

# registr: Registration for Exponential Family Functional Data

2018-01-18

## Summary

Functional data analysis is a set of tools for understanding patterns and variability in data where the basic unit of observation is a curve measured over time, space, or another domain. Classic functional data analysis assumes that each curve is continuous or comes from a Gaussian distribution. However, applications with exponential family functional data – curves that arise from any exponential family distribution, and have a smooth latent mean – are increasingly common.

Often in a functional dataset curves have similar underlying patterns but the main features of each curve, such as the minimum and maximum, have shifts such that the data appear misaligned. This misalignment can obscure patterns shared across curves and produce messy summary statistics. Registration methods reduce variability in functional data and clarify underlying patterns by aligning curves. Our method estimates a map, called a **warping function**, which transforms the domain from so that curves are aligned. The model for registration can be written

$$E[Y_i(h_i^{-1}(t_i^*)) | c_i, h_i^{-1}] = \mu_i(t)$$

$$g[\mu_i(t)] = \alpha(t) + \sum_{k=1}^K c_{ik} \psi_k(t).$$

For subject  $i$ , warping function  $h_i^{-1}$  maps the domain on which curves are misaligned,  $t_i^*$ , to aligned domain  $t$  such that  $h_i^{-1}(t_i^*) = t$ . Then  $Y_i(t_i^*)$  and  $Y_i(h_i^{-1}(t_i^*))$  are the unregistered and registered functional response curves, respectively. The  $\mu_i(t)$  are subject-specific means related to the population-level mean  $\alpha(t)$  and a linear combination of population-level basis functions  $\psi(t)$  and subject-specific scores  $c_i$  through a known link function  $g$ .

The **registr** package estimates warping functions and other parameters in this model via a two-step iterative algorithm which is detailed in Wrobel et al. (2018). The main function is **register\_fpca**, which registers functional data from a specified exponential family distribution. **register\_fpca** reads in a long-format functional dataset and outputs an object of class **registration**.

To enhance computational efficiency, key algorithm components are implemented in C++ using the R libraries **Rcpp** and **RcppArmadillo** (Eddelbuettel and François 2011, Eddelbuettel and Sanderson (2014)). Interactive visualizations are enabled with the **refund.shiny** package (Wrobel and Goldsmith 2015, Wrobel et al. (2016)).

## References

- Eddelbuettel, Dirk, and Romain François. 2011. “Rcpp: Seamless R and C++ Integration.” *Journal of Statistical Software* 40 (8): 1–18. doi:10.18637/jss.v040.i08.
- Eddelbuettel, Dirk, and Conrad Sanderson. 2014. “RcppArmadillo: Accelerating R with High-Performance C++ Linear Algebra.” *Computational Statistics and Data Analysis* 71 (March): 1054–63. <http://dx.doi.org/>

10.1016/j.csda.2013.02.005.

Wrobel, J, and J Goldsmith. 2015. *Refund.shiny: Interactive Plotting for Functional Data Analyses*.

Wrobel, J, S-Y Park, A-M Staicu, and J Goldsmith. 2016. “Interactive Graphics for Functional Data Analyses.” *Stat* 5: 88–101. doi:10.1002/sta4.109.

Wrobel, J, V Zipunnikov, J Schrack, and J Goldsmith. 2018. “Registration for Exponential Family Functional Data.” *Under Review*.