

Optimal Data-Driven Sparse Parameterization of Diffeomorphisms for Population Analysis

Sandy Durrleman, Marcel Prastawa, Guido Gerig, and Sarang Joshi

SCI Institute, University of Utah, 72 S. Central Campus Dr.,
UT-84112 Salt Lake City

Abstract. In this paper, we propose a novel approach for intensity based atlas construction from a population of anatomical images, that estimates not only a template representative image but also a common optimal parameterization of the anatomical variations evident in the population. First, we introduce a discrete parameterization of large diffeomorphic deformations based on a finite set of control points, so that deformations are characterized by a low dimensional geometric descriptor. Second, we optimally estimate the position of the control points in the template image domain. As a consequence, control points move to where they are needed most to capture the geometric variability evident in the population. Third, the optimal number of control points is estimated by using a $\log -L^1$ sparsity penalty. The estimation of the template image, the template-to-subject mappings and their optimal parameterization is done via a single gradient descent optimization, and at the same computational cost as independent template-to-subject registrations. We present results that show that the anatomical variability of the population can be encoded efficiently with these compact and adapted geometric descriptors.

1 Introduction

Fundamental to Computational Anatomy is the estimation of the template and the template-to-subject mappings that characterize anatomical variability in a population. For statistical analysis of a set of images it is crucial to efficiently parametrize the anatomical variability. One approach consists of extracting features of the mappings, like the Jacobian matrix of the deformation [3], or its determinant [9]. More comprehensive approaches rely on the statistical analysis of the displacement fields of the voxels grid by using a log-Euclidean technique [2], or on the Riemannian characterization of the group of diffeomorphisms parametrized via the continuous initial momenta map in the LDDMM setting [14].

The intrinsic problem with such approaches is that the fundamental anatomical variability is parametrized in an infinite dimensional space (practically on the order the size of the imaging modality), which does not reflect the intrinsic dimensionality of the anatomical variations in a finite database of images. Indeed, the estimated deformations are usually constrained to be spatially smooth, thus preventing every voxel from moving independently. Moreover, the need for deformation is not equally distributed over the domain, since intensity-based registrations

are mostly driven by the level sets of the template image. As a consequence, we hypothesize that the anatomical variability can be characterized by a compact geometric parametrization of much smaller dimension. It has already been shown in [7] that smooth vector fields parameterizing diffeomorphisms can be efficiently approximated by a small number of momenta via a “matching pursuit” technique, where momenta stands for vectors attached to control points.

In this paper we propose a control point parameterization of large deformation diffeomorphisms following [10] to drive template-to-subject image registrations. This differs from LDDMM image registration, for which the deformation is parameterized by a continuous map of momenta that are always parallel to the image gradient [12]. Here, we propose to use a finite set of momenta, which are not constrained in their direction. Control points techniques have been widely used for small deformation transformations [8], its use for large deformation matching of images is challenging. In [13], diffeomorphisms were built by a composition of small B-splines transforms without a comprehensive variational formulation. In [1], diffeomorphisms were characterized via a finite set of initial momenta located at the vertices of a “texture mesh”, but no attempt was made to estimate an optimal mesh describing a whole population of images. The inherent difficulty is to find an efficient way to transport information back and forth from source to target. Indeed, control points flow from source to target (via the deformation ϕ), whereas the variability in the population is in the target image domain and hence needs to be pulled back to the source to build the template ($I_{\text{src}} \circ \phi^{-1}$). We solve this issue via a new formulation of the dynamical system which drives the LDDMM registration, borrowed from optimal control theory. It enables us to easily transport points and vectors via simple integration of ODE. One of the striking results of this formulation is that the optimal positions of the control points in the template space can be found *at no additional computational cost*. We demonstrate that the control points are naturally attracted by the contours of the template image, and a regularity term optimizes their spatial distribution to avoid redundancy in the parameterization. The number of control points determines the number of degrees of freedom of the deformations and therefore the accuracy of the template-to-subject deformations and the sharpness of the atlas. To optimize the number of control points for a given atlas sharpness, we use a log L^1 -penalty term on the set of initial momenta, in the spirit of the in vogue sparse statistical methods. Our results show that this prior enables to prune the set of control points without sacrificing the description of the anatomical variability. This sparse and adapted parameterization of the variability seems therefore adapted to the statistical analysis of a collection of images.

We follow the now well established paradigm for atlas estimation by first defining pairwise image matching and then using it for population analysis.

2 Image Matching with Discrete Parameterization of Deformations

Image matching term. Let I_{src} and I_{tar} two images to be matched and ϕ a diffeomorphism of the image domain. The source image is deformed into I_1 via

$I_{\text{src}} \circ \phi^{-1}$. Let $\mathbf{y} = (y_1, \dots, y_M)$ be the concatenation of the location of the voxels in the target image. The deformed image I_1 linearly interpolates the gray levels of the neighboring voxels around positions $\phi^{-1}(y_k)$. We denote $\mathbf{y}(0) = \phi^{-1}(\mathbf{y})$.

The deformation estimation is driven by the L^2 norm between the images:

$$\|I_{\text{src}} \circ \phi^{-1} - I_{\text{tar}}\|^2 = \sum_{k=1}^M (I_{\text{src}}(y_k(0)) - I_{\text{tar}}(y_k))^2 = A(\mathbf{y}(0)), \quad (1)$$

which depends only on the positions $\mathbf{y}(0)$, since the target voxel positions \mathbf{y} are fixed on a regular lattice.

Non-linear diffeomorphic deformations. A standard way to construct non-linear diffeomorphisms is to integrate infinitesimal transformation parametrized via a time-varying vector field $v_t(x)$ over the time interval $[0, 1]$: $\dot{\phi}_t(x) = v_t(\phi_t(x))$, with initial condition $\phi_0(x) = x$. Under the conditions detailed in [4] and satisfied here, the resulting $(\phi_t)_{t \in [0,1]}$ is a flow of diffeomorphisms (for each time $t \in [0, 1]$, ϕ_t is a diffeomorphic deformation).

Let $\mathbf{c}_0 = \{c_1, \dots, c_N\}$ be a finite set of control points. These points move in space according to the deformation: $c_i(t) = \phi_t(c_i)$. By analogy with landmark matching [10], we parametrize the instantaneous velocity field $v_t(x)$ by a set of *time-varying* vectors $\boldsymbol{\alpha}(t) = \{\alpha_1(t), \dots, \alpha_N(t)\}$ as:

$$v_t(x) = \sum_{i=1}^N K(x, c_i(t))\alpha_i(t), \quad (2)$$

where K is an interpolating kernel, assumed hence forth without loss of generality, to be Gaussian: $K(x, y) = \exp(-|x - y|^2/\sigma^2)$.

The positions of the control points $c_i(t)$ depend on the velocities and therefore on the vectors $\boldsymbol{\alpha}(t)$. They satisfy a set of N coupled ODEs: $\dot{c}_i(t) = \dot{\phi}_t(c_i) = v_t(c_i) = \sum_{j=1}^N K(c_i(t), c_j(t))\alpha_j(t)$, with initial condition at $t = 0$: $c_i(0) = c_i$. In a matrix form, this can be written as: $\dot{\mathbf{c}}(t) = f(\mathbf{c}(t), \boldsymbol{\alpha}(t))$ with $\mathbf{c}(0) = \mathbf{c}_0$.

Once these positions have been estimated, the motion of any point x_0 is computed by solving the ODE: $\dot{x}(t) = v_t(x)$, $x(0) = x_0$. It follows that the flow of diffeomorphisms is entirely determined by the time-varying vectors $\boldsymbol{\alpha}(t)$ and the initial control points \mathbf{c}_0 .

In particular, the points $\phi_1^{-1}(y_k)$ in the data term are computed by flowing the positions y_k from $t = 1$ (target space) back to time $t = 0$ (source space): one integrates backward the ODE: $\dot{y}_k(t) = v_t(y_k) = \sum_{i=1}^N K(y_k(t), c_i(t))\alpha_i(t)$ with the final condition $y_k(1) = y_k$. This equation in matrix form becomes $\dot{\mathbf{y}} = g(\mathbf{y}(t), \mathbf{c}(t), \boldsymbol{\alpha}(t))$, $\mathbf{y}(1) = \mathbf{y}$. The solution at time $t = 0$ is $\mathbf{y}(0) = \phi_1^{-1}(\mathbf{y})$.

One defines the regularity of the deformation as its total kinetic energy: $\int_0^1 \|v_t\|_V^2 dt = \int_0^1 \sum_{i=1}^N \sum_{j=1}^N \alpha_i(t)^t K(c_i(t), c_j(t))\alpha_j(t) dt$, using the Sobolev norm of the velocity field associated with the kernel K . We write it as $\int_0^1 L(\mathbf{c}(t), \boldsymbol{\alpha}(t)) dt$.

Criterion minimization. Now, the matching criterion can be written as:

$$E(\mathbf{c}_0, \boldsymbol{\alpha}(t)) = A(\mathbf{y}(0)) + \gamma \int_0^1 L(\mathbf{c}(t), \boldsymbol{\alpha}(t)) dt \quad (3)$$

subject that:

$$\begin{aligned}\dot{\mathbf{c}}(t) &= f(\mathbf{c}(t), \boldsymbol{\alpha}(t)) & \mathbf{c}(0) &= \mathbf{c}_0 \\ \dot{\mathbf{y}}(t) &= g(\mathbf{y}(t), \mathbf{c}(t), \boldsymbol{\alpha}(t)) & \mathbf{y}(1) &= \mathbf{y}\end{aligned}\quad (4)$$

where γ is the scalar trade-off between the regularity of the deformation and the fidelity to data. The minimization of this criterion with respect to the time-varying vectors $\boldsymbol{\alpha}(t)$ leads to the source-to-target deformation parameterized by the control points \mathbf{c}_0 . The minimization with respect to the \mathbf{c}_0 leads to the optimal positions of the control points which parameterize the best matching possible. Both optimizations will be done via a single gradient descent.

It has been shown that the regularity term for $\gamma > 0$ ensures that the flow of diffeomorphisms builds a geodesic path in a certain group of diffeomorphisms [4]. This means that at the minimum the flow of diffeomorphisms is entirely determined by the initial momenta at time $t = 0$: $(\mathbf{c}_0, \boldsymbol{\alpha}(0))$.

We optimize the criterion via a gradient descent. The initial conditions are $\boldsymbol{\alpha}(t) = 0$ for all t (which gives $\phi_t(x) = x$, i.e. no deformation) and the control points \mathbf{c}_0 are given on a regular lattice. A variation of the momenta $\delta\boldsymbol{\alpha}(t)$ and of the control points initial position $\delta\mathbf{c}_0$ induces a variation of the whole path of the control points $\delta\mathbf{c}(t)$ from source ($t = 0$) to target ($t = 1$). Then, this changes the path of the target voxels flowing back to the source: it leads to a variation of the positions $\delta\mathbf{y}(0)$ and hence a variation of the criterion. We show in Appendix A, that the gradient of the criterion with respect to these two variables is:

$$\begin{cases} \nabla_{\boldsymbol{\alpha}} E(t) = 2\gamma\boldsymbol{\alpha}(t) + \boldsymbol{\eta}^c(t) + \tilde{\boldsymbol{\eta}}^y(t) \\ \nabla_{\mathbf{c}_0} E = \boldsymbol{\eta}^c(0) \end{cases}\quad (5)$$

where $\boldsymbol{\eta}^c$ and $\tilde{\boldsymbol{\eta}}^y$ are $3N$ dimensional vectors, which satisfy one forward and one backward integral equations:

$$\begin{aligned}\eta_p^y(t) &= -\nabla_{y_p(0)} A + \int_0^t \sum_{q=1}^N \frac{2}{\sigma^2} K(y_p(s), c_q(s)) \alpha_q(s)^t \eta_p^y(s) (y_p(s) - c_q(s)) ds \\ \eta_i^c(t) &= -\int_t^1 \sum_{j=1}^N \frac{2}{\sigma^2} K(c_i(s), c_j(s)) \left(\alpha_j(s)^t \eta_i^c(s) + \alpha_i(s)^t \eta_j^c(s) + 2\gamma\alpha_i(s)^t \alpha_j(s) \right) (c_i(s) - c_j(s)) \\ &\quad - \sum_{k=1}^M \frac{2}{\sigma^2} K(c_i(s), y_k(s)) \alpha_i(s)^t \eta_k^y(s) (c_i(s) - y_k(s)) ds \\ \nabla_{y_k(0)} A &= 2 \left(I_0(y_k(0)) - I_1(y_k) \right) \nabla_{y_k(0)} I_0\end{aligned}$$

and $\tilde{\boldsymbol{\eta}}^y$ is the solution of the set of N linear equations for all t :

$$\sum_{j=1}^N K(c_i(t), c_j(t)) \tilde{\eta}_j^y(t) = \sum_{k=1}^M K(c_i(t), y_k(t)) \eta_k^y(t)\quad (6)$$

The auxiliary variable $\boldsymbol{\eta}^y$ is computed forward in time: it transports the usual image force $\nabla_{y_k(0)} A$ from the source ($t = 0$) to the target space ($t = 1$). Once at time $t = 1$, this variable is used as a source term in the ODE satisfied by the

Algorithm 1. Image Matching with finite-dimensional parameterization

-
- 1: $\alpha_i(t) \leftarrow 0$ for all $i = 1, \dots, N$ and all t
 - 2: $\mathbf{c}_0 \leftarrow$ initial positions of control points (input)
 - 3: **repeat** {Gradient descent}
 - 4: {Compute path of control points (forward integration)}
 - 5: $c_i(t) = c_i(0) + \int_0^t \sum_{j=1}^N K(c_i(s), c_j(s)) \alpha_j(s) ds$
 - 6: {Compute deformed source image (backward integration)}
 - 7: $y_k(t) = y_k(1) - \int_t^1 \sum_{j=1}^N K(y_k(s), c_j(s)) \alpha_j(s) ds$
 - 8: {Compute gradient of source image}
 - 9: $\nabla_{y_k(0)} E = 2 \left(I_0(y_k(0)) - I_1(y_k) \right) \nabla_{y_k(0)} I_0$
 - 10: {Compute auxiliary variable η^y (forward integration)}
 - 11: $\eta_p^y(t) = -\nabla_{y_p(0)} E - \int_0^t \sum_{q=1}^N \alpha_q^t(s) \eta_p^y(s) \nabla_1 K(y_p(s), c_q(s)) ds$
 - 12: {Compute auxiliary variable η^c (backward integration)}
 - 13: $\eta_i^c(t) = \int_t^1 \sum_{j=1}^N \left(\alpha_j(s)^t \eta_i^c(s) + \alpha_i(s)^t \eta_j^c(s) + 2\gamma \alpha_i(s)^t \alpha_j(s) \right) \nabla_1 K(c_i(s), c_j(s)) +$
 $\sum_{k=1}^M \alpha_i(s)^t \eta_k^y(s) \nabla_2 K(y_k(s), c_i(s)) ds$
 - 14: {Solve the linear system}
 - 15: $\sum_{j=1}^N K(c_i(t), c_j(t)) \tilde{\eta}_j^y(t) = \sum_{k=1}^M K(c_i(t), y_k(t)) \eta_k^y(t)$ ($\tilde{\eta}^y$ is of dimension $3N$)
 - 16: {Compute gradient}
 - 17: $\nabla_i J(t) = 2\gamma \alpha_i(t) + \eta_i^c(t) + \tilde{\eta}_i^y(t)$
 - 18: {Update time-varying momenta}
 - 19: $\alpha_i(t) \leftarrow \alpha_i(t) - \varepsilon \nabla_{\alpha_i} J(t)$
 - 20: {Update initial positions of control points}
 - 21: $c_i(0) \leftarrow c_i(0) - \varepsilon \eta_i^c(0)$
 - 22: **until** Convergence
-

variable η^c , which is integrated backward in time. At time $t = 0$, this variable is used to update the position of the control points in the source image domain, at no additional cost. The overall gradient descent is summarized in Algorithm 1.

Remark 1 (Interpretation in the small deformation setting). To better understand these equations, we linearize this model in time. Then, the flow of diffeomorphisms is reduced to the transform: $\phi(x) = x + v(x)$ parameterized by the fixed momenta $(\mathbf{c}_0, \alpha(0))$. For small deformations, the inverse is approximated by $\phi^{-1}(y_k) = y_k - v(y_k)$. The matching criterion becomes:

$$E(\mathbf{c}_0, \boldsymbol{\alpha}) = \left\| I_{\text{src}} \circ \phi^{-1} - I_{\text{tar}} \right\|^2 + \gamma \|v\|_V^2, \quad (7)$$

whose gradient can be computed straightforwardly as:

$$\begin{aligned} \frac{1}{2} \nabla_{\alpha_i} E &= - \sum_{k=1}^M K(c_i, y_k) (I_{\text{src}}(y_k - v(y_k)) - I_{\text{tar}}(y_k)) \nabla_{y_k - v(y_k)} I_{\text{src}} + \gamma \sum_{j=1}^N K(c_i, c_j) \alpha_j \\ \frac{1}{2} \nabla_{c_i} E &= \sum_{k=1}^M \frac{2}{\sigma^2} K(c_i, y_k) (I_{\text{src}}(y_k - v(y_k)) - I_{\text{tar}}(y_k)) (\nabla_{y_k - v(y_k)} I_{\text{src}})^t \alpha_i (c_i - y_k) \\ &\quad - \gamma \sum_{j=1}^N \frac{2}{\sigma^2} K(c_i, c_j) \alpha_i^t \alpha_j (c_i - c_j) \end{aligned}$$

This is exactly the linearization of the gradient (5), at order 0 for the first equation and at order 1 for the second one (the zeroth order vanishing).

The first equation consists of two terms: the first one is the convolution of the usual image force at the control points, the second one is a regularizer of the estimated momenta, which can be seen as a low-pass filter on the momenta. The second equation is the update rule for the control points positions. The first term shows that they are attracted by the voxels where the gradient of the image is large (i.e. the contours), provided that the momenta α_i pushes in the ‘right’ direction, that of the image force (making the dot product positive). The second term is a repulsion term which moves away two control points which carry momenta pointing in the same direction, thus limiting the redundancy of the parameterization at the scale of the kernel σ .

3 Atlas Estimation with Common Sparse Parameterization of Deformations

3.1 Joint Optimization of the Template Image and the Deformations

Given a set of N_s images, one wants to find a template image I_0 , the optimal set of control points \mathbf{c}_0 in the image domain and the optimal set of N_s time-varying vectors $\alpha_s(t)$, which drive the registration of the template to every image. As a consequence, the criterion to be minimized is given by:

$$E(I_0, \mathbf{c}_0, \alpha_1(t), \dots, \alpha_{N_s}(t)) = \sum_{s=1}^{N_s} \left\{ A_s(\mathbf{y}_s(0)) + \gamma \int_0^1 L(c_s(t), \alpha_s(t)) ds \right\} \quad (8)$$

subject that:

$$\begin{aligned} \dot{\mathbf{c}}_s(t) &= f(\mathbf{c}_s(t), \alpha_s(t)) & \mathbf{c}_s(0) &= \mathbf{c}_0 \\ \dot{\mathbf{y}}_s(t) &= g(\mathbf{y}_s(t), \mathbf{c}_s(t), \alpha_s(t)) & \mathbf{y}_s(1) &= \mathbf{y} \\ A_s(\mathbf{y}_s(0)) &= \|I_0(\mathbf{y}_s(0)) - I_s\|^2 \end{aligned} \quad (9)$$

where we notice that the initial and final condition of the ODEs are shared among the population. This criterion has the same form as in (3). Denoting E_s each term of the sum, the gradient with respect to \mathbf{c}_0 and the $\alpha_s(t)$ is given as:

$$\nabla_{\alpha_s(t)} E = \nabla_{\alpha_s(t)} E_s \quad \nabla_{\mathbf{c}_0} E = \sum_{s=1}^{N_s} \nabla_{\mathbf{c}_0} E_s \quad (10)$$

This means that the weighting vectors are computed for N_s parallel registrations by integrating the ODEs of (5). After the back and forth integrations, the auxiliary variables $\eta_s^c(0)$ are averaged over the subjects to update the common parameterization of the deformations \mathbf{c}_0 .

The gradient with respect to the template image I_0 is the sum of the gradient $\nabla_{I_0} A_s(\mathbf{y}_s(0))$. The value $I_0(y_k(0))$ is computed as $\sum_{p \in \mathcal{N}(y_k(0))} \rho_p(y_k(0)) I_0(\pi_p(y_k(0)))$, where $\mathcal{N}(y)$ denotes the set of 8 neighboring voxels $\pi_p(y)$ of the position y and $\rho_p(y)$ the weights corresponding to the trilinear interpolation. Let

R_s be the residual image $I_0(\mathbf{y}(0)) - I_s$. Then a variation δI_0 of the template image leads to:

$$\begin{aligned} \frac{1}{2}\delta A(\mathbf{y}(0)) &= \sum_i R_s(y_i) \sum_{p \in \mathcal{N}(y_i(0))} \rho_p(y_i(0)) \delta I_0(\pi_p(y_i(0))) \\ &= \sum_j \left(\sum_{\{i; \exists k, \pi_k(y_i(0))=y_j\}} \rho_k(y_i(0)) R_s(y_i) \right) \delta I_0(y_j) = \sum_j \nabla_{I_0} A_s(y_j) \delta I_0(y_j) \end{aligned}$$

where one multiplies the gray levels for each voxel in the source domain instead of the target domain. This shows that the gradient of A is the partial volume projection [6] of the residual image. This is computed by flowing the voxel y_k back in the source domain and distributing its gray level at the 8 neighboring voxels with the same weight as for a trilinear interpolation. The accumulation of such values for every voxel and every subject gives the gradient.

3.2 Sparsity Enforced by $\log -L^1$ Penalty Term

The number of control points determines the number of degrees of freedom of the deformations. The more control points, the more accurate the registrations, the sharper the template image. Consequently, the optimization of the previous criterion tends to use as many control points as possible. However, as we will see in our experiments, from a certain point, adding more momenta only marginally increase the atlas sharpness at the cost of adding much more noise in the description of the variability. To find the optimal number of degrees of freedom for an accurate description of the variability, we use a numerically stabilized $\log -L^1$ penalty motivated by [5] :

$$E = \sum_{s=1}^{N_s} \left\{ A_s(\mathbf{y}_s(0)) + \gamma \int_0^1 L(c_s(t), \boldsymbol{\alpha}_s(t)) ds + \gamma_{\text{sp}} \sum_{i=1}^N \log^c (\|\boldsymbol{\alpha}_i^s(0)\|) \right\} \quad (11)$$

where N denotes the total number of control points, $\boldsymbol{\alpha}_i^s$ the i th initial momentum vector of the s th subject. For numerical stability we truncate the log function near 0 via: $\log^c(x) = \max(\log(x), \log(c)) - \log(c)$ a positive penalty. This penalty function has almost no effect on large momenta, whereas it enforces the small momenta to converge to zero. The constant c is a threshold to avoid numerical instability, set typically at the voxel size. Every 5 iterations of the gradient descent, control points with an initial momenta smaller than c are pruned.

This penalty induces only a small change in the algorithm. The quantity $\gamma_{\text{sp}} \boldsymbol{\alpha}_i^s / \|\boldsymbol{\alpha}_i^s\|^2$ is added to each $\nabla_{\boldsymbol{\alpha}_i} E^s$ as soon as $\|\boldsymbol{\alpha}_i^s\| > c$.

4 Experiments

Our method is independent of the dimension of the images. In this section, we focus on 2D images for a better visualization and understanding of the results.

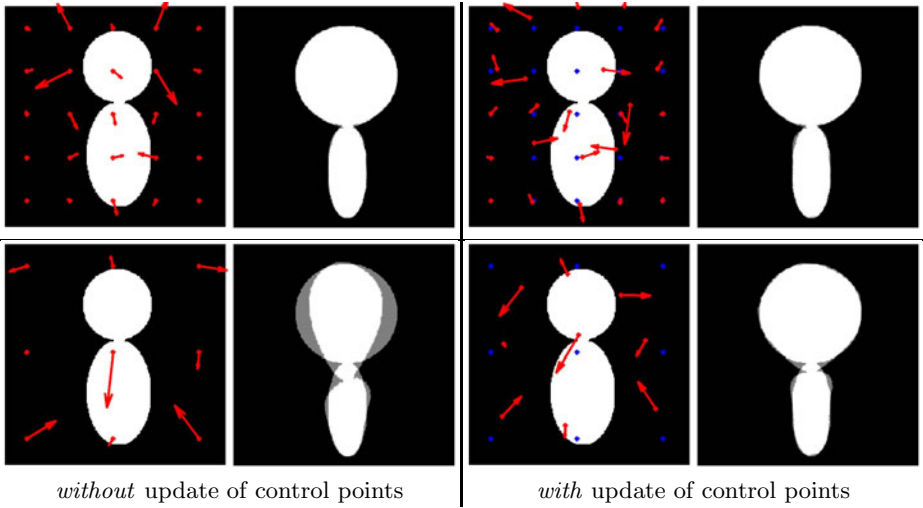


Fig. 1. Synthetic image matching with 25 (top) and 9 (bottom) control points. On the left of each panel: the source image with the initial momenta (red arrows). On the right the superimposition of the deformed source and target image. First row shows that a *discrete* parameterization is sufficient for a perfect matching. Second row shows that moving the control points to their optimal positions gives a much better representation of the shape differences for a fixed number of parameters. $\sigma = 50$ voxels, $\gamma = 10^{-2}$.

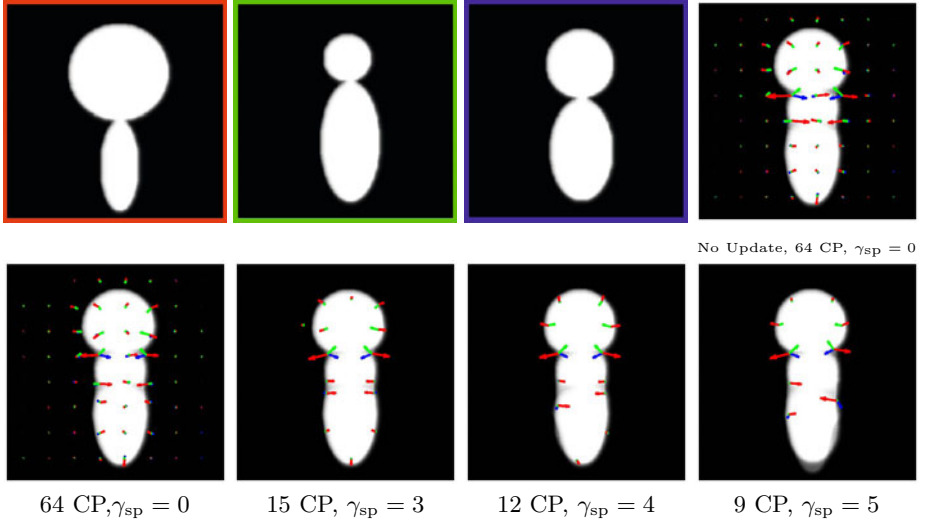


Fig. 2. Atlas from 3 images (top left). Template and initial momenta shown without (top right, $A = 163$) and with (bottom left, $A = 127$) update of control points. The sparsity term drastically reduces the number of control points, for a comparable atlas sharpness: data term $A = 134$ and 206 for $\gamma_{sp} = 3$ and 4 . For $\gamma_{sp} = 5$, the atlas sharpness worsens ($A = 393$), as noticeable at the bottom of the image. $\gamma = 0.1$, $\sigma = 30$ voxels.

In Fig. 1, we show the matching of two synthetic 2D images of size 256×256 . It shows that a *discrete* parameterization (with the maximum number of degrees of freedom: one control point every σ) enables a perfect matching, as would do a parameterization by a *continuous* momenta map as in [4]. Using much fewer momenta, the matching is less accurate. But, moving the control points at their optimal position near the contours drastically increases the matching accuracy.

In Fig. 2, we construct an atlas from 3 synthetic images. The sparsity prior enables to select the most important momenta and to give a compact representation of the shape variability. Fig. 3 shows that the initial number of control points can be divided by 5 without sacrificing much of the atlas sharpness.

In Fig. 4, we show the motion of the control points to the contours of the source image during a registration between MRI slices of size 176×256 . In Fig. 5, we construct an atlas from 5 of such images. The sparsity prior shows that the main variations in the population are located at the skull, near the ventricles and the main sulci, like the frontal sulcus. These structures are indeed the most salient and variable at the scale of analysis: $\sigma = 5$ voxels.

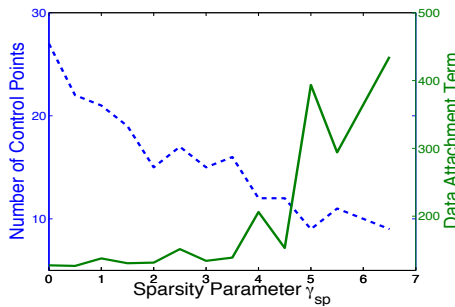


Fig. 3. Graph of the the number of control points and the fidelity-to-data term with respect to the sparsity parameter γ_{sp} with data of Fig. 2. There is a whole range $\gamma_{sp} \in]0, 4]$ for which the number of control points can be divided by 2.4 to 5.3 (from 27 to 12 compared to the initial 64 control points) for an almost constant atlas sharpness.

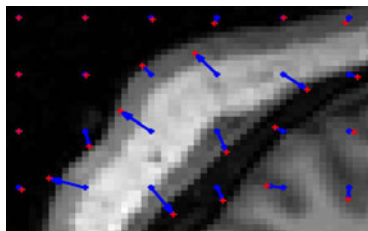


Fig. 4. Registration of brain MRI slices: close-up on the left posterior area of the source image. Initial control points in blue moved to the positions shown in red. They moved toward the contours of the image. They do not move in absence of image force in homogeneous areas. $\sigma = 5$ voxels and $\gamma = 10^{-3}$.

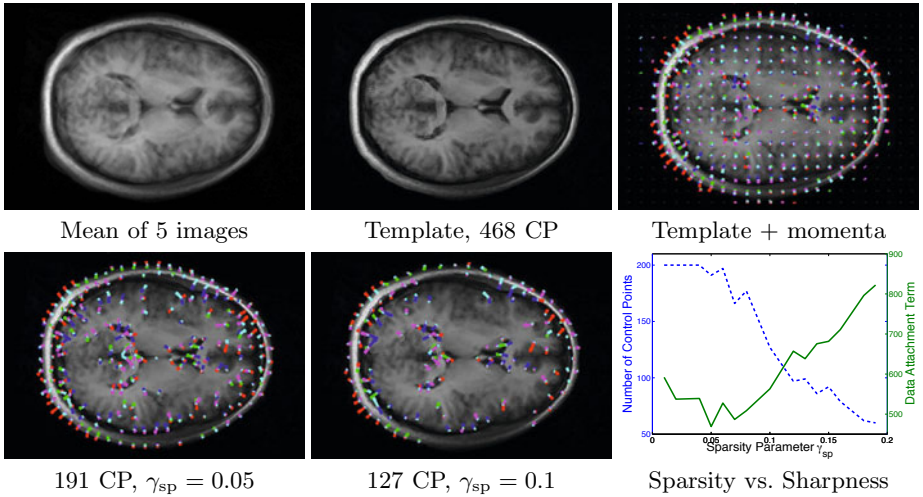


Fig. 5. Atlas from 5 brain images. Top row: estimated atlas with update of control points but without sparsity enforced. Bottom row: with enforced sparsity, the control points focus on the skull, the ventricles and the major sulci. The graph shows that we can achieve an equivalent description of the variability with 40.8% to 27.1% of the original 468 momenta, for $\gamma_{sp} \in]0, 0.1]$. $\sigma = 5$, $\gamma = 5 \cdot 10^{-3}$.

5 Discussion and Conclusion

In this paper, we present a new method for parameterizing large and dense image deformations via a discrete set of control points. Given a set of images, we estimate the template image, the template-to-subject deformations and their parameterization via the optimal placement of the control points and the optimal number of them according to a sparse prior. The whole estimation is posed as a single optimization problem and is solved by a single gradient descent. This is more controllable and more efficient than usual alternated minimizations. The algorithm requires only a solution of ODEs and linear systems; no heuristic rule is used to update the control points. By contrast, a death/birth procedure of control points could have been investigated, but at the cost of more heuristic and arbitrary priors. Our results show that the anatomical variability can be efficiently described by small number of well-placed momenta. We expect to show in the future that these new parameterizations substantially improve the statistical analysis of sets of 3D anatomical images. Future work will extend this framework to include the construction of geodesic diffeomorphisms by integration of Hamiltonian systems, as initiated in [1,11] Future work will also focus on the automatic estimation of the best trade-offs between atlas sharpness, sparsity and regularity of the deformations by adding priors on γ and γ_{sp} .

Acknowledgments. This work was supported by NIH grants: NIBIB (5R01 EB007 688), NCRRR (P41 RR023953), ACE-IBIS (RO1 HD055741), and NA-MIC (U54 EB005149).

References

1. Allassonnière, S., Trouvé, A., Younes, L.: Geodesic shooting and diffeomorphic matching via textured meshes. In: Rangarajan, A., Vemuri, B.C., Yuille, A.L. (eds.) EMMCVPR 2005. LNCS, vol. 3757, pp. 365–381. Springer, Heidelberg (2005)
2. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-euclidean framework for statistics on diffeomorphisms. In: Larsen, R., Nielsen, M., Sparring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 924–931. Springer, Heidelberg (2006)
3. Ashburner, J., Hutton, C., Frackowiak, R., Johnsrude, I., Price, C., Friston, K.: Identifying global anatomical differences: deformation-based morphometry. *Human Brain Mapping* 6(5-6), 348–357 (1998)
4. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *IJCV* 61, 139–157 (2005)
5. Candès, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted L^1 minimization. *Journal of Fourier Analysis and Applications* 14(5), 877–905 (2008)
6. Durrleman, S.: Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution. Thèse de sciences (phd thesis), Université de Nice-Sophia Antipolis (March 2010)
7. Durrleman, S., Pennec, X., Trouvé, A., Ayache, N.: Statistical models of sets of curves and surfaces based on currents. *Med. Im. Anal.* 13(5), 793–808 (2009)
8. Glasbey, C.A., Mardia, K.V.: A penalised likelihood approach to image warping. *Journal of the Royal Statistical Society, Series B* 63, 465–492 (2001)
9. Gogtay, N., Lu, A., Leow, A.D., Klunder, A.D., Lee, A.D., Chavez, A., Greenstein, D., Giedd, J.N., Toga, A.W., Rapoport, J.L., Thompson, P.M.: 3D growth pattern abnormalities visualized in childhood-onset schizophrenia using tensor-based morphometry. *Proc. Natl. Acad. Sci.* 105(41), 15979–15984 (2008)
10. Joshi, S., Miller, M.: Landmark matching via large deformation diffeomorphisms. *IEEE Transaction on Image Processing* 9(8), 1357–1370 (2000)
11. Marsland, S., McLachlan, R.: A hamiltonian particle method for diffeomorphic image registration. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 396–407. Springer, Heidelberg (2007)
12. Miller, M., Trouvé, A., Younes, L.: Geodesic shooting for computational anatomy. *Journal of Mathematical Imaging and Vision* 24(2), 209–228 (2006)
13. Rueckert, D., Aljabar, P., Heckemann, R.A., Hajnal, J.V., Hammers, A.: Diffeomorphic Registration Using B-Splines. In: Larsen, R., Nielsen, M., Sparring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 702–709. Springer, Heidelberg (2006)
14. Singh, N., Fletcher, P., Preston, J., Ha, L., King, R., Marron, J., Wiener, M., Joshi, S.: Multivariate statistical analysis of deformation momenta relating anatomical shape to neuropsychological measures. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 529–537. Springer, Heidelberg (2010)

A Differentiation of the Criterion

A variation of the momenta $\delta\alpha(t)$ and the initial position of the control points $\delta\mathbf{c}_0$ induces a variation of the path of the control points $\delta\mathbf{c}(t)$ and then of the voxel positions $\delta\mathbf{y}(t)$. This induces the variation of the criterion:

$$\delta E = \nabla_{\mathbf{y}(0)} A^t \delta \mathbf{y}(0) + \gamma \int_0^1 (\partial_1 L(t) \delta \mathbf{c}(t) + \partial_2 L(t) \delta \alpha(t)) dt \quad (12)$$

The variations $\delta\mathbf{c}(t)$ and $\delta\mathbf{y}(t)$ satisfy the linearized ODEs:

$$\begin{aligned}\dot{\delta\mathbf{c}}(t) &= \partial_1 f(\mathbf{c}(t), \boldsymbol{\alpha}(t))\delta\mathbf{c}(t) + \partial_2 f(\mathbf{c}(t), \boldsymbol{\alpha}(t))\delta\boldsymbol{\alpha}(t) \\ \dot{\delta\mathbf{y}}(t) &= \partial_1 g(\mathbf{y}(t), \mathbf{c}(t), \boldsymbol{\alpha}(t))\delta\mathbf{y}(t) + \partial_2 g(\mathbf{y}(t), \mathbf{c}(t), \boldsymbol{\alpha}(t))\delta\mathbf{c}(t) + \partial_3 g(\mathbf{y}(t), \mathbf{c}(t), \boldsymbol{\alpha}(t))\delta\boldsymbol{\alpha}(t)\end{aligned}$$

with $\delta\mathbf{c}(0) = \delta\mathbf{c}_0$ and $\delta\mathbf{y}(1) = 0$. Let $R_{st} = \exp(\int_s^t \partial_2 f(u)du)$ and $V_{st} = \exp(\int_s^t \partial_1 g(u)du)$. The solution of these linear ODEs with source terms are:

$$\delta\mathbf{c}(t) = R_{0t}\delta\mathbf{c}_0 + \int_0^t R_{st}\partial_2 f(s)\delta\boldsymbol{\alpha}(s)ds \text{ and } \delta\mathbf{y}(t) = -\int_t^1 V_{st}\left(\partial_2 g(s)\delta\mathbf{c}(s) + \partial_3 g(s)\delta\boldsymbol{\alpha}(s)\right)ds$$

Plugging these equations into (12) and using Fubini's theorem leads to:

$$\begin{aligned}\delta E &= \int_0^1 \left(\gamma\partial_1 L(t) + \nabla_{\mathbf{y}(0)} A^t V_{t0}\partial_2 g(t)\right)R_{0t}dt\delta\mathbf{c}_0 \\ &+ \int_0^1 \left(\gamma\partial_2 L(t) + \int_t^1 \gamma\partial_1 L(s)R_{ts}\partial_2 f(t)ds + \nabla_{\mathbf{y}(0)} A^t \left(V_{t0}\partial_3 g(t) + \int_t^1 V_{s0}\partial_2 g(s)R_{ts}\partial_2 f(t)ds\right)\right)\delta\boldsymbol{\alpha}(t)dt\end{aligned}$$

Therefore the gradient of E with respect to the L^2 metric is given as:

$$\nabla_{\boldsymbol{\alpha}} E(t) = \gamma\partial_2 L(t)^t + \partial_3 g(t)^t \boldsymbol{\eta}^y(t) + \partial_2 f(t)^t \boldsymbol{\eta}^c(t) \text{ and } \nabla_{\mathbf{c}_0} E = \boldsymbol{\eta}^c(0) \quad (13)$$

where we have denoted:

$$\boldsymbol{\eta}^y(t) = V_{t0}^t \nabla_{\mathbf{y}(0)} A \text{ and } \boldsymbol{\eta}^c(t) = \int_t^1 R_{ts}^t \left(\gamma\partial_1 L(s)^t + \partial_2 g(s)^t \boldsymbol{\eta}^y(s)\right)ds \quad (14)$$

Since $V_{t0} = \text{id} - \int_0^t V_{s0}\partial_1 g(s)ds$ and $R_{ts} = \text{id} + \int_t^s R_{us}\partial_1 f(u)du$, we have:

$$\begin{aligned}\boldsymbol{\eta}^y(t) &= -\nabla_{\mathbf{y}(0)} A - \int_0^t \partial_1 g(s)^t \boldsymbol{\eta}^y(s)ds \\ \boldsymbol{\eta}^c(t) &= \int_t^1 \left(\gamma\partial_1 L(s)^t + \partial_2 g(s)^t \boldsymbol{\eta}^y(s) + \partial_1 f(s)^t \boldsymbol{\eta}^c(s)\right)ds\end{aligned} \quad (15)$$

Given the definitions of f, g and L , the L^2 -gradient can be written as:

$$\nabla_{\boldsymbol{\alpha}}^{L^2} E(t) = 2\gamma\mathbf{k}(\mathbf{c}(t), \mathbf{c}(t))\boldsymbol{\alpha}(t) + \mathbf{k}(\mathbf{c}(t), \mathbf{c}(t))\boldsymbol{\eta}^c(t) + \mathbf{k}(\mathbf{y}(t), \mathbf{c}(t))\boldsymbol{\eta}^y(t) \quad (16)$$

where $\mathbf{k}(\mathbf{y}, \mathbf{c})$ denotes the $3M$ -by- $3N$ block matrix whose (i, j) th-block is $K(y_i, c_j)$.

The Sobolev gradient associated to the metric $\mathbf{k}(\mathbf{c}(t), \mathbf{c}(t))$ is therefore:

$$\nabla_{\boldsymbol{\alpha}}^{\text{Sob}} E(t) = 2\gamma\boldsymbol{\alpha}(t) + \boldsymbol{\eta}^c(t) + \mathbf{k}(\mathbf{c}(t), \mathbf{c}(t))^{-1}\mathbf{k}(\mathbf{y}(t), \mathbf{c}(t))\boldsymbol{\eta}^y(t) \quad (17)$$

which requires to solve a linear system. If control points gets closer than σ , then the matrix $\mathbf{k}(\mathbf{c}(t), \mathbf{c}(t))$ is badly conditioned [6]. In this case, we adjust σ in this matrix to the minimal distance between control points and use the gradient $\mathbf{k}^{\text{adj}}(\mathbf{c}(t), \mathbf{c}(t))^{-1}\nabla^{L^2} E$, which interpolates between the L^2 gradient ($\sigma_{\text{adj}} \rightarrow 0$) and the Sobolev gradient ($\sigma_{\text{adj}} \rightarrow \sigma$).