

# Hierarchical Principal Manifold Estimation

Robert Zielinski<sup>1</sup>, Kun Meng<sup>2</sup>, and Ani Eloyan<sup>1</sup>

<sup>1</sup>Department of Biostatistics, Brown University

<sup>2</sup>Division of Applied Mathematics, Brown University

August 11, 2023

## Abstract

## 1 Introduction

Neurodegenerative diseases such as Alzheimer’s disease (AD) and Parkinson’s disease are highly complex, chronic conditions that exhibit substantial heterogeneity between patients. To enable researchers to identify potential biomarkers of progression for these conditions, large multi-site longitudinal observational studies have been conducted that collect information from a wide range of participants over long durations across several modalities, including neuroimaging [Cite ADNI and PPMI studies]. Neuroimaging data in particular have been shown to be sensitive to developments in the early stages of disease progression, often before diagnosis occurs [Find citation from reading].

While observations from neuroimaging data are typically high-dimensional, it is commonly assumed that the data lie along a low-dimensional manifold [Relevant citations demonstrating this]. Under this assumption, manifold learning algorithms can be used to recover this underlying manifold and parameterize this high-dimensional input data in the learned low-dimensional space. Most commonly-used manifold learning algorithms, including Isomap (Tenenbaum, Silva, and Langford 2000), Locally Linear Embedding (Roweis and Saul 2000), Laplacian Eigenmaps (Belkin and Niyogi 2003), and Diffusion Maps (Coifman and Lafon 2006) construct a graph of similarities between observations, which is then used to determine the structure of the manifold. Meng and Eloyan 2021 presents an alternative principal manifold framework for manifold estimation which extends to higher dimensions the principal curve framework introduced in Trevor Hastie and Stuetzle 1989, in which principal curves are defined as one-dimensional curves that pass through the middle of a high-dimensional dataset.

In neuroimaging settings, manifold learning algorithms are most often used to estimate the underlying manifold at the population level, with vectorized image intensities being used as the input data [Citations]. Yue et al. 2016 used manifold learning to estimate a low-dimensional parameterization of the corpus callosum at the individual level. However, in many cases, the outcome of interest is framed as a comparison between different groups of individuals (e.g. treatment groups in a clinical trial). This goal is particularly relevant in the case of the complex neurodegenerative conditions mentioned above, where the high degree of heterogeneity in patient experiences have led to attempts to identify subgroups of patients based on a range of criteria such as symptom expression, onset, or rate of progression [Citations]. Fundamentally, if a low dimensional manifold is thought to effectively represent data related to the patient experience of a neurodegenerative condition, and the experience of that condition differs by group, then the low-dimensional representation of that experience should also differ by group.

A method to accurately and efficiently estimate the manifolds underlying data at the group level must take advantage of the grouped structure of observations found in the dataset to which it is to be applied. To develop an estimate of the manifold underlying the hippocampi of participants in the ADNI study, it is necessary to recognize the basic nesting structure of the data, in which brain images are nested within participants, which may belong to a latent disease subgroup for which patients tend to have more related symptom experiences or underlying physiological developments, which are in turn nested within study groups based on Alzheimer’s disease status, which are finally nested within the greater study population. The goal of this article is to develop a manifold learning approach that appropriately accounts for this hierarchical structure, estimates the underlying manifold at each level of the hierarchy, and allows for statistical testing of differences between those low-dimensional manifolds.

There have been several approaches taken to address hierarchical structure in manifold learning problems. Bhatia et al. 2012 introduced a modification of Laplacian eigenmaps to estimate manifolds from recursively subdivided regions of medical images while enforcing similar alignment of the manifolds estimated from related regions. Freiche et al. 2022 builds on this approach to incorporate multiple imaging modalities and automate the tuning of the algorithm’s hyperparameters.

Gao, Mishne, and Scheinost 2019 proposed a two-step diffusion maps method for hierarchical data with variation over several domains, where diffusion maps is applied first to reduce the images for each individual to  $d_1$  dimensions, resulting in individual-specific manifold estimates, then diffusion maps is applied again on the concatenated parameterizations to reduce over the individuals to  $d_2$  dimensions, thus estimating a group-level manifold. This approach allows the first step of dimension reduction to capture the most

important qualities of each image separately before reducing across subjects, but the two step procedure ensures that the manifold estimated at the individual level has no relation to the manifold estimated at the group level, posing problems for the interpretability of the embedding mappings and comparisons of manifolds between levels.

Guerrero et al. 2017 noted that the grouping behavior of observations recorded longitudinally from the same individual adversely impacted the creation of the neighborhood graph between data points when estimating a manifold for the entire population. The authors proposed a constraint specifying that, a given observation’s collection of nearest neighbors may include at most one point from each individual or other specified grouping. This encourages a more even representation of the entire space of observations in the neighborhood graph, and may be applied to any manifold learning algorithm that features a neighborhood graph to specify proximities between observations. However, it simply mitigates the effects of the nesting structure on a population-level manifold estimate rather than enabling a detailed comparison of manifold estimates for each group.

This article builds on the principal manifold estimation (PME) algorithm introduced in Meng and Eloyan 2021, which uses an iterative procedure to specify an embedding mapping from the low-dimensional manifold to a high-dimensional space that takes the form of a smoothing spline. Drawing from literature in hierarchical and additive models, we replace the smoothing spline in the PME algorithm with an additive smoothing spline model (Gelman and Hill 2007, T.J. Hastie and R.J. Tibshirani 1990, Gelman, Carlin, et al. 2014). While the specific model structure may change to meet the needs of the application, Brumback and Rice 1998 and Schulam and Saria 2015 provide examples of the general model framework used. This ultimately yields a nonlinear dimension reduction algorithm that takes advantage of a statistically principled model to address nesting structures present in the input data. As it is based on the PME framework, the algorithm outputs an explicit representation of the embedding function, making it possible to both generate new data in high-dimensional space using low-dimensional parameters and map new input data to the estimated manifold. Smoothing spline analysis of variance allows the statistical comparison of the estimated manifolds between groups or individuals (Wang 1998).

The following sections are organized as follows: Section 2 specifies an additive smoothing spline model for use in the analysis of neuroimaging data from the ADNI study, and provides a comprehensive description of the hierarchical PME algorithm developed in this paper. Section 3 applies this method to several simulated datasets using a variety of underlying manifold structures and tests performance on the MNIST dataset of handwritten digits. Section 4 demonstrates how the model specified in Section 2 may be used for the

estimation of a manifold underlying the surfaces of subcortical structures of participants of the ADNI dataset, while Section 5 provides a concluding discussion of the proposed algorithm and opportunities for further study.

## 2 Methods

### 2.1 Additive Smoothing Spline Model

We begin this section with a brief review of additive models. Consider a regression scenario where output variable  $\mathbf{y} = (y_1, \dots, y_n)^T$  are thought to be influenced by covariate vectors  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ . In this situation, the standard multiple linear regression model assumes that the dependence of  $y_i$  on  $\mathbf{x}_i$  takes the linear form

$$y_i = \alpha + x_{i1}\beta_1 + \dots + x_{ip}\beta_p + \epsilon_i,$$

where  $\epsilon$  has mean 0 and variance  $\sigma^2$ . Additive models are a direct extension of this model, where the additive structure of the model is retained, but the assumption of linearity is discarded. A simple additive model takes the form

$$y_i = \alpha + \sum_{j=1}^p f_j(X_{ij}) + \epsilon_i,$$

where again  $\epsilon_i$  has mean 0 and variance  $\sigma^2$ , but the  $f_j$ s represent arbitrary functions which are commonly smooth. While the functions in the model above only take as parameters the value of one covariate, with each covariate being represented in the model once, this does not need to be the case. The structure of additive models seeks to balance benefits of interpretability found in additive models with the flexibility of nonlinear, and frequently, nonparametric functional forms. Further information can be found in T.J. Hastie and R.J. Tibshirani [1990](#).

With this background in place, we turn to the motivating situation of analyzing the subcortical surfaces of ADNI participants using MR images, with a focus on the hippocampus. ADNI is a longitudinal observational study, where several imaging modalities are used to record neuroimages from participants over time. Participants can generally be considered to belong to three groups based on cognitive state: elderly healthy controls, those displaying symptoms of mild cognitive impairment, and those diagnosed with AD. In practice, participants may transition between these categorizations as their symptoms progress. However, because patients may be diagnosed with AD at different stages, accurately modeling these transitions is challenging and is beyond the scope of this article. For simplicity, we use the groupings recorded at the time

each participant was recruited to the study.

Individual participants are typically scheduled for imaging at six month intervals, though variability exists in terms of the recorded intervals. Also of note, the numbers of observations contributed by each participant varies, so groups may be unbalanced. We denote the total number of participants as  $N$ , and use  $n_i$  to represent the number of images present for each individual. Meanwhile,  $G$  denotes the number of observed groups under consideration.

Within a segmented structural MRI, the voxels estimated to mark the surface of the hippocampus can be assumed to lie along a two-dimensional manifold that has been wrapped into the three-dimensional image space. More generally, we let  $d$  denote the intrinsic dimension of the low-dimensional manifold, and  $D$  represent the dimension of the high-dimensional input space. For the  $j$ th image from the  $i$ th individual belonging to group  $g$  with  $i = 1, \dots, N$ ,  $j = 1, \dots, n_i$ , and  $g = 1, \dots, G$ , we denote the  $k$ th voxel in the high-dimensional space as  $\mathbf{x}_{ijk} \in \mathbb{R}^3$ , and its associated low-dimensional parameterization as  $\mathbf{y}_{ijk} \in \mathbb{R}^2$ . We let  $t_{ij}$  denote the time at which the  $j$ th image for individual  $i$  is recorded. The embedding function which maps from the  $d$ -dimensional manifold to  $D$ -dimensional ambient space is represented as  $\psi(\mathbf{y}_{ijk})$ , which can be modeled as

$$\mathbf{x}_{ijk} = \psi(\mathbf{y}_{ijk}) = f_p(\mathbf{y}_{ijk}, t_{ij}) + f_{g(i)}(\mathbf{y}_{ijk}, t_{ij}) + f_i(\mathbf{y}_{ijk}, t_{ij}) + f_{ij}(\mathbf{y}_{ijk}) + \epsilon.$$

Each function in this additive model represents a different level of nesting in the data, with  $f_p(\mathbf{y}, t)$  estimating the embedding at the population level, and  $f_g(\mathbf{y}, t)$ ,  $f_i(\mathbf{y}, t)$ , and  $f_{ij}(\mathbf{y})$  describing deviations from the population estimates specific to the group, individual, and image in question, respectively. As estimating longitudinal changes in the underlying manifold may be of interest,  $f_p(\mathbf{y}, t)$ ,  $f_g(\mathbf{y}, t)$  and  $f_i(\mathbf{y}, t)$  are each denoted as varying-coefficient models (Trevor Hastie and Robert Tibshirani 1993). Thus, the model is capable of estimating separate longitudinal changes for population, group, and individual estimates. In cases where longitudinal changes are not of interest, each varying-coefficient model may be constrained to be held constant with respect to time. As the  $ij$ th image is recorded at a single time point,  $f_{ij}(\mathbf{y}, t)$  may not be dependent on time. We specify  $f_{ij}(\mathbf{y})$  as a smoothing spline, while the functions  $f_p(\mathbf{y}, t)$ ,  $f_g(\mathbf{y}, t)$ , and  $f_i(\mathbf{y}, t)$  are each varying-coefficient smoothing spline models, estimated by using a smoothing spline to smooth over smoothing spline coefficients calculated at each time point available in the data.

As mentioned previously, this proposed additive spline model structure has several benefits.

## **2.2 Hierarchical Principal Manifold Estimation**

# **3 Application to Simulated and MNIST Data**

## **3.1 Application to Simulated Data**

## **3.2 Application to MNIST Data**

# **4 Application to ADNI Data**

# **5 Discussion**

## References

- Alexander, Monica and Leontine Alkema (2018). “Global Estimation of Neonatal Mortality Using a Bayesian Hierarchical Splines Regression Model”. In: *Demographic Research* 38, pp. 335–372. ISSN: 1435-9871. JSTOR: [26457049](#). URL: <https://www.jstor.org/stable/26457049> (visited on 07/06/2023).
- Belkin, Mikhail and Partha Niyogi (June 1, 2003). “Laplacian Eigenmaps for Dimensionality Reduction and Data Representation”. In: *Neural Computation* 15.6, pp. 1373–1396. ISSN: 0899-7667. DOI: [10.1162/089976603321780317](#). URL: <https://doi.org/10.1162/089976603321780317> (visited on 02/24/2023).
- Bhatia, Kanwal K. et al. (2012). “Hierarchical Manifold Learning”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*. Ed. by Nicholas Ayache et al. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 512–519. ISBN: 978-3-642-33415-3. DOI: [10.1007/978-3-642-33415-3\\_63](#).
- Brumback, Babette A. and John A. Rice (Sept. 1, 1998). “Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves”. In: *Journal of the American Statistical Association* 93.443, pp. 961–976. ISSN: 0162-1459. DOI: [10.1080/01621459.1998.10473755](#). URL: <https://doi.org/10.1080/01621459.1998.10473755> (visited on 08/09/2023).
- Coifman, Ronald R. and Stéphane Lafon (July 1, 2006). “Diffusion Maps”. In: *Applied and Computational Harmonic Analysis*. Special Issue: Diffusion Maps and Wavelets 21.1, pp. 5–30. ISSN: 1063-5203. DOI: [10.1016/j.acha.2006.04.006](#). URL: <https://www.sciencedirect.com/science/article/pii/S1063520306000546> (visited on 08/08/2023).
- Deng, Jiansong et al. (July 1, 2008). “Polynomial Splines over Hierarchical T-meshes”. In: *Graphical Models* 70.4, pp. 76–86. ISSN: 1524-0703. DOI: [10.1016/j.gmod.2008.03.001](#). URL: <https://www.sciencedirect.com/science/article/pii/S1524070308000039> (visited on 07/06/2023).
- Forsey, David R. and Richard H. Bartels (Apr. 1995). “Surface Fitting with Hierarchical Splines”. In: *ACM Transactions on Graphics* 14.2, pp. 134–161. ISSN: 0730-0301, 1557-7368. DOI: [10.1145/221659.221665](#). URL: <https://dl.acm.org/doi/10.1145/221659.221665> (visited on 07/06/2023).
- Freiche, Benoît et al. (2022). “Characterizing Myocardial Ischemia and Reperfusion Patterns with Hierarchical Manifold Learning”. In: *Statistical Atlases and Computational Models of the Heart. Multi-Disease, Multi-View, and Multi-Center Right Ventricular Segmentation in Cardiac MRI Challenge*. Ed. by Esther Puyol Antón et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 66–74. ISBN: 978-3-030-93722-5. DOI: [10.1007/978-3-030-93722-5\\_8](#).

- Gao, Siyuan, Gal Mishne, and Dustin Scheinost (2019). “A Hierarchical Manifold Learning Framework for High-Dimensional Neuroimaging Data”. In: *Information Processing in Medical Imaging*. Ed. by Albert C. S. Chung et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 631–643. ISBN: 978-3-030-20351-1. DOI: [10.1007/978-3-030-20351-1\\_49](https://doi.org/10.1007/978-3-030-20351-1_49).
- Gelman, Andrew, John B. Carlin, et al. (2014). *Bayesian Data Analysis*. 3rd. Texts in Statistical Science. Boca Raton: CRC Press/Taylor & Francis Group.
- Gelman, Andrew and Jennifer Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge: Cambridge University Press.
- Gerber, Samuel et al. (Oct. 1, 2010). “Manifold Modeling for Brain Population Analysis”. In: *Medical Image Analysis*. Special Issue on the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2009 14.5, pp. 643–653. ISSN: 1361-8415. DOI: [10.1016/j.media.2010.05.008](https://doi.org/10.1016/j.media.2010.05.008). URL: <https://www.sciencedirect.com/science/article/pii/S1361841510000617> (visited on 08/02/2023).
- Green, P. J. and B.W. Silverman (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Monographs on Statistics and Applied Probability 58. Springer.
- Guerrero, R. et al. (Mar. 1, 2017). “Group-Constrained Manifold Learning: Application to AD Risk Assessment”. In: *Pattern Recognition* 63, pp. 570–582. ISSN: 0031-3203. DOI: [10.1016/j.patcog.2016.09.023](https://doi.org/10.1016/j.patcog.2016.09.023). URL: <https://www.sciencedirect.com/science/article/pii/S0031320316303016> (visited on 01/17/2023).
- Hastie, T.J. and R.J. Tibshirani (1990). *Generalized Additive Models*. Monographs on Statistics and Applied Probability 43. London: Chapman and Hall.
- Hastie, Trevor and Werner Stuetzle (June 1989). “Principal Curves”. In: *Journal of the American Statistical Association* 84.406, pp. 502–516. ISSN: 0162-1459, 1537-274X. DOI: [10.1080/01621459.1989.10478797](https://doi.org/10.1080/01621459.1989.10478797). URL: <http://www.tandfonline.com/doi/abs/10.1080/01621459.1989.10478797> (visited on 11/29/2022).
- Hastie, Trevor and Robert Tibshirani (1993). “Varying-Coefficient Models”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 55.4, pp. 757–779. ISSN: 2517-6161. DOI: [10.1111/j.2517-6161.1993.tb01939.x](https://doi.org/10.1111/j.2517-6161.1993.tb01939.x). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1993.tb01939.x> (visited on 08/10/2023).



- He, Ying et al. (2006). “Manifold T-Spline”. In: *Geometric Modeling and Processing - GMP 2006*. Ed. by Myung-Soo Kim and Kenji Shimada. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 409–422. ISBN: 978-3-540-36865-6. DOI: [10.1007/11802914\\_29](https://doi.org/10.1007/11802914_29).
- Hendricks, Wallace and Roger Koenker (Mar. 1, 1992). “Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity”. In: *Journal of the American Statistical Association* 87.417, pp. 58–68. ISSN: 0162-1459. DOI: [10.1080/01621459.1992.10475175](https://doi.org/10.1080/01621459.1992.10475175). URL: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1992.10475175> (visited on 07/06/2023).
- Meng, Kun and Ani Eloyan (2021). “Principal Manifold Estimation via Model Complexity Selection”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 83.2, pp. 369–394. ISSN: 1467-9868. DOI: [10.1111/rssb.12416](https://doi.org/10.1111/rssb.12416). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/rssb.12416> (visited on 11/29/2022).
- Roweis, Sam T. and Lawrence K. Saul (Dec. 22, 2000). “Nonlinear Dimensionality Reduction by Locally Linear Embedding”. In: *Science* 290.5500, pp. 2323–2326. DOI: [10.1126/science.290.5500.2323](https://doi.org/10.1126/science.290.5500.2323). URL: <https://www.science.org/doi/10.1126/science.290.5500.2323> (visited on 02/24/2023).
- Schulam, Peter and Suchi Saria (2015). “A Framework for Individualizing Predictions of Disease Trajectories by Exploiting Multi-Resolution Structure”. In: *Advances in Neural Information Processing Systems*. Vol. 28. Curran Associates, Inc. URL: <https://proceedings.neurips.cc/paper/2015/hash/285e19f20beded7d215102b49d5c09a0-Abstract.html> (visited on 08/09/2023).
- (2016). “Integrative Analysis Using Coupled Latent Variable Models for Individualizing Prognoses”. In: *Journal of Machine Learning Research* 17.232, pp. 1–35. ISSN: 1533-7928. URL: <http://jmlr.org/papers/v17/15-436.html> (visited on 07/06/2023).
- Tenenbaum, Joshua B., Vin de Silva, and John C. Langford (Dec. 22, 2000). “A Global Geometric Framework for Nonlinear Dimensionality Reduction”. In: *Science* 290.5500, pp. 2319–2323. DOI: [10.1126/science.290.5500.2319](https://doi.org/10.1126/science.290.5500.2319). URL: <https://www.science.org/doi/10.1126/science.290.5500.2319> (visited on 02/24/2023).
- Wang, Yuedong (1998). “Mixed Effects Smoothing Spline Analysis of Variance”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60.1, pp. 159–174. ISSN: 1467-9868. DOI: [10.1111/1467-9868.00115](https://doi.org/10.1111/1467-9868.00115). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00115> (visited on 08/09/2023).

- Xie, Z. and G.E. Farin (Jan. 2004). “Image Registration Using Hierarchical B-splines”. In: *IEEE Transactions on Visualization and Computer Graphics* 10.1, pp. 85–94. ISSN: 1941-0506. DOI: [10.1109/TVCG.2004.1260760](https://doi.org/10.1109/TVCG.2004.1260760).
- Yue, Chen et al. (July 2, 2016). “Parameterization of White Matter Manifold-Like Structures Using Principal Surfaces”. In: *Journal of the American Statistical Association* 111.515, pp. 1050–1060. ISSN: 0162-1459. DOI: [10.1080/01621459.2016.1164050](https://doi.org/10.1080/01621459.2016.1164050). pmid: 28090127. URL: <https://doi.org/10.1080/01621459.2016.1164050> (visited on 02/24/2023).