

ShapeLME: Longitudinal statistical analysis tutorial

Introduction

Deformetrica gave us a way to estimate smooth subject specific shape models. Each subject model lets us re-evaluate shapes at known time-points. Given a longitudinal data set, a more natural question is to understand the behavior of an entire group (i.e., healthy group, groups at risk of HD) and also to have the ability to distinguish between two different groups. The optimization problem for diffeomorphic longitudinal modeling is challenging and existing approaches haven't yet provided a consistent way to compare two different trajectories. But the estimation of diffeomorphic subject specific models remove extraneous and unwanted variability in data.

To do statistics, this package **ShapeLME** computes longitudinal models to summarize the behavior of a specific group. To do longitudinal analysis, we first make sure that the output of Deformetrica processing gives shapes in longitudinal correspondence. With these correspondences, the setup is ripe to estimate linear mixed-effects for shape as summaries [2, 1].

Diffeomorphic trajectories are good at capturing smooth anatomical shape changes, while the particle optimization framework excels at finding compact statistical shape representations with increased statistical power. The idea here is to leverage the strengths of both approaches, to provide an integrated solution, characterized by improved statistical performance in the analysis of both scalar and shape trajectory data derived from noisy segmentations [3].

Installation details for ShapeLME are outlined in the README file. Once installed, we are ready to do some statistical analysis.

From shape meshes to point correspondences

Each observation of each subject are estimated and stored in vtk format with information about mesh points, polygonal connectivity surface normals, and every shape observation is evaluated in correspondence with every other shape. We extract just the point correspondence positions alone and store it in another directory in lpts format.

From vtk to lpts

- **data-path:** path to **regressed observed shape meshes** - our naming: `reg_at_obs_time_pts`
- **output-data-path:** path to output point correspondence files - our naming: `shape_lpts_for_stats`

Run the script:

- **python write_vtk_to_lpts_script.py**

Estimating longitudinal trajectories

The linear mixed-effects model, originally developed by Laird and Ware [2] has the ability to incorporate multiple covariates such as risk of onset group membership (CTRL, LOW, MED, HIGH), sex, among other information and estimate an associated model. This model was extended to study longitudinal shape by Datar, et al. [1].

For the tutorial, we control for sex by considering only female subjects and estimate models for CTRL and HIGH risk groups only. Once the script is run, output directories for each structure (left/right caudate/putamen) is created and fixed and random effects are respectively estimated and stored as a text file.

Note: Deformetrica estimates diffeomorphic models for entire shape complexes (such as the striatal complex of left/right caudate/putamen), whereas here we separate out each structure, center them, align each structure to a template structure, and estimate longitudinal models per structure. This doesn't have to be the case. Here too, one could consider the whole shape complex, with an appropriate alignment, and estimate models that way as long as there is longitudinal correspondence across subjects and time-points. This will be incorporated in future iterations of this software, but for now, we estimate models for each structure separately.

Estimating mixed-effects

- **input-data-path:** the lpts directory path obtained as output above - `shape_lpts_for_stats`
- **output-data-path:** path to fixed and random effects files for each structure - `mixed_effects_results`

Run the script:

- **python estimate_mixed_effects_script.py**

Script details When we run the script, 5 files get created - 3 of which are created to be used as input to the estimation - `lpts_file.structure.txt`, `design_file.txt`, `tp_file.txt`; and 2 files created post estimation - fixed and random effects files. The design file can be modified depending on the data and model choice.

- In our tutorial example, we estimate linear models - i.e., intercept and slope parameters, and the only other covariate is risk group with two options (i.e., CTRL (0), HIGH (1)). Therefore, the first two columns of `design_file.txt` represents a baseline of intercept and baseline for slope parameters (say CTRL is the baseline). The next two columns represent *group * intercept*, *group * slope*, where *group* = 0, 1. In essence, columns 3 and 4 represent perturbation from CTRL in intercept-slope parameters respectively to get to HIGH, and that covers all the data.

- If suppose we had another covariate say sex (both males and females in data), we would need add two more columns to account for intercept slope perturbations to go from a set baseline (say female) to the alternative (male).
- If a covariate had more than 2 alternatives (say group = *CTRL, LOW, MED, HIGH* has 4 options), we would need 4 columns - one baseline intercept, and 3 different columns to represent perturbations from a set intercept baseline. And we need a similar setup for slope.

So depending on model choice and design file setup, the estimated fixed-effects file will have same number of columns as the design file representing longitudinal estimates of the associated covariate parameter.

- The number of rows of fixed-effects file is $3 \times \text{number_of_mesh_points_per_shape}$, 3 being x,y,z coordinates.
- The choice of a simple linear model and enforcing randomness only in intercept and slope from group level to subject level is estimated in the 2 columns of the random-effects file.
- The number of rows here is $3 \times \text{number_of_mesh_points_per_shape} \times \text{number_of_subjects}$.

Visualizing shapes and trajectories

The baseline shapes along with the respective group evolutions of each category can be visualized in Paraview. To do this, we convert the respective point and velocity files back to vtk format keeping the original neighborhood and connectivity information of the input mesh data. Once thats done, we then load the respective vtk files into Paraview and compare the visualizations there. The output of this script is a sequence of shape files (in vtk format) evaluated at different points along the estimated lonitudinal trajectory that are parameterized by fixed-effects. The time axis is taken to be between min and max of the given data range. The baseline is considered to be the minimum age of the data span.

- **input-path:** the mixed effects results directory
- **output-data-path:** output vtk shape sequences path for each structure - our naming: `output_shape_seq.vtk`

Run the script:

- **python create_shape_sequences_script.py**

Here's a visualization of two different views of the left caudate

Create your own visualizations in Paraview

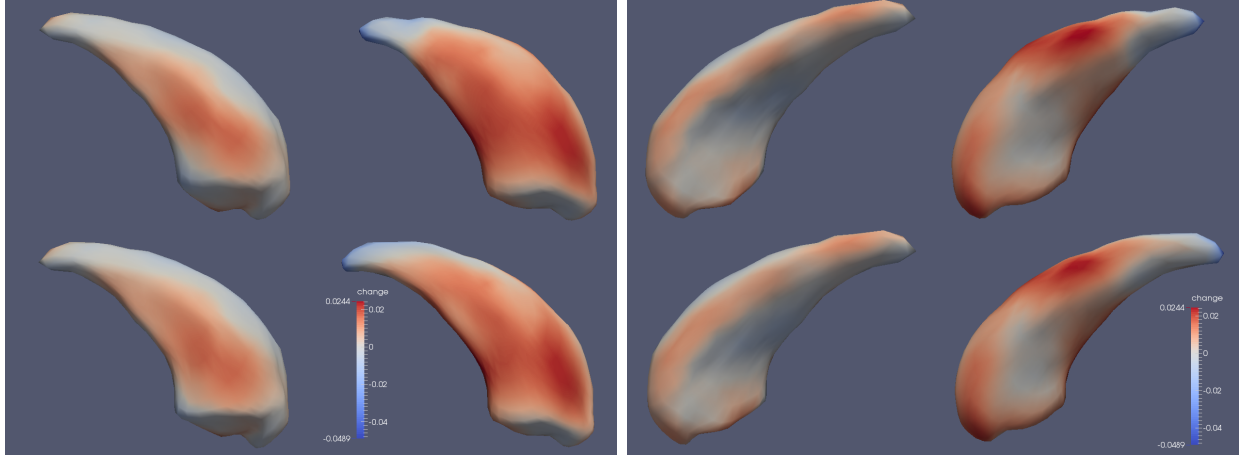


Figure 1: Two views (left and right) of left caudate fixed-effects estimates: CTRL (top), HIGH (bottom). Colormap indicates local expansion (red) and contraction (blue)

- To create a similar visualizaion as in 1, load the first vtk files in the estimated output shape sequence for each group. This will be denoted by ctrl_0.vtk, hd_0.vtk.
- Load up two windows and link their cameras so as to see both groups in the same perspective. Run two filters - generate surface normals and calculator on both files.
- In generate surface normals, compute cell normals without splitting and with consistency on. This is an interpolation scheme where normals are computed everywhere and smooth structure is obtained. In calculator, compute the dot product of velocity and surface normals. A negative dot product indicates local contraction, and a positive dot product represents local expansion. Color this measure from blue to red represent contraction to expansion.
- Another visualization to create is to load all shapes in the sequence and see them as a video.

Hypothesis testing

We finally test the hypothesis for significant differences between CTRL and HIGH risk groups by a non-parametric permutation test using the Hotelling t^2 statistic. The input to this are the point correspondence files, and the output is a p-value. The tutorial does two different hypothesis tests

- Are there significant differences in shape between CTRL and HIGH groups at baseline, and
- Are there significant differences in shape change between CTRL and HIGH groups over time.

As explained before, the first two columns of the estimated fixed-effects parameters represent intercept and slope of a baseline category - CTRL in our case. The next two columns represent perturbations of intercept and slope from baseline to go to the alternative group - HIGH. The question then becomes are the parameters of columns 3 & 4 significant as they attest to the differences between two groups - at baseline (intercept) and over time (slope). The non-parametric permutation tests with the Hotelling t^2 statistic is explained in Section 3.4 in [1].

- **input-data-path:** the lpts directory path obtained as output above - shape_lpts_for_stats
- **output-data-path:** text file listing hypothesis results - we store it in mixed_effects_results

Run the script:

- **python group_hypothesis_testing_script.py**

References

- [1] M. Datar, P. Muralidharan, A. Kumar, S. Gouttard, J. Piven, G. Gerig, R.T. Whitaker, and P.T. Fletcher. Mixed-effects shape models for estimating longitudinal changes in anatomy. In *MICCAI STIA*, volume 7570, pages 76–87, 2012.
- [2] Nan Laird and James H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.
- [3] P. Muralidharan, J. Fishbaugh, H.J. Johnson, S. Durrleman, J.S. Paulsen, G. Gerig, and P.T. Fletcher. Diffeomorphic shape trajectories for improved longitudinal segmentation and statistics. In *Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2014.