Supplementary Material to the MICCAI 2012 paper entitled "Topology Preserving Atlas Construction from Shape Data without Correspondence using Sparse Parameters"

Stanley Durrleman, Marcel Prastawa, Julie R. Korenberg
Sarang Joshi, Alain Trouvé and Guido Gerig

1 Objective Function for Atlas Construction

In this document, we provide the differentiation of the L^2 part of the criterion for shape atlas construction. With the notations of the main paper [1], the differentiable part of the criterion (omitting the sparsity penalty) writes:

$$E(\mathbf{X}_0, \mathbf{c}_0, \{\boldsymbol{\alpha}_0^i\}) = \sum_{i=1}^{N_{\text{subj}}} \left\{ \frac{1}{2\sigma^2} D(\mathbf{X}_0^i(1), \mathbf{X}_i) + L(\mathbf{S}_0^i) \right\} \tag{1}$$

subject to:

$$\begin{cases} \dot{\mathbf{S}}^i(t) = F(\mathbf{S}^i(t)) & \mathbf{S}^i(0) = \{\mathbf{c}_0, \boldsymbol{\alpha}_0^i\} \\ \dot{\mathbf{X}}_0^i(t) = G(\mathbf{X}_0^i(t), \mathbf{S}^i(t)) & \mathbf{X}_0^i(0) = \mathbf{X}_0 \end{cases} \tag{2}$$

where

$$L(\mathbf{S}_0^i) = \sum_{p,q}^{N_c} {\alpha_{0,p}^i}^t K(c_{0,p}, c_{0,q}) \alpha_{0,q}^i \tag{3}$$

Assuming the ambient space is of dimension 3, $\mathbf{X}$ is a vector of length $3N_x$, where N_x is the number of points in the template shape, $\mathbf{c}$ and $\boldsymbol{\alpha}$ are two vectors of length $3N_c$ each, where N_c is the number of control points, so that $\mathbf{S}$ is a vector of length $6N_c$.

$F(\mathbf{S}) = \begin{pmatrix} F^c(\mathbf{c}, \boldsymbol{\alpha}) \\ F^\alpha(\mathbf{c}, \boldsymbol{\alpha}) \end{pmatrix}$ is a vector of length $6N_c$, which is decomposed into two vectors of size $3N_c$. The kth coordinate (among N_c) of F^c and F^α is the

3D vector:

$$\begin{aligned} F^c(\mathbf{S})_k &= \sum_{p=1}^{N_c} K(c_k(t), c_p(t))\alpha_p(t) \\ F^\alpha(\mathbf{S})_k &= -\sum_{p=1}^{N_c} \alpha_k(t)^t \alpha_p(t) \nabla_1 K(c_k(t), c_p(t)) \end{aligned} \tag{4}$$

$G(\mathbf{X}, \mathbf{S})$ is a vector of size $3N_x$. Its kth coordinate (among N) is the 3D vector:

$$G(\mathbf{X}, \mathbf{S})_k = \sum_{p=1}^{N_c} K(x_k(t), c_p(t))\alpha_p(t) \tag{5}$$

2 Differentiation of the Objective Function

2.1 Gradient of the Objective Function

In this section, we aim to prove that the gradient of the objective function with respect to the initial state $\mathbf{S}_{0,i} = \{\mathbf{c}, \boldsymbol{\alpha}_i\}$ and the template shape $\mathbf{X}_0$ is given as:

$$\begin{aligned} \nabla_{\boldsymbol{\alpha}_0^i} E &= \xi^{\alpha,i}(0) + \nabla_{\boldsymbol{\alpha}_0^i} L(\mathbf{c}_0, \boldsymbol{\alpha}_0^i) \\ \nabla_{\mathbf{c}_0} E &= \sum_{i=1}^{N_{\text{subj}}} \left\{ \xi^{c,i}(0) + \nabla_{\mathbf{c}_0} L(\mathbf{c}_0, \boldsymbol{\alpha}_0^i) \right\} \\ \nabla_{\mathbf{X}_0} E &= \sum_{i=1}^{N_{\text{subj}}} \theta^i(0) \end{aligned} \tag{6}$$

where the auxiliary variables $\xi^i(t) = \{\xi^{c,i}(t), \xi^{\alpha,i}(t)\}$ (of the same size as $\mathbf{S}^i(t)$) and $\theta^i(t)$ (of the same size as $\mathbf{X}_0$) satisfy the linear ODEs:

$$\begin{cases} \dot{\theta}^i(t) = -(\partial_1 G(\mathbf{X}^i(t), \mathbf{S}^i(t)))^t \theta^i(t) \\ \theta^i(1) = \nabla_{\mathbf{X}^i(1)} D_X(\mathbf{X}^i(1), \mathbf{X}_i) \end{cases} \tag{7}$$

$$\begin{cases} \dot{\xi}^i(t) = -\Big(\partial_2 G(\mathbf{X}^i(t), \mathbf{S}^i(t))^t \theta^i(t) + d_{\mathbf{S}^i(t)} F^t \xi^i(t) \Big) \\ \xi^i(1) = 0 \end{cases} \tag{8}$$

2.2 Proof

The differentiation of the criterion (1) can be done for each subject i independently. Therefore, we differentiate only one term of the sum in (1) for a generic subject's index i that we omit in the following for clarity purposes.

A small perturbation $\delta\mathbf{S}_0$ of the initial state of the system induces a perturbation of the motion of the particles $\delta\mathbf{S}(t)$, which, in turn, induces a perturbation of the template mapping $\delta\mathbf{X}(t)$ and of the criterion δE, which writes thanks to the chain rule:

$$\delta E = \left(\nabla_{\mathbf{X}(1)} D\right)^t \delta\mathbf{X}(1) + \left(\nabla_{\mathbf{S}_0} L\right)^t \delta\mathbf{S}_0. \tag{9}$$

According to (2), the perturbations $\delta\mathbf{S}(t)$ and $\delta\mathbf{X}(t)$ satisfy the linearized ODEs:

$$\begin{aligned} \dot{\delta\mathbf{S}}(t) &= d_{S(t)}F\delta\mathbf{S}(t) & \delta\mathbf{S}(0) &= \delta\mathbf{S}_0 \\ \dot{\delta\mathbf{X}}(t) &= \partial_1 G\delta\mathbf{X}(t) + \partial_2 G\delta\mathbf{S}(t) & \delta\mathbf{X}(0) &= \delta\mathbf{X}_0 \end{aligned}$$

The first ODE is linear. Its solution is given by:

$$\delta\mathbf{S}(t) = \exp\left(\int_0^t d_{\mathbf{S}(u)}F du\right)\delta\mathbf{S}_0. \tag{10}$$

The second ODE is linear with source term. Its solution is given by:

$$\delta\mathbf{X}(t) = \int_0^t \exp\left(\int_u^t \partial_1 G ds\right)\partial_2 G(u)\delta\mathbf{S}(u)du + \exp(\int_0^1 \partial_1 G(s)ds)\delta\mathbf{X}_0 \tag{11}$$

Plugging (10) into (11) and then into (9) leads to:

$$\begin{cases} \nabla_{\mathbf{S}_0} E = \displaystyle\int_0^1 R_{0t}^t \partial_2 G(\mathbf{X}(t), S(t))^t V_{t1}^t \nabla_{\mathbf{X}(1)} D + \nabla_{\mathbf{S}_0} L \\ \nabla_{\mathbf{X}_0} E = V_{01}^t \nabla_{\mathbf{X}(1)} D \end{cases}, \tag{12}$$

where we denoted $R_{st} = \exp\left(\int_s^t d_{\mathbf{S}(u)}F du\right)$ and $V_{st} = \exp\left(\int_s^t \partial_1 G(\mathbf{X}(u), S(u))du\right)$.

Let us denote $\theta(s) = {V_{s1}}^t \nabla_{\mathbf{X}(1)} D$, $g(s) = \partial_2 G(s)^t \theta(s)$ and $\xi(t) = \int_t^1 {R_{0s}}^t g(s)ds$, so that the gradient (12) can be re-written as:

$$\begin{cases} \nabla_{\mathbf{S}_0} E = \displaystyle\int_0^1 {R_{0s}}^t g(s)ds + \nabla_{\mathbf{S}_0} L = \xi(0) + \nabla_{\mathbf{S}_0} L \\ \nabla_{\mathbf{X}_0} E = \theta(0) \end{cases}.$$

Now, we need to make explicit the computation of the auxiliary variables $\theta(t)$ and $\xi(t)$. By definition of V_{t1}, we have $V_{11} = \mathrm{Id}$ and $dV_{t1}/dt = V_{t1}\partial_1 G(t)$, which implies that $\theta(0) = \nabla_{X(1)} D$ and $\dot{\theta}(t) = -\partial_1 G(t)^t \theta(t)$, namely (7).

For $\xi(t)$, we notice that $R_{ts} = \mathrm{Id} - \int_t^s \frac{dR_{us}}{du} du = \mathrm{Id} + \int_t^s R_{us} d_{\mathbf{S}(u)} F(u) du$. Therefore, using Fubini's theorem, we get:

$$\begin{aligned} \xi(t) &= \int_t^1 {R_{ts}}^t g(s)ds \\ &= \int_t^1 g(s) + d_{\mathbf{S}(s)}F^t \int_s^1 {R_{su}}^t g(u)duds \\ &= \int_t^1 g(s) + d_{\mathbf{S}(s)}F^t \xi(s)ds. \end{aligned}$$

This last equation is nothing but the integral form of the ODE given in (8).

3 Gradient in coordinates

In this section, we write the gradient in coordinates. Expanding the variables $\mathbf{S}^i(t) = \{\mathbf{c}_{0,k}(t), \alpha^i_{0,k}(t)\}$, $\mathbf{X}^i(t) = \{X^i_k(t)\}$, $\theta^i(t) = \{\theta^i_k(t)\}$ and $\xi^i(t) = \{\xi^{c,i}_k(t), \xi^{\alpha,i}_k(t)\}$, we have

$$\nabla_{\mathbf{c}_{0,k}} E = \sum_{i=1}^{N_{\text{subj}}} \xi^{c,i}_k(0) + \nabla_{\mathbf{c}_{0,k}} L(\mathbf{S}^i_0)$$
$$\nabla_{\alpha^i_{0,k}} E = \sum_{i=1}^{N_{\text{subj}}} \xi^{\alpha,i}_k(0) + \nabla_{\alpha^i_k} L(\mathbf{S}^i_0)$$

where the gradient of the regularity term is given as (from now on, we omit the subject's index i for clarity purposes):

$$\nabla_{\alpha_k} L = 2 \sum_{p=1}^{N_c} K^{\mathrm{g}}(\mathbf{c}_k, \mathbf{c}_p)\alpha_p$$
$$\nabla_{\mathbf{c}_k} L = 2 \sum_{p=1}^{N_c} {\alpha_p}^t \alpha_k \nabla_1 K^{\mathrm{g}}(\mathbf{c}_k, \mathbf{c}_p)$$

and the other terms are computed as follows.

3.1 Computation of $\dot{\theta}(t)$

The term $\partial_1 G(\mathbf{X}(t), \mathbf{S}(t))$ is a block-matrix of size $3N_c \times 3N_x$ whose $(k,p)th$ 3×3 block is given as:

$$d_{X_k} G(\mathbf{X}(t), \mathbf{S}(t))_p = \sum_{j=1}^{N_c} \alpha_j(t) \nabla_1 K(X_p(t), c_j(t))^t \delta(p-k)$$

so that the vector $\theta(t)$ is updated according to:

$$-\dot{\theta}_k(t) = \sum_{p=1}^{N_c} \alpha_p(t)^t \theta_k(t) \nabla_1 K(X_k(t), c_p(t)) \tag{13}$$

3.2 Computation of $\dot{\xi}(t) = (\dot{\xi}^c(t), \dot{\xi}^\alpha(t))$

The terms $\partial_{\mathbf{c}^{\mathrm{g}}} G(\mathbf{X}(t), \mathbf{S}(t))$ and $\partial_{\boldsymbol{\alpha}} G(\mathbf{X}(t), \mathbf{S}(t))$ are both matrices of size $3N_x \times 3N_c$, whose (k,p) block is given respectively by:

$$d_{c_k} G_p = \alpha_k \left(\nabla_1 K(c_k, X_p)\right)^t$$
$$d_{\alpha_k} G_p = K(c_k, X_i)\mathrm{I}_3$$

The differential of the function $F(\mathbf{S}) = \begin{pmatrix} F^c(\mathbf{c}, \boldsymbol{\alpha}) \\ F^\alpha(\mathbf{c}, \boldsymbol{\alpha}) \end{pmatrix}$ can be decomposed into 4 blocks as follows:

$$d_{\mathbf{S}(t)}F = \begin{pmatrix} \partial_{\mathbf{c}} F^c & \partial_{\boldsymbol{\alpha}} F^c \\ \partial_{\mathbf{c}} F^\alpha & \partial_{\boldsymbol{\alpha}} F^\alpha \end{pmatrix} \tag{14}$$

Therefore, the update rules for the auxiliary variables $\xi^c(t)$ and $\xi^\alpha(t)$ are given as:

$$\begin{cases} -\dot{\xi}_k^c(t) = \sum_{p=1}^{N_x} \alpha_k(t)^t \theta_p(t) \nabla_1 K(c_k(t), X_p(t)) + (\partial_{\mathbf{c}} F^c)^t \xi^c(t)_k + (\partial_{\mathbf{c}} F^\alpha)^t \xi^\alpha(t)_k \\ -\dot{\xi}_k^\alpha(t) = \sum_{p=1}^{N_x} K(c_k(t), X_p(t)) \theta_p(t) + (\partial_\alpha F^c)^t \xi^c(t)_k + (\partial_\alpha F^\alpha)^t \xi^\alpha(t)_k \end{cases}$$

with

$$\begin{aligned} (\partial_c F^c)^t \xi^c(t)_k = &\quad \sum_{p=1}^{N_c} \Big(\alpha_p(t)^t \xi_k^c(t) + \alpha_k(t)^t \xi_p^c(t) \Big) \nabla_1 K(c_k(t), c_p(t)) \\ (\partial_c F^\alpha)^t \xi^\alpha(t)_k = &\quad \sum_{p=1}^{N_c} \alpha_k(t)^t \alpha_p(t) \nabla_{1,1} K(c_k(t), c_p(t))^t \Big(\xi_p^\alpha(t) - \xi_k^\alpha(t) \Big) \\ (\partial_\alpha F^c)^t \xi^c(t)_k = &\quad \sum_{p=1}^{N_c} K(c_k(t), c_p(t)) \xi_j^c(t) \\ (\partial_\alpha F^\alpha)^t \xi^\alpha(t)_k = &\quad \sum_{p=1}^{N_c} \nabla_1 K(c_k(t), c_p(t))^t \Big(\xi_p^\alpha(t) - \xi_k^\alpha(t) \Big) \alpha_p(t) \end{aligned}$$

In these equations, we supposed the kernel symmetric: $K^{\mathrm{g}}(x, y) = K^{\mathrm{g}}(y, x)$. If the kernel is a scalar isotropic kernel of the form $K^{\mathrm{g}} = f(\|x - y\|^2)\mathrm{I}$, then we have:

$$\begin{aligned} &\nabla_1 K^{\mathrm{g}}(x, y) = 2f'(\|x - y\|^2)(x - y) \\ &\nabla_{1,1} K^{\mathrm{g}}(x, y) = 4f''(\|x - y\|^2)(x - y)(x - y)^t + 2f'(\|x - y\|^2)\mathrm{I} \end{aligned}$$

The ODEs are integrated by using a Euler scheme with prediction/correction scheme. This has the same accuracy as a Runge Kutta scheme of order 2.

Given the current value of the position of the vertices of the template shape $\mathbf{S}_0^i$, the position of the control points $\mathbf{c}_0$ and the momentum vectors α_0^i, one generates the subject-specific motions of the couples (control point, momentum vector) $\mathbf{S}^i(t)$ and the deformations of the template shape $\mathbf{X}^i(t)$ by integrating the ODEs (2) forward in time. Given these trajectories, one computes the gradient of the data term $\nabla_{\mathbf{X}^i(1)} D$ that indicates in which direction to move the vertices of the deformed template to decrease the most the residual data term. Then, one propagates this information to the control points and momenta by computing $\theta^i(1)$ and $\xi^i(1)$ and flowing these variables from time $t = 1$ back to time $t = 0$. Their values at time $t = 0$ is used to update the initial position of the control points $\mathbf{c}_0$, the initial momentum vectors $\boldsymbol{\alpha}_0^i$ and the positions of the vertices of the template shape $\mathbf{X}_0$.

The computation of the gradient uses back and forth integration to exchange information from the template space, where the parameters to be updated live, to the subjects' space, where the data terms live. The gradient is entirely driven by the gradient of the data term $\nabla_{\mathbf{X}^i(1)} D(\mathbf{X}^i(1), \mathbf{X}_i)$. The differentiation of the residual current norm is performed as explained in [2]. Would we have point correspondence across the shapes and would we choose the sum of squared differences for D ($D(\mathbf{X}^i(1), \mathbf{X}_i) = \sum_{k=1}^{N_x} (\mathbf{X}^i(1)_k - \mathbf{X}_k^i)^2$), then we would have simply:

$$\nabla_{\mathbf{X}_k^i(1)} D(\mathbf{X}_k^i(1), \mathbf{X}_i) = 2(\mathbf{X}_k^i(1) - \mathbf{X}_k^i).$$

References

[1] Durrleman, S., Prastawa, M., Korenberg, J.R., Joshi, S., Trouvé, A., Gerig, G.: In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI'12). LNCS, Springer (September 2012), to appear

[2] Vaillant, M., Glaunès, J.: Surface matching via currents. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 381–392 (2005)