,l}$. The label $\tau_{l,i}$ is then sampled from the distribution $\sum_{t=1}^m w_{l,i}(t) \delta_t$ and the deformation is the mode of the drawn label $\beta_{l,i} = \beta_{l,i,\tau_{l,i}}^*$.

3.2 Using a stochastic version of the EM algorithm : SAEM-MCMC

An alternative to the computation of the E-step in a complex nonlinear context is to use the stochastic approximation EM algorithm (SAEM) [14] coupled with an MCMC procedure [15] and a truncation on random boundaries. Our model belongs to the exponential density family which means that: $q(y, \beta, \tau, \eta) = \exp[-\psi(\eta) + \langle S(\beta, \tau), \phi(\eta) \rangle]$, where the sufficient statistic S is a Borel function on $\mathbb{R}^{2k_g} \times \{1, \dots, m\}$ taking its values in an open subset \mathcal{S} of \mathbb{R}^m and ψ, ϕ two Borel functions on $\Theta \times \varrho$ (the dependence on y is omitted for sake of simplicity).

We introduce the following function: $L : \mathcal{S} \times \Theta \times \varrho \rightarrow \mathbb{R}$ as $L(s; \eta) = -\psi(\eta) + \langle s, \phi(\eta) \rangle$. It has been proved in [1] that there exists a critical function $\hat{\eta} : \mathcal{S} \rightarrow \Theta \times \varrho$ which satisfies: $\forall \eta \in \Theta \times \varrho, \forall s \in \mathcal{S}, L(s; \hat{\eta}(s)) \geq L(s; \eta)$. Then, each iteration of this algorithm consists of the following four steps.

Simulation step: The missing data are drawn using a transition probability of a convergent Markov chain having the posterior distribution as stationary distribution: $(\beta_{l+1}, \tau_{l+1}) \sim \Pi_{\eta_l}((\beta_l, \tau_l), \cdot)$

Stochastic approximation step: Since the model is exponential, the stochastic approximation is done on the sufficient statistics using the simulated values of the missing data: $s_{l+1} = s_l + \Delta_{l+1}(S(\beta_{l+1}, \tau_{l+1}) - s_l)$, where $(\Delta_l)_l$ is a decreasing sequence of positive step-sizes.

Truncation step: A truncation is done on the stochastic approximation.

Maximisation step: The parameters are updated: $\eta_{l+1} = \hat{\eta}(s_{l+1})$.

Concerning the choice of Π_{η} used in the simulation step, as we aim to simulate (β_i, τ_i) through a transition kernel whose stationary distribution is $q(\beta, \tau | y_i, \eta)$, we simulate τ_i with a kernel whose stationary distribution is $q(\tau | y_i, \eta)$ and then β_i through a transition kernel that has $q(\beta | \tau, y_i, \eta)$ as stationary distribution. Given any initial deformation field $\xi_0 \in \mathbb{R}^{2k_g}$, we run, for each component t , J_l iterations of a hybrid Gibbs sampler $\Pi_{\eta, t}$ using the conditional prior distribution $\beta^j | \beta^{-j}$ as the proposal for the j^{th} coordinate, β^{-j} referring to β without its j^{th} coordinate. So that we get J_l elements $\xi_{t, i} = (\xi_{t, i}^{(k)})_{1 \leq k \leq J_l}$ of an ergodic homogeneous Markov chain whose stationary distribution is $q(\cdot | y_i, t, \eta)$. Denoting $\xi_i = (\xi_{t, i})_{1 \leq t \leq m}$, we simulate τ_i through the discrete density with weights

given by: $\hat{q}_{\xi_i}(t | y_i, \eta) \propto \left(\frac{1}{J_l} \sum_{k=1}^{J_l} \left[\frac{f_t(\xi_{t, i}^{(k)})}{q(y_i, \xi_{t, i}^{(k)}, t | \eta)} \right] \right)^{-1}$, where $f_t(\xi)$ is the density of the

Gaussian distribution $\mathcal{N}(0, \Gamma_{g, t})$. Then, we update β_i by re-running J_l times the hybrid Gibbs sampler Π_{η, τ_i} starting from a random initial point β_0 . It has been proved in [12], that the sequence $(\eta_l)_l$ generated through this algorithm converges a.s. toward a critical point of the penalised likelihood of the observations.

4 Single component model

We focus here on the single component model ($m = 1$). The unobserved variables are only the deformation fields β and the parameters are reduced to $\theta = (\alpha, \sigma^2, \Gamma_g)$. In this particular setting, denoting by P the distribution governing the observations and by $\Theta_* = \{ \theta_* \in \Theta \mid E_P(\log q(y | \theta_*)) = \sup_{\theta \in \Theta} E_P(\log q(y | \theta)) \}$, it has been proved in [1] that the MAP estimator $\hat{\theta}_n$ exists a.s. and converges toward an element in Θ_* . From the algorithmical viewpoint, the FAM algorithm does not require any changes. Indeed, each E step only corresponds to a single computation of the mode of the posterior density. Fortunately, the stochastic algorithm can be simplified. In the simulation step, only a single iteration of the Markov chain is needed for each iteration of the SAEM algorithm: $\beta_{l+1} \sim \Pi_{\theta_l}(\beta_l, \cdot)$ yielding a non homogeneous Markov chain. It has been proved in [11], that the sequence $(\theta_l)_l$ generated converges almost surely toward a critical point of the penalised likelihood of the observations.

5 Experiments

5.1 Estimation results

We illustrate this theoretical framework with the USPS handwritten digit database which corresponds to non noisy gray level images. In addition, we compare the two algorithmical approaches on 2D medical images of a part of the corpus calosum (the splenium) and part of the cerebellum.

Figure 2 shows the templates estimated from a training set (Figure 2-(a)) of 20 or 40 images per digit with both algorithms for the models with one and two components per class respectively. The results are quite similar, in particular the two components present the same features for both algorithms. Topologically different shapes are separated (cf digits 7 and 2) and the other digit clusters are relevant. While estimating a single component, the templates are a good representant of the shapes existing in the training set.

Concerning the geometrical variability, Figure 3, left image, presents some synthetic examples drawn with respect to the model with the estimated parameters. In spite of some artefacts described below, the kind of deformations learnt applied to the estimated templates looks like the elements of the training set.

Last but not least, one could wonder how those algorithms deal with noisy images. In [1], this particular case has been shown to fail with the FAM algorithm with a toy example. Whereas, in [11,12], the authors have proved the theoretical convergence of the two stochastic algorithms (for the mixture and simple models). This supports the fact that the estimated parameters should be less sensitive to the noise that can appear in the data. This is what we show in Figure 4 for a database of 20 images per digit which is partly presented. The results are related to the theory. Indeed, the FAM algorithm is stuck in some local maximum of the likelihood whereas the stochastic algorithm reaches a better estimator for the parameters. This illustrate the power of the stochastic approach to solve this problem. Both the template and the geometrical distribution are well estimated. The results are presented in Figure 4 and in the right image of Figure 3 where we can notice that the noise is separated from the estimation of the photometry and the geometry variability.

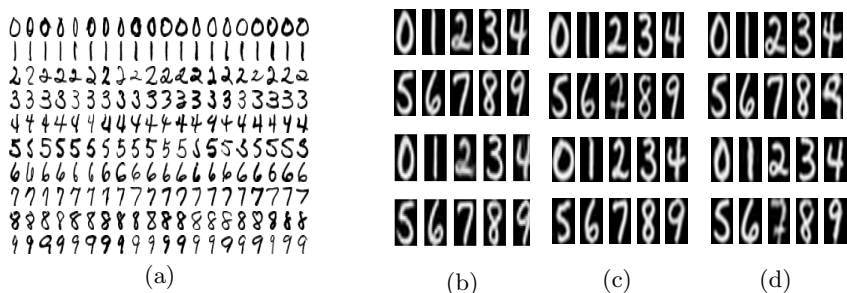


Fig. 2. (a) Some images of the USPS training set: 20 images per class. (b,c,d): Top row : FAM Algorithm, Bottom row : SAEM-MCMC algorithm. (b): one component prototype. (c-d): 2 component prototypes.

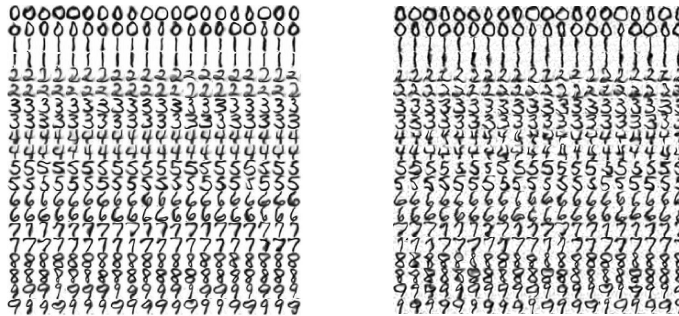


Fig. 3. 40 synthetic examples per class generated with the estimated parameters: 20 with the direct deformations and 20 with the inverse deformations. Left: from the non-noisy database estimated parameters. Right: from the noisy database estimated parameters. Note that the variability of digit is well reproduced, both in the case of highly deformable digits (e.g. 2 and 4) or in more constrained situations (e.g. 7 and 1).



Fig. 4. Left: Two images per digit of the noisy database. Middle: Estimated prototypes in a noisy setting $\sigma^2 = 1$. Left : with the FAM algorithm. Right : with the SAEM-MCMC coupling procedure.

The computational times of both algorithms for the simple model are very similar. The gradient descent required to compute the mode at each iteration last as long as one run of the Gibbs sampler used in the simulation step. For the general model, the SAEM-MCMC algorithm is longer since it requires the computation of many iterations of m Markov chains which can actually be easily parallelised. In addition, the number J_i of iterations of the Markov chain can be fixed all along the algorithm.

We also test the algorithms on some medical images. The database we consider has 47 2D images, each of them representing the splenium (back of the corpus calosum) and a part of the cerebellum.

The results of the estimation are presented in Figure 5 where we can see the improvement from the gray level mean to our estimations. The second image, corresponding to the deterministic algorithm result, shows a well contrasted splenium whereas the cerebellum remains a little bit blurry (note that it is still much better than the simple mean). The third image, corresponding to the

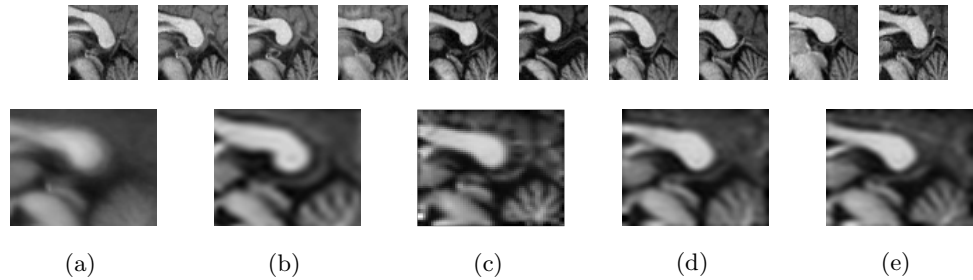


Fig. 5. Top row : Ten images of the training set representing the splenium and a part of the cerebellum. Bottom row : Results from the template estimation. (a,b,c) : gray level mean image of the 47 images and templates estimated with the FAM and the stochastic algorithms (respectively) on the simple model. (d,e) : two component estimated templates.

stochastic EM algorithm result, presents some real improvement again. Indeed, the splenium is still very contrasted, the background is not blurry and overall, the cerebellum is well reconstructed with several branches. The two anatomical shapes are relevant representants of the ones observed in the training set.

The estimation has been done while enabling the decomposition of the database into two components. The two estimated templates are presented in Figure 5 (d) and (e). The differences can be seen in particular on the shape of the splenium, where the fornix is more or less close to the boundary of the image and the thickness of the splenium varies. The number of branches in the two cerebella also tends to be different from one image to the other (4 in the first component and 5 in the second one). The estimation suffers from the small number of images we have. To be able to explain the huge variability of the two anatomical shapes, more components would be interesting but at the same time more images so that the components will not end up empty.

5.2 Optimisation on the representation, model and algorithms

Despite the fact that many parameters (e.g. the noise variance) are self-calibrated during the estimation process, the algorithm depends on some hyper-parameters we would like to discuss briefly.

Data representation Issues. The first point to be explain is the effect of the representation of the data, in particular the spline representation of both the template and the deformations (cf Section 2). We have chosen Gaussian kernels. The influence of their two scales can be seen on the template estimation. Indeed, choosing a too small geometric scale leads to very localised deformations around *fixed* control points and the resulting template is more blurry. In Figure 6, we present the results on a 20 handwritten digit images learning process. On the opposite side, a very large scale induces very smooth deformations which would no more be relevant for the kind of deformations required to explain the database.

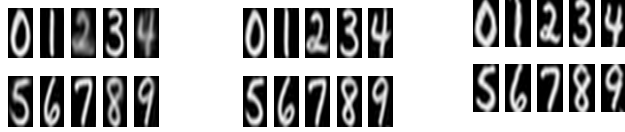


Fig. 6. Estimated prototypes (20 images per digit), $\sigma_g = 0.2$ (Left), $\sigma_g = 0.3$ (Right) with images in $[-1, 1]^2$.

Concerning the photometric scale, it is straightforward that a large scale will drive to blurry template. This is particularly noticeable on digit 1 where the thickness significantly increases (cf Figure 7 two left images).

In addition, the effects of increasing scale can also be noticed on the learnt covariance matrix. Given a fatty template, the deformations required to fit the database will be forced to contract the template. This phenomena is thus important in the learnt covariance matrix. When we generate new data thanks to the estimated parameters, we can see, as in Figure 7 right images, that the template is contracted, which is relevant, but also enlarged since the distribution on β is symmetric (this particular point is detailed in the next paragraph). Those large images are not typical from the training set.

Model distribution Issues. One question is the relevance of the Gaussian distribution chosen for the deformation field. It is natural to think that the mean of the deformations around an atlas is close to zero whereas the symmetry of the distribution is much more arguable. The probability for a deformation field $+\beta$ equals its opposite one $-\beta$. In Figure 3, we show the effects of the action of both fields on the learnt 10 digits templates. For example, digits 3 and 9 present for some samples irregular images whereas the opposite deformation leads to an image which is very similar to one or more element of the training set.

Another issue about the model is the choice of the prior hyper-parameters. In particular, the effect of the inverse Wishart prior a_g on the geometric covariance matrix is important. Indeed, if we would like to satisfy the theoretical requirements to the algorithms, we have to chose $a_g \geq 4k_g + 1$. however, the update formula is a barycenter between the expectation of the empirical covariance matrix and the prior with weights n and a_g respectively [1]. Since we are working with small sample sizes, this condition makes the update of Γ_g very constrained



Fig. 7. Two left images: Estimated prototypes of digit 1 (20 images per class) for different hyper-parameters. Left: smaller geometry and larger photometric scales. Right: larger geometry and smaller photometric scales. Right images: Synthetic examples corresponding to the two previous templates of digit 1.

close to the prior Σ_g . This does not enable the geometry to be well estimated and the effects can be seen directly on the template but also on the classification rate [1].

Stochastic Algorithm Issues. The FAM algorithm is deterministic and does not depend on any choice. Unfortunately, the stochastic algorithm requires several choices to optimise.

To optimise the choice of the transition kernel Π_η , we run the algorithm with different kernels and compare the evolution of the simulated hidden variables as well as the results on the estimated parameters. Some kernels, as an ordinary Hastings Metropolis algorithm using as proposal the prior or a standard random walk added to the current value, do not allow to visit well the entire support of the unobserved variable. From this point of view the hybrid Gibbs sampler we used has better properties and gives nice estimation results.

To prove the convergence of the stochastic algorithms, we have to suppose that as soon as the stochastic approximation wanders outside an increasing compact set, the unobserved variable needs to be projected inside a given compact set (this is the truncation on random boundaries). In practice however, this step is never required, the results presented were obtained without this control.

Finally, the initialisation of the parameters can lead to undesirable effects. For example, if the first value of the photometric parameter α is set to 0, at the first iteration of the Gibbs sampler, the proposal will be accepted with probability one. Since the candidate coordinates are simulated according to the conditional a priori, the resulting vector β leads to a variation which does not correspond to a relevant digit deformation. This implies some oscillations on the updated template. The next simulated deformation variable will try to take these oscillations into account to get closer and closer to the oscillating template, staying in its orbit. The results can be observed in Figure 6 (Right) specially for digit 1.

5.3 Results on classification rates

To get an objective way of comparing our algorithms and showing the performances, we use our model to propose a classifier which can easily be run on the USPS testset. We use the same approximations for the classification process, either a mode approximation of the posterior density or some MCMC methods to approximate the expectation required to compute the best class. Running the estimation with a FAM algorithm on all USPS database with 15 components and using a “mode” classifier gives a classification error rate of 3.5%. This is comparable to other classifiers results. The importance of the coupled photometric and geometric estimation is emphasised in [1].

Since the drawback of this method can be better proved in the presence of noise, we add an independent Gaussian noise of variance 1 on both the training set and the testset and run both estimations (with one component per class) and both classifications. We run the parameters estimation through the “SAEM-like” algorithm presented in the previous section. The classification error rate obtained are 22.52% when the classification uses the mode approximation and 17.07% using some MCMC methods. These results are a lot worse if the parameters are estimated with the FAM process. For example, the classification error reaches 40.71% when the classification is also done via the mode approximation.

6 Conclusion

We have presented some applications of the coherent statistical framework with BME-Template models described in [1, 11, 12]. This framework is fairly versatile and could be derived in many other important situations in CA. The possibility to work with mixture of deformable templates in a principled statistical way is also a quite enjoyable and unique feature of this setting. Reported experiments show that the deterministic FAM algorithm, despite its simplicity, performed significantly worse especially under noisy conditions than more sophisticated stochastic alternative. The introduction of such MCMC methods are still quite challenging in the 3D setting or for large deformation ([16] for a “FAM like” template estimation) but from an algorithmic point of view, there is a continuous interpolation from deterministic to stochastic algorithms (just increasing the number of MCMC steps) so that there is no sharp complexity gaps between to two approaches. Increasingly available computational power will make such stochastic approaches more and more appealing in the future.

References

1. Allasonnière, S., Amit, Y., Trouvé, A.: Toward a coherent statistical framework for dense deformable template estimation. *J.R.S.S.* **69** (2007) 3–29
2. Toga, W.A., Thompson, P.M.: The role of image registration in brain mapping. *Image and Vision Computing Journal* **19**(1–2) (2001) 3–24
3. Miller, M.I., Younes, L.: Group actions, homeomorphisms, and matching: A general framework. *Int. J. Comput. Vision* **41**(1–2) (2001) 61–84
4. Miller, M.I., Trouvé, A., Younes, L.: Geodesic shooting for computational anatomy. *J. Math. Imaging Vision* **24** (2006) 209–228
5. Amit, Y., Grenander, U., Piccioni, M.: Structural image restoration through deformable templates. *JASA* **86** (1991) 376–387
6. Grenander, U., Miller, M.: *Pattern theory: from representation to inference*. Oxford; New York: Oxford University Press (2007)
7. Glasbey, C.A., Mardia, K.V.: A penalised likelihood approach to image warping. *Journal of the Royal Statistical Society, Series B* **63** (2001) 465–492
8. Marsland, S., Twining, C.J., Taylor, C.J.: A minimum description length objective function for groupwise non-rigid image registration. *Im. Vis. Comp.* (2007)
9. Van Leemput, K.: Probabilistic Brain Atlas Encoding Using Bayesian Inference. *MICCAI* **1** (2006) 704–711
10. Lindstrom, M., Bates, D.: Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**(3) (1990) 673–687
11. Allasonnière, S., Kuhn, E., Trouvé, A.: Bayesian deformable models bulding via stochastic approximation algorithm: A convergence study. in revision
12. Allasonnière, S., Kuhn, E.: Stochastic algorithm for parameter estimation for dense deformable template mixture model. submitted
13. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* **1** (1977) 1–22
14. Delyon, B., Lavielle, M., Moulines, E.: Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.* **27**(1) (1999) 94–128
15. Kuhn, E., Lavielle, M.: Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM Probab. Stat.* **8** (2004) 115–131 (electronic)
16. Ma, J., Miller, M., Trouvé, A., Younes, L.: Bayesian template estimation in computational anatomy. To appear in *NeuroImage* (2008)