

AN EFFICIENT PARALLEL ALGORITHM FOR HIERARCHICAL GEODESIC MODELS IN DIFFEOMORPHISMS

Nikhil Singh, Jacob Hinkle, Sarang Joshi, P. Thomas Fletcher*

Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, Utah

ABSTRACT

We present a novel algorithm for computing hierarchical geodesic models (HGMs) for diffeomorphic longitudinal shape analysis. The proposed algorithm exploits the inherent parallelism arising out of the independence in the contributions of individual geodesics to the group geodesic. The previous serial implementation severely limits the use of HGMs to very small population sizes due to computation time and massive memory requirements. The conventional method makes it impossible to estimate the parameters of HGMs on large datasets due to limited memory available on-board current GPU computing devices. The proposed parallel algorithm easily scales to solve HGMs on a large collection of 3D images of several individuals. We demonstrate its effectiveness on longitudinal datasets of synthetically generated shapes and 3D magnetic resonance brain images (MRI).

Index Terms— LDDMM, HGM, Vector Momentum, Diffeomorphisms, Longitudinal Analysis

1. INTRODUCTION

A longitudinal study of neuroanatomical aging, development, or disease progression necessitates modeling anatomical changes over time. A convenient representation of anatomical variability is via maps of diffeomorphisms, which are topology-preserving, smooth, and invertible transformations of a template image [1, 2]. Previous work proposes hierarchical geodesic models (HGMs) as generalizations of classical hierarchical models to diffeomorphisms [3]. Such models are effective at describing the longitudinal changes in anatomical shape within a population. HGMs utilize the metric in the space of diffeomorphisms to define a group geodesic given a population of geodesics. They are applicable to commonly occurring unbalanced designs in medical imaging data, where measurements are staggered, i.e., not every individual is measured at the same time points. Although mathematically elegant, estimating the parameters of these models involves massive computation. Naive implementations of the algorithm presented in [3] put severe limitations

on its applicability for population studies. In this paper, we present an efficient parallel algorithm for estimating the initial conditions of the group geodesic under the HGM.

2. METHODOLOGY

2.1. Preliminaries

Let Ω be the coordinate space of the image, I . A diffeomorphism, $\phi(t)$, is constructed by the integration of an ordinary differential equations (ODE) on Ω defined via a smooth, time-indexed velocity field, $v(t)$. The deformation of an image I by ϕ is defined as the action of the diffeomorphism, given by $\phi \cdot I = I \circ \phi^{-1}$. The choice of a self-adjoint differential operator, L , determines the right-invariant Riemannian structure on the collection of velocity fields with the norm defined as, $\|v\|^2 = \int_{\Omega} (Lv(x), v(x)) dx$.

The tangent space at identity, $V = T_{\text{Id}}\text{Diff}(\Omega)$, consists of all vector fields with finite norm. Its dual space, $V^* = T_{\text{Id}}^*\text{Diff}(\Omega)$, consists of vector-valued distributions over Ω . The velocity, $v \in V$, maps to its dual deformation momenta, $m \in V^*$, via the operator L such that $m = Lv$ and $v = Km$. The operator $K : V^* \rightarrow V$ denotes the inverse of L . Note that constraining ϕ to be a geodesic with initial momentum, $m(0)$, implies that ϕ , m , and I , all evolve in a way entirely determined by the metric L , and that the deformation is determined entirely by the initial deformation momenta, $m(0)$. The initial velocity, $v(0) \in V$, or equivalently, the initial momentum, $m(0) \in V^*$, uniquely parameterize the the geodesic path $\phi(t)$ [2, 4, 5].

2.2. Hierarchical Geodesic Model (HGM)

We are given a population of N individuals with M_i measurements for the i th individual. There can be a variable number of measurements for each individual. Denote H_{ij} as the j th measured image of the i th individual at time t_{ij} .

Figure 1 shows a schematic of the HGM. We model geodesic trend for an individual with a diffeomorphism, $\xi_i(t)$ (brown). The initial image, or intercept, $J_i(0)$, and the initial momenta, or slope, $n_i(0)$, fully parameterize the trajectory for the i th individual. At the group level, we model the group geodesic trend with the diffeomorphism, $\psi(t)$, (red) starting

*This research is supported by NIH grants U01NS082086, 5R01EB007688, U01 AG024904, R01 MH084795 and P41 RR023953, and NSF Grant 1054057.

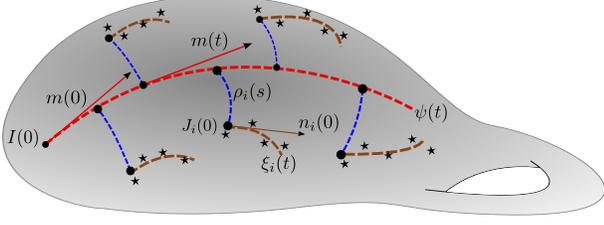


Fig. 1. Hierarchical geodesic modeling in diffeomorphisms.

at identity, parameterized by initial momenta, $m(0)$. Let ϕ_i denote the diffeomorphism that matches individual baseline $J_i(0)$ from identity and ρ_i denote the residual geodesic between $\psi(t_i)$ and ϕ_i : $\rho_i = \phi_i \circ \psi^{-1}(t_i)$. The initial momenta $p_i(0)$ parameterizes the residual ρ_i .

The two stages for parameter estimation of this model are given as:

Individual level: Given M_i observed images H_{ij} at time points t_{ij} for an individual such that $j = 1, \dots, M_i$, the geodesic that passes closest, in the least squares sense, to the data minimizes the energy functional

$$\mathcal{I}(J_i(0), n_i(0)) = \frac{1}{2} \|n_i(0)\|_K^2 + \frac{1}{2\sigma_i^2} \sum_{j=1}^{M_i} \|J_i(t_{ij}) - H_{ij}\|_{L^2}^2,$$

where $J_i(0)$ and $n_i(0)$ are the initial ‘‘intercept’’ and ‘‘slope’’ to be estimated that completely parameterize the geodesic for the i^{th} individual. Here, $J_i(t) = \xi_i(t) \cdot J_i(0)$, and $\|\cdot\|_K$ is the norm defined by the kernel, K , in the dual space of momenta, as per the metric induced by Sobolev operator, L , on velocity fields. Geodesic regression [6, 7] solves the initial conditions of each geodesic to provide estimates of $J_i(0)$ and $n_i(0)$ for $i = 1, \dots, N$, that minimize each of the N energy functionals. In the discussion that follows, for clarity and ease of notation, we will use $J_i = J_i(0)$ to denote the initial ‘‘intercept’’ and $n_i = n_i(0)$ to denote initial ‘‘slope’’ for an individual.

Group level: At the group level (Figure 1), the idea is to estimate the average geodesic, $\psi(t)$, that is a representative of the population of geodesic trends denoted by the initial intercept-slope pair, (J_i, n_i) , for N individuals, $i = 1, \dots, N$. The required estimate for $\psi(t)$ must span the entire range of time along which the measurements are made for the population and must minimize residual diffeomorphisms ρ_i from $\psi(t)$.

The energy functional for group and residual geodesics based on data log-likelihood takes the form

$$\begin{aligned} \mathcal{G}(\psi, \rho_i, m(0), p_i(0), I(0)) &= \frac{1}{2} \|m(0)\|_K^2 \\ &+ \frac{1}{2\sigma_I^2} \sum_{i=1}^N (\|p(0)_i\|_K^2 + \|\rho_i \cdot \psi(t_i) \cdot I(0) - J_i\|_{L^2}^2) \\ &+ \frac{1}{2\sigma_S^2} \sum_{i=1}^N \|\rho_i \cdot \psi(t_i) \cdot m(0) - n_i\|_K^2. \end{aligned} \quad (1)$$

Here, σ_I^2 and σ_S^2 represent the variances corresponding to the likelihood for the intercept and slope terms respectively. Notice, it is a combination of group actions of diffeomorphism on both the initial image, $I(0)$, and the initial momenta, $m(0)$, of the group geodesic. This group action on momenta also coincides with the co-adjoint transport in the group of diffeomorphisms. This optimization problem corresponds to jointly estimating the group geodesic flow, ψ , and residual geodesic flows, ρ_i , and the group baseline template, $I(0)$.

Evaluating gradients of \mathcal{G} : The gradients of \mathcal{G} are evaluated using an optimal control based approach. The dynamics of this system are controlled by adding time-dependent constraints to enforce all paths to be geodesics. In particular, it results in $N + 1$ geodesics: 1 for the group, ψ , and N for the residuals, ρ_i (for $i = 1, \dots, N$).

The variation of the energy functional, \mathcal{G} , with respect to all time dependent variables results in ODEs in the form of dependent adjoint equations with boundary conditions and added jump conditions. We refer the reader to the original article by Singh et. al [3] for details about these gradients and jump conditions. The energy functional at the group level is jointly minimized such that the group estimates, $I(0), m(0)$, and all the N residual estimates, $\rho_i(1), p_i(0)$, are updated at each iteration of gradient descent. The resulting adjoint system for the group geodesic takes the form

$$\left. \begin{aligned} -\dot{\hat{m}} + \text{ad}_v \hat{m} + \hat{v} &= -0 \\ \dot{\hat{I}} - \nabla \cdot (\hat{I}v) &= -0 \\ \text{ad}_m^* m + \hat{I} \nabla I - L\hat{v} &= 0 \end{aligned} \right\}, \quad (2)$$

with boundary conditions: $\hat{I}(1) = 0$, and $\hat{m}(1) = 0$, and with added jumps at measurements, t_i , such that,

$$\left. \begin{aligned} \hat{I}(t^{i+}) - \hat{I}(t^{i-}) &= \frac{1}{\sigma_I^2} |D\rho_i| (I(t_i) \circ \rho_i^{-1} - J_i) \circ \rho_i \\ \hat{m}(t^{i+}) - \hat{m}(t^{i-}) &= \frac{1}{\sigma_S^2} \text{Ad}_{\rho_i^{-1}} (K \star (\text{Ad}_{\rho_i^{-1}}^* m(t_i) - n_i)) \end{aligned} \right\}. \quad (3)$$

Finally, the gradients for update of the initial group momentum is

$$\delta_{m(0)} \tilde{\mathcal{G}} = K \star m(0) - \hat{m}(0). \quad (4)$$

The variation of $\tilde{\mathcal{G}}$ with respect to the group initial image, $\delta_{I(0)} \tilde{\mathcal{G}}$, can be directly computed from the energy functional, $\tilde{\mathcal{G}}$. Since, $\rho_i \cdot \psi(t_i) \cdot I(0) = I(0) \circ \psi^{-1}(t_i) \circ \rho_i^{-1}(1) = I(0) \circ \phi^{-1}$, a change of variable for ϕ_i followed by taking the derivative with respect to $I(0)$ results in the closed form solution for optimum initial image, $I(0)$, for the group geodesic as

$$I(0) = \frac{\sum_{i=1}^N J^i \circ \phi_i |D\phi_i|}{\sum_{i=1}^N |D\phi_i|}. \quad (5)$$

2.3. Parallel Algorithm for the HGM

The estimation of the initial conditions of the group geodesic, as presented above, is computationally intensive and also has massive memory requirements. A naive serial computation of gradient updates results in a very slow algorithm. Additionally, a single GPU-based implementations easily hit the limits of the available memory in the state-of-the-art computing architectures even for a small population study. In this section, we discuss a fast and parallel GPU-based algorithm that easily scales to big longitudinal studies.

Equation (4) suggests that the gradient depends upon the adjoint variable, $\hat{m}(0)$, corresponding to momenta, m at $t = 0$. At a given iteration of gradient descent, $\hat{m}(0)$ must be computed by the backward integration of the adjoint system (2). To realize the parallelism in the computation, we must note that in each iteration of the optimization algorithm: a) The backward integration of the adjoint system (2) is conditional on the estimates of geodesic paths, ψ and ρ_i 's. b) The jumps added to $\hat{m}(t)$ as per (3) during this integration are independent to each other, and c) Integration is a linear operator. In fact, the objective function in (1) is separable for N individuals. Thus, the jumps are also linearly separable.

The above imply that the $\hat{m}(0)$ is a result of accumulating the integrated jumps that are independent and linearly separable, given the current estimates of the group and residual geodesics. The backward integration thus lends itself to a division into parallel computations of the jumps independently, followed by their independent backward integrations along the group geodesic. This computation is divided over L subsets of the full population. Each of the L processes compute the adjoint variable for $\frac{N}{L}$ individuals and results in its own version of $\hat{m}(0)$, denoted as the $\hat{m}_l(0)$. This results in $\hat{m}_l(0)$ (for $l = 1 \dots L$) that represent effects of the pull by only the respective subset of individuals. Due to linearity of integration, the $\hat{m}(0)$ is the sum of the adjoints computed over the L subsets such that: $\hat{m}(0) = \sum_l^L \hat{m}_l(0)$.

Note that the image update step in (5) is trivially parallel since it does not involve any backward integration and only relies on current estimates of the geodesics. Both the numerator and the denominator in (5) can be parallelly computed along with the L sub-processes. If we denote $A = \sum_{i=1}^N J^i \circ \phi_i |D\phi_i|$ and $B = \sum_{i=1}^N |D\phi_i|$ such that A_l and B_l are the accumulated sums only on the l th subset, the accumulated, $A = \sum_l^L A_l$ and $B = \sum_l^L B_l$.

Pseudocode for this parallel computation is detailed in Algorithm 1. Step 3 in this algorithm computes the geodesics all the way to baseline points of the individual subjects along their respective residual geodesics. Step 4 performs the backward integration of adjoint variables starting from these end points to the initial baseline time of the group geodesic. Both of these steps work parallelly as L processes on L subsets of the population.

Algorithm 1: HGMPParallel

```

input : Individual image and momenta pairs  $(J_i, n_i)$ .
output: Group image and momenta pair  $(I(0), m(0))$ .
1 begin
2   while not converged do
3     // spawn  $L$  processes using MPI
4      $(\psi(t_l), \rho_l) \leftarrow \text{FwdEvolveGeodesics}(I(0), m(0))$ 
5      $(\hat{m}_l(0), A_l, B_l) \leftarrow \text{BwdIntegrateAdj}(J_l, n_l, \psi(t_l), \rho_l)$ 
6
7     // E.g. MPIReduce  $\hat{m}_l(0), A_l, B_l$ 
8      $\hat{m}(0) \leftarrow \text{SumAcrossProcesses}(\hat{m}_l(0))$ 
9      $A \leftarrow \text{SumAcrossProcesses}(A_l)$ 
10     $B \leftarrow \text{SumAcrossProcesses}(B_l)$ 
11    // Updates as per Eqs. (4), (5)
12     $m(0) \leftarrow \text{UpdateMomenta}(m(0), \hat{m}(0))$ 
13     $I(0) \leftarrow \text{UpdateImage}(A, B)$ 
14  end
15  return  $(I(0), m(0))$ 
16 end

```

3. RESULTS

We present our evaluation and compare the performance of our proposed parallel algorithm for HGM with its corresponding serial implementation. We report the group estimate results both a population of 2D synthetic shapes (size 128x128) and a population of 3D MRI (size 128x128x128).

We generated the synthetic data using the forward model similar to the procedure presented in [3]. We first generated a ground truth group geodesic in diffeomorphisms by solving the image matching problem. We generated random perturbations from the group trend to generate a population of individual geodesics. Please see [3] for details. We generated 100 such randomly perturbed trends from the group trend. For the 3D MRI experiments, longitudinal 3D-MRI sequences for 69 cognitively normal (non-demented) individuals were downloaded from the OASIS database. The individuals span a total range of 30 years from the age of 60 to 90 years. Each individual is scanned for about 2-4 timepoints with at least one year difference between their timepoints. At the individual level of HGM, 69 geodesic regressions are performed independently on the time-series of scans for 69 individuals. At the group level, the initial conditions of the average geodesic are estimated based on the estimated initial conditions of 69 individuals from individual level.

For fair comparisons we use identical integration schemes and identical number of iterations of constant stepsize gradient descent based optimization for all runs. For all our implementations, we use the message passing interface (MPI) library of python (mpi4py). The GPU computing devices used in these tests consists of NVIDIA Tesla 1070's. The on-device GPU memory for each Tesla 1070 is 4GB. A multi-node clus-

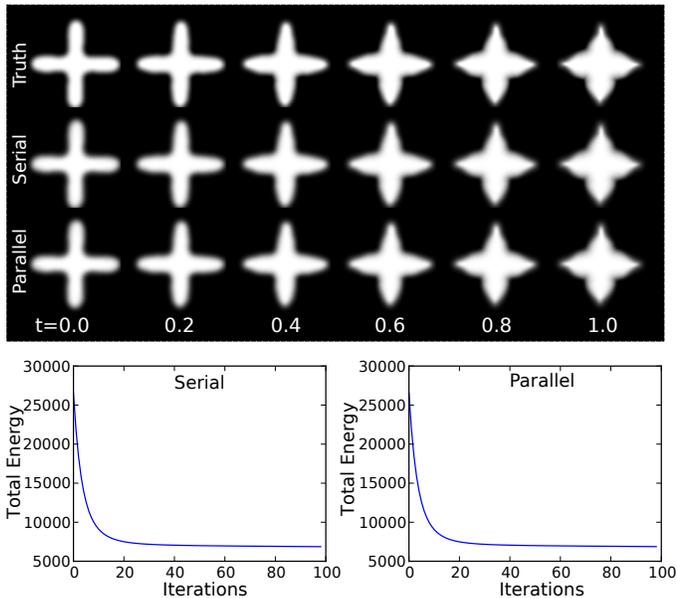


Fig. 2. Comparison of serial and 100 threads parallel HGM.

ter consisting of computing servers with these Teslas is used for all our tests. In our experiments, the parallel algorithm uses 100 parallel processes with 2D synthetic shapes while 69 parallel processes with 3D images. This means that we assign one individual per process in the group geodesic estimation stage of HGM.

Table 1. Runtime comparison.

Group	Serial GPU	Parallel GPU	Serial CPU	Parallel CPU
2D Toy	288s	19s	2440s	73s
3D Brain	—	11m 48s	66h 30m	132m 12s

We notice that the parallel HGM outpaces the serial HGM for both the CPU as well as the GPU based implementations (Table 1) on all experiments. This difference in performance gets especially magnified for our experiments with 3D MR brain images. Our parallel algorithm using the GPUs impressively finishes within 12 minutes. As we discussed in Sec. 2, the Tesla 1070 is not able to meet the massive memory requirement for the serial HGM (blank in Table 1). This demonstrates a severe practical limitation of the previous algorithm.

We also notice that both the implementations result in identical estimates of group geodesics (Figure 2). Along with the group estimates, the energies of the objective function during gradient descent iterations also match closely. Figure 3 depicts the initial conditions, i.e., the initial image and momenta directions for the group geodesic estimated at 60 years of age. The last column shows the smoothly deformed grid corresponding to 30 years of deformation (from age 60 to 90),

which spans the whole time range for the population. The deformation momenta at the age of 60 depict realistic directions of atrophy in the average representation of longitudinal changes in the population.

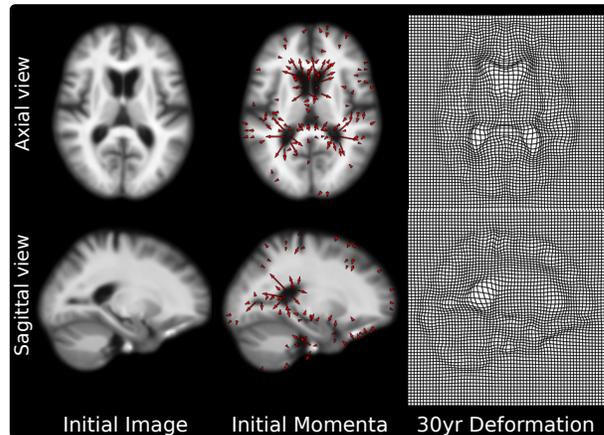


Fig. 3. Parallel HGM on 3D MRI for 69 individuals.

The source code for our CPU and GPU based implementations of the parallel HGM can be found at: <https://bitbucket.org/scicompanat/vectormomentum>

4. REFERENCES

- [1] M.F. Beg, M.I. Miller, A. Trouvé, and L. Younes, “Computing large deformation metric mappings via geodesic flows of diffeomorphisms,” *IJCV*, v.61, p.139–157, 2005.
- [2] L. Younes, F. Arrate, and M. I. Miller, “Evolution equations in computational anatomy,” *NeuroImage*, vol. 45, no. 1 Suppl, pp. S40–S50, 2009.
- [3] N. Singh, et al.: “A hierarchical geodesic model for diffeomorphic longitudinal shape analysis,” in *IPMI 2013* vol. 7917, pp. 560–571, 2013.
- [4] V. I. Arnol’d, “Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l’hydrodynamique des fluides parfaits,” *Ann. Inst. Fourier*, vol. 16, pp. 319–361, 1966.
- [5] M. I. Miller, A. Trouvé, and L. Younes, “Geodesic shooting for computational anatomy,” *Journal of Mathematical Imaging and Vision*, vol. 24, pp. 209–228, 2006.
- [6] M. Niethammer, Y. Huang, and F-X. Vialard, “Geodesic regression for image time-series,” in *MICCAI 2011*, vol. 6892 pp. 655–662. Springer, 2011.
- [7] N. Singh, et al.: “A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction,” in *ISBI 2013* pp. 1219–1222, 2013.