

Hierarchical Principal Manifold Estimation

Robert Zielinski¹, Kun Meng², and Ani Eloyan¹

¹Department of Biostatistics, Brown University

²Division of Applied Mathematics, Brown University

October 3, 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 (Mueller et al. [2005](#), Marek et al. [2011](#)). Neuroimaging data in particular have been shown to be sensitive to developments in the early stages of disease progression, often before diagnosis occurs (Jack et al. [2010](#)).

While observations from neuroimaging data are typically high-dimensional, it is commonly assumed that the data lie along a low-dimensional manifold. Under this assumption, manifold learning algorithms can be used to recover this underlying manifold and parameterize the 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](#) present an alternative principal manifold framework for manifold estimation which extends to higher dimensions the principal curve framework introduced by 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 applied at the population level, with vectorized image intensities being used as the input data. 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 has led to attempts to identify subgroups of patients based on a range of criteria such as symptom expression, onset, or rate of progression (Ferreira, Nordberg, and Westman 2020, Venuto et al. 2023). 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. To develop an estimate of the manifold underlying the hippocampi of participants in the Alzheimer’s Disease Neuroimaging Initiative (ADNI) study, it is necessary to recognize the basic nesting structure of the data, in which brain images are nested within participants belong to a latent disease subgroup where patients tend to have more related symptom experiences or underlying physiological developments. These subgroups are in turn nested within study groups based on Alzheimer’s disease status, which are finally nested within the greater study population. In this paper, we develop a manifold learning approach that appropriately accounts for this hierarchical structure when estimating the manifolds underlying such data. We further propose an approach for estimation of the underlying manifold at each level of the hierarchy, allowing for statistical testing of differences between those low-dimensional manifolds.

There have been attempts 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 built 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 dimension of the images for each individual, resulting in individual-specific manifold estimates, then diffusion maps is applied again on the concatenated parameterizations to reduce over the individuals, 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 does not ensure that the manifold estimated at the individual level is structurally related

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, when using graph-based embedding methods to estimate a manifold from population data, the grouping behavior of observations recorded longitudinally from the same individual adversely impacted the creation of the neighborhood graph between data points. 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 given 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 represent 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.

In this work, we build on the principal manifold estimation (PME) algorithm introduced by 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, Hastie and 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 additive spline models used for different applications. The use of additive spline models ultimately yields a nonlinear dimension reduction algorithm that takes advantage of a statistically principled model to address nesting structures of 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 may allow 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 the MNIST dataset of handwritten digits, and compares the results to those produced using PME. The performance of the two approaches is then compared when they are applied to a series of simulated datasets of varying complexities. Section 4 discusses proposed work, which will demonstrate 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. This will be followed by 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 the dependent variable $\mathbf{y} = (y_1, \dots, y_n)^T$ is thought to be associated with independent covariate vectors $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$, for $i = 1, \dots, n$, with n representing the number of observations. 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 ϵ has mean 0 and variance σ^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

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

where again ϵ_i has mean 0 and variance σ^2 , however for each j , f_j represents an arbitrary function commonly assumed to be smooth. While the functions in model 1 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, as functions specifying interactions between covariates may also be included. The structure of additive models seeks to balance benefits of interpretability found in linear regression models with the flexibility of nonlinear, and frequently, nonparametric functional forms. Further information can be found in [Hastie and Tibshirani 1990](#).

With this background in place, we turn to the motivating example of analyzing the subcortical surfaces of ADNI participants using MR images, with a focus on the hippocampus, although the framework can be applied to modeling other structures in the brain. ADNI is a longitudinal observational study, where several imaging modalities are used to record brain images 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 manuscript. 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 the recorded intervals. Also of note, the numbers of observations contributed by each participant varies, so groups may be unbalanced. We denote the number of groups under consideration using G , the number of individuals included in the study as I , the number of images available as N , and the number of observations available for the j th image as Q_j , with $j = 1, 2, \dots, N$. The individual to which the j th image belongs is denoted as $i(j)$, the group of the $i(j)$ th individual is represented by $g(j)$, and we use t_j to denote the time at which the j th image was recorded, with all times being relative to the $i(j)$ th individual's baseline visit. This time benchmark may be changed to age, time from diagnosis, time from symptom onset, or any other measure of time, depending on the setting. Finally, we use N_t , N_i , and N_g to denote the number of images available for time point t , individual i , and group g , respectively.

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 with $j = 1, \dots, N$, we denote the q th voxel in the D -dimensional space as $\mathbf{x}_{jq} \in \mathbb{R}^D$, its associated d -dimensional parameterization as $\mathbf{y}_{jq} \in \mathbb{R}^d$, and its weight as w_{jq} . We use \mathbf{x}_g , \mathbf{x}_i , \mathbf{x}_j , and \mathbf{x}_t to denote the input voxels from all images in the g th group, i th individual, j th image, and t th time point, respectively, and use similar notation for all other relevant terms.

The embedding function which maps from the d -dimensional manifold to D -dimensional ambient space is represented as $\psi(\mathbf{y}_{jq})$, which can be modeled as

$$\mathbf{x}_{jq} = \psi(\mathbf{y}_{jq}, t_j) = f_p(\mathbf{y}_{jq}, t_j) + f_{g(j)}(\mathbf{y}_{jq}, t_j) + f_{i(j)}(\mathbf{y}_{jq}, t_j) + f_j(\mathbf{y}_{jq}) + \epsilon_{jq}. \quad (2)$$

The embedding map $\psi(\mathbf{y})$ is associated with a projection index $\kappa_\psi(\mathbf{x}) : \mathbb{R}^D \rightarrow \mathbb{R}^d$, which, informally, outputs the d -dimensional parameterization \mathbf{y} that minimizes $\|\mathbf{x} - \psi(\mathbf{y})\|_{\mathbb{R}^D}$ and follows the generalized definition of projection indices given in Meng and Eloyan 2021.

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 $f_j(\mathbf{y})$ 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, where the coefficients of each function are dependent on time (Hastie and 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 j th image is recorded at a single time point, $f_j(\mathbf{y}, t)$ may not be dependent on time. We specify $f_j(\mathbf{y})$ 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.

The smoothing spline functions $f(\mathbf{y})$ minimize

$$\mathcal{K}_\lambda(f) = \sum_{j=1}^N \sum_{q=1}^{Q_j} w_{jq} \|\mathbf{x}_{jq} - f(\mathbf{y}_{jq})\|_{\mathbb{R}^d}^2 + \lambda \|\nabla^{\otimes 2} f\|_{L^2(\mathbb{R}^d)}^2,$$

where $\nabla^{\otimes 2} f$ denotes the Hessian matrix of f , and λ is a given parameter used to penalize curvature in f . Thus, $f(\mathbf{y})$ takes the form

$$f_l(\mathbf{y}) = \sum_{j=1}^N \sum_{q=1}^{Q_j} s_{jq,l} \times \eta_d(\|\mathbf{y} - \mathbf{y}_{jq}\|) + \sum_{p=1}^{d+1} \alpha_{p,l} \times \phi_p(\mathbf{y}), \quad l = 1, 2, \dots, D, \quad (3)$$

with $\phi_p(\mathbf{y})$ representing the basis of the linear space of polynomials on \mathbb{R}^d with degree ≤ 1 and $\eta_d(y) = y^{4-d} \log y$ when $\log y \neq 0$ and d is even, $\eta_d(y) = 0$ when $\log y = 0$ and d is even, and $\eta_d(y) = y^{4-d}$ when d is odd.

Without accounting for the additive structure of the model, and after concatenating all relevant values into $Q = \sum_{j=1}^N Q_j$ -dimensional vectors of input values \mathbf{y} , weights \mathbf{w} , and outputs \mathbf{x} , we can estimate f by obtaining the values of $s_l \in \mathbb{R}^Q$ and $\alpha_l \in \mathbb{R}^{d+1}$ for $l = 1, \dots, D$ that minimize $\|\mathbf{W}^{\frac{1}{2}}(\mathbf{x}_l - \mathbf{E}s_l - \mathbf{R}\alpha_l)\|_{\mathbb{R}^Q}^2 + \lambda \|\mathbf{E}^{\frac{1}{2}}s_l\|_{\mathbb{R}^Q}^2$ under the constraint that $\mathbf{R}^T s_l = 0$, where

- $\mathbf{E} = (E_{qr})_{1 \leq q, r \leq Q}$ is defined by $E_{qr} = \eta_d(\|\mathbf{y}_q - \mathbf{y}_r\|)$.
- $\mathbf{R} = (R_{mp})_{1 \leq m \leq Q, 1 \leq p \leq d+1}$ is defined by $R_{mp} = \phi_p(\mathbf{y}_m)$.
- $\mathbf{W} = \text{diag}(w_1, \dots, w_Q)$.

With v_l representing Lagrange multipliers, the solutions for s_l and α_l can be found by solving

$$\begin{pmatrix} 2\mathbf{E}\mathbf{W}\mathbf{E} + 2\lambda\mathbf{E} & 2\mathbf{E}\mathbf{W}\mathbf{R} & \mathbf{R} \\ 2\mathbf{R}^T\mathbf{W}\mathbf{E} & 2\mathbf{R}^T\mathbf{W}\mathbf{R} & \mathbf{0} \\ \mathbf{R}^T & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} s_l \\ \alpha_l \\ v_l \end{pmatrix} = \begin{pmatrix} 2\mathbf{E}\mathbf{W}\mathbf{x}_l \\ 2\mathbf{R}^T\mathbf{W}\mathbf{x}_l \\ \mathbf{0} \end{pmatrix}. \quad (4)$$

The optimal value of λ can be found using cross-validation, where with images being randomly split into K folds, functions may be evaluated by the mean squared error, written as

$$MSD(\lambda) = \frac{1}{K} \sum_{k=1}^K \frac{1}{N_u} \sum_{j=1}^{N_u} \frac{1}{Q_j} \sum_{q=1}^{Q_j} \|\mathbf{x}_{jq} - f_{\lambda}^{(k)}(\mathbf{y}_{jq})\|^2. \quad (5)$$

Here, $f_{\lambda}^{(k)}(\mathbf{y})$ denotes the estimated spline function given tuning parameter λ and with the data for the k th fold excluded.

The coefficients and tuning parameter that minimize $\mathcal{K}_{\lambda}(f)$ suffice for estimating $f_j(\mathbf{y})$, in which the coefficients are not time-varying. To reach estimates for the varying-coefficient models, equation 4 is solved using the data available at each time point within the relevant study level. To estimate $f_p(\mathbf{y}, t)$, we use $\{\mathbf{x}_t\}_{t=1}^T$ and $\{\mathbf{y}_t\}_{t=1}^T$, yielding fitted function $\{\hat{f}_t(\mathbf{y})\}_{t=1}^T$. Using the maximum and minimum values of \mathbf{y}_t in each dimension, a new d -dimensional grid of $Q' = \max(\{Q_j\}_{j=1}^N)$ knots, $\{\mathbf{y}'_q\}_{q=1}^{Q'}$, is generated with the knots being evenly spread throughout the parameter space. These parameters serve as common knots that allow the spline coefficients to be compared across time points. They are then used to compute a new set of corresponding outcome values $\{\mathbf{x}'_t\}_{t=1}^T$. The spline functions are reestimated as $\{\hat{f}'_t\}_{t=1}^T$ using $\{\mathbf{x}'_t\}_{t=1}^T$ and \mathbf{y}' . These newly generated knots are weighted equally, so an unweighted smoothing spline model is used. Following Green and Silverman 1994, we let

- $\mathbf{E}' = (E'_{qr})_{1 \leq q, r \leq Q'}$ defined by $E'_{qr} = \eta_d(\|\mathbf{y}'_q - \mathbf{y}'_r\|)$.
- $\mathbf{R}' = (R'_{mp})_{1 \leq m \leq Q', 1 \leq p \leq d+1}$ is defined by $R'_{mp} = \phi_p(\mathbf{y}'_m)$.

Then the coefficients that minimize $(\mathbf{x}'_t - \mathbf{E}'\mathbf{s}_t - \mathbf{R}'^T\alpha_t)^T(\mathbf{x}'_t - \mathbf{E}'\mathbf{s}_t - \mathbf{R}'^T\alpha_t) + \gamma\mathbf{s}_t^T\mathbf{E}'\mathbf{s}_t$ can be found by solving

$$\begin{pmatrix} \mathbf{E}' + \lambda_t\mathbf{I} & \mathbf{R}'^T \\ \mathbf{R}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{s}'_t \\ \alpha'_t \end{pmatrix} = \begin{pmatrix} \mathbf{x}'_t \\ \mathbf{0} \end{pmatrix}. \quad (6)$$

The coefficients found at each time point t , \mathbf{s}'_t and α'_t are concatenated in matrix $\omega_t = (\mathbf{s}'_t, \alpha'_t)^T$. This coefficient matrix is then used as the output value in a cubic smoothing spline function using the time points as input values. This function takes the form

$$g_\gamma(t) = \sum_{i=1}^T \delta_i \|t - t_i\|^3 + \sum_{p=1}^2 \tau_p \phi(t)_p.$$

We define matrix \mathbf{A} as $\mathbf{A}_{qr} = \|t_q - t_r\|^3$, matrix \mathbf{T} as $\mathbf{T}_{pr} = \phi_p(t_r)$, $\delta = (\delta_1, \dots, \delta_T)^T$, and $\tau = (\tau_1, \tau_2)^T$. Then, the coefficients of $g_\gamma(t)$, given a value of γ , can be found by solving

$$\begin{pmatrix} \mathbf{A} + \gamma \mathbf{I} & \mathbf{T}^T \\ \mathbf{T} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \delta \\ \tau \end{pmatrix} = \begin{pmatrix} \omega \\ \mathbf{0} \end{pmatrix}. \quad (7)$$

As with the tuning process for λ , the optimal value of γ can be identified using cross-validation, with either leave-one-out or k -folds cross-validation being appropriate depending on the number of time points available. If using leave-one-out cross-validation, the mean squared error can be calculated as

$$MSD(\gamma) = \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \frac{1}{Q_j} \sum_{q=1}^{Q_j} \|\mathbf{x}_{jq} - f_\gamma^{(t)}(\mathbf{y}_{jq}, t_j)\|^2. \quad (8)$$

We then let $\hat{\gamma}$ be the value of γ that minimizes $MSD(\gamma)$.

With a proper estimate of the $\hat{\gamma}$, we may denote the coefficients of $g(t)$ as $\Omega = (\delta^T, \tau^T)$, and those estimated by $g(t)$ as $\omega(t) = (\mathbf{s}_{\hat{\gamma}}(t)^T, \alpha_{\hat{\gamma}}(t)^T)^T$. We use these coefficients in equation 3 in place of the constant coefficients. Therefore, the varying-coefficient models have the following outputs:

$$f(\mathbf{y}, t) = \sum_{q=1}^{Q'} \mathbf{s}_{\hat{\gamma}, l}(t) \eta_d(\|\mathbf{y} - \mathbf{y}'_m\|) + \sum_{p=1}^{d+1} \alpha_{\hat{\gamma}, p, l}(t) \phi_p(\mathbf{y}), \quad l = 1, \dots, D. \quad (9)$$

The fitting process for the varying-coefficient smoothing spline model is given formally as Algorithm 1.

While the procedure outlined in Algorithm 1 yields an appropriate estimator when working with unnested data, this approach must be adapted to work within the additive model structure of model 2. As discussed in Hastie and Tibshirani 1990, the additive model $\psi(\mathbf{y}, t)$ can be estimated using a backfitting procedure, which uses the model's additive nature to iteratively fit each distinct smoothing spline function until convergence. Specifically, we use the fact that, under the model,

$$f_p(\mathbf{y}_{jq}, t_j) = E \left[\mathbf{x}_{jq} - f_{g(j)}(\mathbf{y}_{jq}, t_j) - f_{i(j)}(\mathbf{y}_{jq}, t_j) - f_j(\mathbf{y}_{jq}) \right],$$

with a similar result holding for each of the smooth functions making up the additive model. The model in its entirety is considered to have converged when the maximum of the distances between the coefficients

Algorithm 1: Varying-Coefficient Smoothing Spline Model

Data: D -dimensional output values $\{\mathbf{x}_{jq}\}_{j=1,q=1}^{N,Q_j}$, d -dimensional input values $\{\mathbf{y}_{jq}\}_{j=1,q=1}^{N,Q_j}$, weights $\{w_{jq}\}_{j=1,q=1}^{N,Q_j}$, observation times $\{t_j\}_{j=1}^N$, candidate tuning parameters $\{\lambda_l\}_{l=1}^L$, $\{\gamma_p\}_{p=1}^P$, and number of cross-validation folds K .

Result: Analytic formula of $\hat{f}(\mathbf{y}, t)$, coefficients $\Omega = (\delta^T, \tau^T)$, and optimal tuning parameter $\hat{\gamma}$.

```

1 for  $t = 1, 2, \dots, T$  do
2   Set  $U_t$  equal to the number of images included in the data at time  $t$ ;
3   for  $l = 1, 2, \dots, L$  do
4     Solve equation 4 to estimate  $f_{\lambda_l,t}(\mathbf{y})$  with inputs  $\mathbf{x}_t, \mathbf{y}_t$ , and  $w_t$ ;
5     for  $u = 1, 2, \dots, U_t$  do
6       Solve equation 4 to estimate  $f_{\lambda_l,t}^{(-u)}(\mathbf{y})$  with inputs  $\mathbf{x}_t^{(-u)}, \mathbf{y}_t^{(-u)}, w_t^{(-u)}$ ;
7       Estimate  $MSD(\lambda_l)_u$  using equation 5;
8     end
9     Set  $MSD(\lambda_l) = \frac{1}{U} \sum_{u=1}^U MSD(\lambda_l)_u$ ;
10  end
11   $\hat{\lambda}_t = \arg \min_{\lambda} MSD(\lambda)$ ;
12   $f_t(\mathbf{y}) = f_{\hat{\lambda}_t}(\mathbf{y})$ ;
13 end
14 Let  $Q' = \max(Q_j)$ ,  $\{\mathbf{y}'_q\}_{q=1}^{Q'}$  be a grid spanning the range of values of  $\{\mathbf{y}_{jq}\}_{j=1,q=1}^{N,Q_j}$ ;
15 for  $t = 1, 2, \dots, T$  do
16   Compute  $\{\mathbf{x}'_{q,t}\}_{q=1}^{Q'} = \{f_t(\mathbf{y}'_q)\}_{q=1}^{Q'}$ ;
17   Solve equation 6 to compute  $\mathbf{s}'_t$  and  $\alpha'_t$ ;
18   Set  $\omega_t = [\mathbf{s}'_t^T, \alpha'_t]^T$ ;
19 end
20 for  $p = 1, 2, \dots, P$  do
21   Compute  $g_{\gamma_p}(t)$  by solving equation 7;
22   for  $t = 1, 2, \dots, T$  do
23     Compute  $g_{\gamma_p}^{(t)}$  by solving 7;
24   end
25   Estimate  $MSD(\gamma_p)$  using 8;
26 end
27  $\hat{\gamma} = \arg \min_{\gamma} MSD(\gamma)$ ;
28  $\hat{f}(\mathbf{y}, t) = f_{\hat{\gamma}}(\mathbf{y}, t)$ , where the form of  $\hat{f}(\mathbf{y}, t)$  is given in 9.

```

of functions estimated for the same element in consecutive steps is below a given threshold. Details of the backfitting algorithm to fit the proposed model is given in Algorithm 2.

This proposed additive spline model structure has several benefits. First, it maintains interpretability at each level of the model. It is possible to visually inspect the embedding maps of the population and each group and individual to better understand how they interact. While the model described here is intended to be applicable to the specific situation involving the modeling of data from the ADNI study, this general modeling approach can be conveniently adapted to a variety of scenarios. Additionally, basing our handling of the data’s hierarchical structure on a statistical model allows us to assess statistical properties of the estimated manifold, where less statistically principled approaches make this option difficult. Of particular interest, Wang 1998 and Brumback and Rice 1998 explore connections between smoothing spline models and mixed-effects models. The smoothing spline analysis of variance proposed by Wang 1998 may be applicable to the model structure proposed here, which would enable a detailed decomposition of the variation in manifold structure by level. For the time being, this extension is left to future work.

A clear concern with the use of this procedure in this case is the number of functions to be estimated, which can quickly increase the computational burden of fitting the models. However, the nesting of the different levels of the data helps to limit the number of function combinations that must be fit, helping to keep computational requirements manageable. Brumback and Rice 1998 used restricted maximum likelihood estimation to reach an estimate of their proposed additive spline model, which may potentially provide computational improvements. However, in the case that the model should be modified to meet the needs of a new dataset, the backfitting algorithm will be more easily adaptable.

2.2 Hierarchical Principal Manifold Estimation

With the structure of the additive spline model established, we now incorporate it into the larger manifold learning process. This procedure is based on the Principal Manifold Estimation (PME) algorithm introduced in Meng and Eloyan 2021, which iterates between estimating a function to embed observations in D -dimensional space and reparameterizing those observations in the d -dimensional space to estimate the underlying manifold. The PME algorithm is largely split into three distinct steps: data reduction, fitting, and tuning. The split between these three steps enforces a process by which candidate functions are fit for each element in a vector of candidate smoothing values in the fitting stage, then the tuning step is used to

Algorithm 2: Backfitting Algorithm for Additive Spline Model

Data: D -dimensional output values $\{\mathbf{x}_{jq}\}_{j=1,q=1}^{N,Q_j}$, d -dimensional input values $\{\mathbf{y}_{jq}\}_{j=1,q=1}^{N,Q_j}$, observation times $\{t_j\}_{j=1}^N$, weights $\{w_{jq}\}_{j=1,q=1}^{N,Q_j}$, group and individual indicator values $\{g(j)\}_{j=1}^N$ and $\{i(j)\}_{j=1}^N$, candidate tuning parameters $\{\lambda_l\}_{l=1}^L$ and $\{\gamma_p\}_{p=1}^P$, ϵ and K .

Result: Analytic formulas of \hat{f}_p , $\{\hat{f}_g\}_{g=1}^G$, $\{\hat{f}_i\}_{i=1}^I$, and $\{\hat{f}_j\}_{j=1}^N$.

```

1 Estimate  $f_p^{(0)}(\mathbf{y}, t)$  and  $\Omega_p^{(0)}$  using Algorithm 1 with inputs  $\{\mathbf{x}_{jq}\}_{j=1,q=1}^{N,Q_j}$ ,  $\{\mathbf{y}_{jq}\}_{j=1,q=1}^{N,Q_j}$ ,  $\{w_{jq}\}_{j=1,q=1}^{N,Q_j}$ ,  $\{t_j\}_{j=1}^N$ ,  $\{\lambda_l\}_{l=1}^L$ ,  $\{\gamma_p\}_{p=1}^P$ , and  $K$ ;
2 for  $g = 1, 2, \dots, G$  do
3   Estimate  $f_g^{(0)}(\mathbf{y}, t)$  and  $\Omega_g^{(0)}$  using Algorithm 1 with inputs  $\{\mathbf{x}_{jq} - f_p^{(0)}(\mathbf{y}_{jq}, t_j)\}_{j=1,q=1}^{N,Q_j}$ ,  $\{\mathbf{y}_{jq}\}_{j=1,q=1}^{N,Q_j}$ ,  $\{w_{jq}\}_{j=1,q=1}^{N,Q_j}$ ,  $\{t_j\}_{j=1}^N$ ,  $\{\lambda_l\}_{l=1}^L$ ,  $\{\gamma_p\}_{p=1}^P$ ,  $K$ ;
4 end
5 for  $i = 1, 2, \dots, I$  do
6   Estimate  $f_i^{(0)}(\mathbf{y}, t)$  and  $\Omega_i^{(0)}$  using Algorithm 1 with inputs  $\{\mathbf{x}_{jq} - f_p^{(0)}(\mathbf{y}_{jq}, t_j) - f_{g(j)}^{(0)}(\mathbf{y}_{jq}, t_j)\}_{j=1,q=1}^{N_i,Q_j}$ ,  $\{\mathbf{y}_{jq}\}_{j=1,q=1}^{N_i,Q_j}$ ,  $\{w_{jq}\}_{j=1,q=1}^{N_i,Q_j}$ ,  $\{t_j\}_{j=1}^{N_i}$ ,  $\{\lambda_l\}_{l=1}^L$ ,  $\{\gamma_p\}_{p=1}^P$ , and  $K$ ;
7 end
8 for  $j = 1, 2, \dots, N$  do
9   Estimate  $f_j^{(0)}(\mathbf{y})$  and  $\omega_j^{(0)}$  by solving 4 with inputs  $\{\mathbf{x}_{jq} - f_p^{(0)}(\mathbf{y}_{jq}, t_j) - f_{g(j)}^{(0)}(\mathbf{y}_{jq}, t_j) - f_{i(j)}^{(0)}(\mathbf{y}_{jq}, t_j)\}_{q=1}^{Q_j}$ ,  $\{\mathbf{y}_{jq}\}_{q=1}^{Q_j}$ ,  $\{w_{jq}\}_{q=1}^{Q_j}$ ,  $\{\lambda_l\}_{l=1}^L$ ,  $K$ ;
10 end
11  $n \leftarrow 0$ ;
12 while  $\max \left( \|\Omega_p^{(n)} - \Omega_p^{(n-1)}\|, \left\{ \|\Omega_g^{(n)} - \Omega_g^{(n-1)}\| \right\}_{g=1}^G, \left\{ \|\Omega_i^{(n)} - \Omega_i^{(n-1)}\| \right\}_{i=1}^I, \left\{ \|\omega_j^{(n)} - \omega_j^{(n-1)}\| \right\}_{j=1}^N \right) \geq \epsilon$  do
13   Estimate  $f_p^{(n+1)}(\mathbf{y}, t)$  and  $\Omega_p^{(n+1)}$  using Algorithm 1 with inputs  $\{\mathbf{x}_{jq} - f_{g(j)}^{(n)}(\mathbf{y}_{jq}, t_j) - f_{i(j)}^{(n)}(\mathbf{y}_{jq}, t_j) - f_j^{(n)}(\mathbf{y}_{jq})\}_{j=1,q=1}^{N,Q_j}$ ,  $\{\mathbf{y}_{jq}\}_{j=1,q=1}^{N,Q_j}$ ,  $\{w_{jq}\}_{j=1,q=1}^{N,Q_j}$ ,  $\{t_j\}_{j=1}^N$ ,  $\{\lambda_l\}_{l=1}^L$ ,  $\{\gamma_p\}_{p=1}^P$ , and  $K$ ;
14   for  $g = 1, 2, \dots, G$  do
15     Estimate  $f_g^{(n+1)}(\mathbf{y}, t)$  and  $\Omega_g^{(n+1)}$  using Algorithm 1 with inputs  $\{\mathbf{x}_{jq} - f_p^{(n+1)}(\mathbf{y}_{jq}, t_j) - f_{i(j)}^{(n)}(\mathbf{y}_{jq}, t_j) - f_j^{(n)}(\mathbf{y}_{jq})\}_{j=1,q=1}^{N_g,Q_j}$ ,  $\{\mathbf{y}_{jq}\}_{j=1,q=1}^{N_g,Q_j}$ ,  $\{w_{jq}\}_{j=1,q=1}^{N_g,Q_j}$ ,  $\{t_j\}_{j=1}^{N_g}$ ,  $\{\lambda_l\}_{l=1}^L$ ,  $\{\gamma_p\}_{p=1}^P$ ,  $K$ ;
16   end
17   for  $i = 1, 2, \dots, I$  do
18     Estimate  $f_i^{(n+1)}(\mathbf{y}, t)$  and  $\Omega_i^{(n+1)}$  using Algorithm 1 with inputs  $\{\mathbf{x}_{jq} - f_p^{(n+1)}(\mathbf{y}_{jq}, t_j) - f_{g(j)}^{(n+1)}(\mathbf{y}_{jq}, t_j) - f_j^{(n)}(\mathbf{y}_{jq})\}_{j=1,q=1}^{N_i,Q_j}$ ,  $\{\mathbf{y}_{jq}\}_{j=1,q=1}^{N_i,Q_j}$ ,  $\{w_{jq}\}_{j=1,q=1}^{N_i,Q_j}$ ,  $\{t_j\}_{j=1}^{N_i}$ ,  $\{\lambda_l\}_{l=1}^L$ ,  $\{\gamma_p\}_{p=1}^P$ , and  $K$ ;
19   end
20   for  $j = 1, 2, \dots, N$  do
21     Estimate  $f_j^{(n+1)}(\mathbf{y})$  and  $\omega_j^{(n+1)}$  by solving 4 with inputs  $\{\mathbf{x}_{jq} - f_p^{(n+1)}(\mathbf{y}_{jq}, t_j) - f_{g(j)}^{(n+1)}(\mathbf{y}_{jq}, t_j) - f_{i(j)}^{(n+1)}(\mathbf{y}_{jq}, t_j)\}_{q=1}^{Q_j}$ ,  $\{\mathbf{y}_{jq}\}_{q=1}^{Q_j}$ ,  $\{w_{jq}\}_{q=1}^{Q_j}$ ,  $\{\lambda_l\}_{l=1}^L$ ,  $K$ ;
22   end
23    $n \leftarrow n + 1$ ;
24 end
25  $\hat{f}_p \leftarrow f_p^{(n)}$ ,  $\{\hat{f}_g\}_{g=1}^G \leftarrow \{f_g^{(n)}\}_{g=1}^G$ ,  $\{\hat{f}_i\}_{i=1}^I \leftarrow \{f_i^{(n)}\}_{i=1}^I$ , and  $\{\hat{f}_j\}_{j=1}^N \leftarrow \{f_j^{(n)}\}_{j=1}^N$ .

```

select the smoothing value that results in the lowest error, as measured by mean squared distance,

$$\mathcal{D}(\hat{f}) = \frac{1}{N} \sum_{i=1}^N \|x_i - \hat{f}_\lambda(\pi_{\hat{f}_\lambda}(x_i))\|_{\mathbb{R}^D}^2,$$

where \hat{f}_λ represents the estimated embedding map given smoothing parameter λ , $\pi_{\hat{f}_\lambda}(x)$ denotes the projection index that obtains a d -dimensional parameterization of D -dimensional input x , and N represents the total number of observations in the dataset.

While this approach to tuning the model parameters is beneficial because it allows a complete comparison of all parameter values under consideration, the additive model described in equation 2 requires many more parameters, meaning that taking the same tuning approach would call for finding an optimal point in a very high-dimensional space. This would greatly increase the computational complexity of the algorithm, likely to the point of infeasibility in an practical situation. Instead, the proposed algorithm avoids this problem by taking a greedy approach to tuning the model, where the smoothing parameters for each model are chosen individually during the spline model’s tuning step, rather than simultaneously.

Thus, the hierarchical PME algorithm proposed here is composed of initialization, data reduction, and fitting steps. The first step uses the PME algorithm to estimate an embedding map from one image. The HDMDE algorithm, detailed in Meng and Eloyan 2021, is then used to reduce the data from each image into a set of component centers and associated weights. The embedding map estimated using the PME algorithm can be used to provide initial parameterizations of the component centers for every image. These parameterizations are then used as starting points in the fitting step.

Like the PME algorithm, the fitting step uses an iterative approach to alternate between using the parameterizations to estimate an embedding function with the form of the additive spline model in equation 2, then using this newly generated embedding function to reestimate the most appropriate parameterizations for the reduced set of components identified in the data reduction process. Once the error of the embedding function as measured by the sum of squared distances measured at the component centers declines by less than a given threshold ϵ , the iterative procedure stops and the embedding function and associated parameterizations are finalized. The details of this process are given in Algorithm 3.

3 Application to MNIST Data and Simulations

To demonstrate the potential benefits of the proposed hierarchical estimation approach when compared to the existing PME algorithm, we implemented the hierarchical PME algorithm using a simplified version of

Algorithm 3: Hierarchical Principal Manifold Estimation Algorithm

Data: Data points $\{\mathbf{v}_{jq}\}_{j=1,q=1}^{N,Q'_j}$ in \mathbb{R}^D ; group and individual indicator values $\{g(j)\}_{j=1}^N$ and $\{i(j)\}_{j=1}^N$; a positive integer $N_0 < \min \left(\left\{ Q'_j \right\}_{j=1}^N \right)$; $\alpha, \epsilon, \epsilon^* \in (0, 1)$; candidate tuning parameters $\{\lambda_l\}_{l=1}^L, \{\gamma_p\}_{p=1}^P$; $itr \geq 1$, which is the maximum number of iterations allowed.

Result: Analytic formulas of $\hat{f}_p, \left\{ \hat{f}_g \right\}_{g=1}^G, \left\{ \hat{f}_i \right\}_{i=1}^I$, and $\left\{ \hat{f}_j \right\}_{j=1}^N$.

- 1 **Initialization:** Apply PME with input $\left(\left\{ \mathbf{v}_{1m} \right\}_{q=1}^{Q'_1}, N_0, \alpha, \epsilon, \epsilon^*, \{\lambda_l\}_{l=1}^L, itr \right)$ and obtain $f^{(1)}$;
 - 2 $\psi^{(1)} \leftarrow f^{(1)}$;
 - 3 **Data Reduction:** Apply HDMDE algorithm with inputs $\left(\left\{ \left\{ \mathbf{v}_{jq} \right\}_{q=1}^{Q_j} \right\}_{j=1}^N, N_0, \epsilon, \alpha \right)$ and obtain $\{Q_j\}_{j=1}^N, \{\mathbf{x}_{jq}\}_{j=1,q=1}^{N,Q_j}$, and $\{w_{jq}\}_{j=1,q=1}^{N,Q_j}$;
 - 4 **Parameterization:** Use the projection index of $\psi^{(1)}, \kappa_{\psi^{(1)}}$, to parameterize $\{\mathbf{x}_{jq}\}_{j=1,q=1}^{N,Q_j}$ by d -dimensional parameters $\{y_{jq}\}_{j=1,q=1}^{N,Q_j}$;
 - 5 $\mathcal{E} \leftarrow 2 \times \epsilon^*, n \leftarrow 1$, and $\mathcal{D}(\psi^{(1)}) \leftarrow \sum_{j=1}^N \sum_{q=1}^{Q_j} w_{jq} \|\mathbf{x}_{jq} - \psi^{(1)}(\kappa_{\psi^{(1)}}(\mathbf{x}_{jq}))\|_{\mathbb{R}^D}^2$;
 - 6 **while** $\mathcal{E} \geq \epsilon^*$ **and** $n < itr$ **do**
 - 7 Compute $f_p^{(n+1)}, \left\{ f_g^{(n+1)} \right\}_{g=1}^G, \left\{ f_i^{(n+1)} \right\}_{i=1}^I$, and $\left\{ f_j^{(n+1)} \right\}_{j=1}^N$ by using Algorithm 2 with inputs $\{\mathbf{x}_{jq}\}_{j=1,q=1}^{N,Q_j}, \left\{ \kappa_{\psi^{(n)}}(\mathbf{x}_{jq}) \right\}_{j=1,q=1}^{N,Q_j}, \{t_j\}_{j=1}^N, \{w_{jq}\}_{j=1,q=1}^{N,Q_j}, \{g(j)\}_{j=1}^N, \{i(j)\}_{j=1}^N, \{\lambda_l\}_{l=1}^L, \{\gamma_p\}_{p=1}^P, \epsilon$, and K ;
 - 8 $\left\{ \psi^{(n+1)}(\mathbf{y}, t) \leftarrow f_p^{(n+1)}(\mathbf{y}, t) + f_{g(j)}^{(n+1)}(\mathbf{y}, t) + f_{i(j)}^{(n+1)}(\mathbf{y}, t) + f_j^{(n+1)}(\mathbf{y}) \right\}_{j=1}^N$;
 - 9 $\mathcal{D}(\psi^{n+1}) \leftarrow \sum_{j=1}^N \sum_{q=1}^{Q_j} w_{jq} \|\mathbf{x}_{jq} - \psi^{n+1}(\kappa_{\psi^{n+1}}(\mathbf{x}_{jq}))\|_{\mathbb{R}^D}^2$;
 - 10 $\mathcal{E} \leftarrow |\mathcal{D}(\psi^{(n+1)}) - \mathcal{D}(\psi^{(n)})| / \mathcal{D}(\psi^{(n)})$;
 - 11 $n \leftarrow n + 1$;
 - 12 **end**
 - 13 $\hat{f}_p \leftarrow f_p^{(n)}, \left\{ \hat{f}_g \leftarrow f_g^{(n)} \right\}_{g=1}^G, \left\{ \hat{f}_i \leftarrow f_i^{(n)} \right\}_{i=1}^I, \left\{ \hat{f}_j \leftarrow f_j^{(n)} \right\}_{j=1}^N$.
-

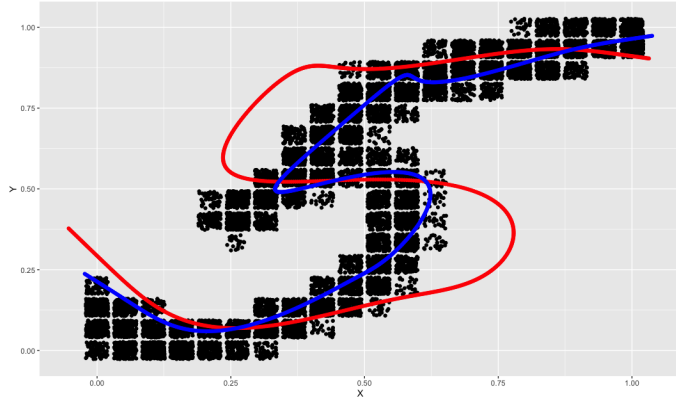
the additive spline model presented in Section 2 for use on images from the MNIST database. The MNIST database includes 70,000 28×28 images of handwritten numerical digits written by different individuals, with 60,000 in a training set and an additional 10,000 in a test set. To test the hierarchical algorithm, we considered a reduced set of 100 images of the numeral “5”, where the individual pixels of each image with intensity greater than 0 can be considered to create a 2-dimensional data cloud, through which a 1-dimensional curve passes. Each representation of the numeral must follow the same general path to be recognizable as a “5”, but differences in individual handwriting styles result in variation between images. We have no meaningful way of grouping images by individual or subgroup in this case, and there are no temporal relationships between images, so we assume the embedding function used in Algorithm 3 takes the form

$$\mathbf{x}_{jq} = \psi(\mathbf{y}_{jq}) = f_p(\mathbf{y}_{jq}) + f_j(\mathbf{y}_{jq}) + \epsilon_{jq}.$$

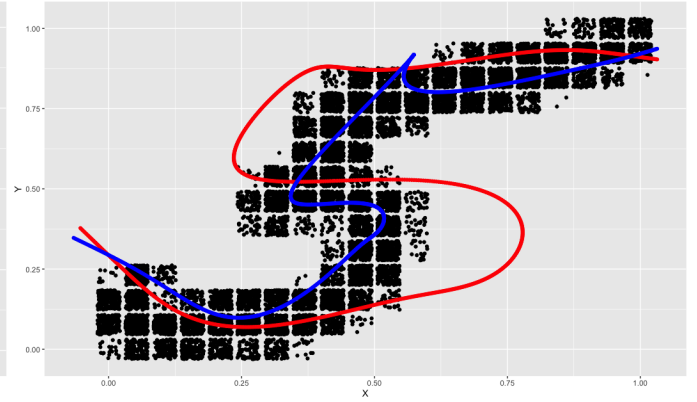
Plots depicting how the population-level manifold estimate differs from the manifolds estimated for each image are shown for four images in Figure 1. Here, we see that the population estimate, represented in red, clearly captures the shape of a numeral “5” and does not show any clear bias compared to the data points included in the plots. Meanwhile, the estimates for the individual images, shown in blue, are more responsive to the image data to which it is fit, indicating that the additive model fit successfully to the data, and providing a simple demonstration of the improvements made available by incorporating hierarchical information into the manifold learning procedure.

4 Proposed Work

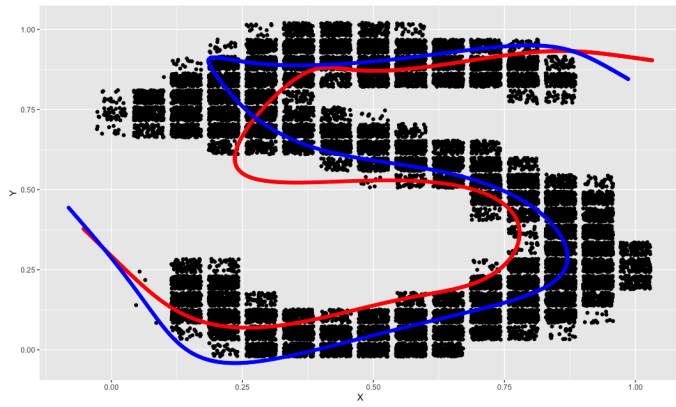
Further work will be dedicated to testing the proposed approach in several settings. First, more extensive performance comparisons between the PME algorithm and its hierarchical extension will be conducted when applied to the MNIST dataset. The method will next be applied to a series of simulated datasets that were generated to reflect the hierarchical structure present in the ADNI dataset. Again, the estimates of the hierarchical approach will be compared to alternative methods, with the fit of each method to the data being compared. Finally, the hierarchical approach will be applied to structural MRI data from the ADNI dataset. Specifically, the algorithm can be used to estimate the manifolds underlying the hippocampus of study participants. Of interest in this case will be a comparison of the varying-coefficient models by group, which may provide a visual confirmation that the algorithm’s estimates are detecting the most relevant information included in the data.



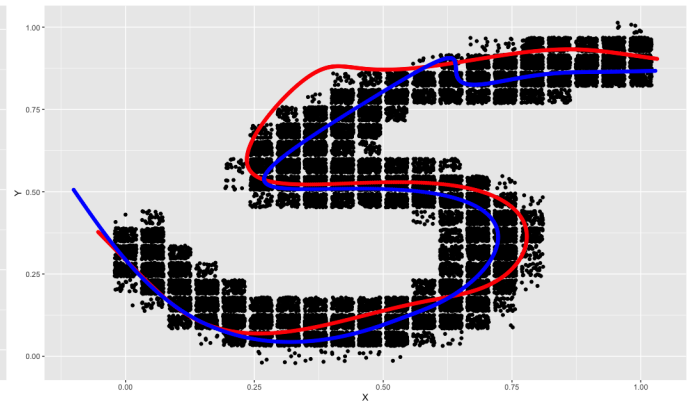
(a) MNIST Image 1



(b) MNIST Image 2



(c) MNIST Image 3



(d) MNIST Image 4

Figure 1: Comparison of PME, Population, and Image-Level Results for MNIST Data

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).
- Ferreira, Daniel, Agneta Nordberg, and Eric Westman (Mar. 10, 2020). “Biological Subtypes of Alzheimer Disease: A Systematic Review and Meta-Analysis”. In: *Neurology* 94.10, pp. 436–448. ISSN: 1526-632X. DOI: [10.1212/WNL.0000000000009058](#). pmid: [32047067](#).
- 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 Es-

- ther 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](https://doi.org/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, 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 (1990). *Generalized Additive Models*. Monographs on Statistics and Applied Probability 43. London: Chapman and Hall.
- (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).
- Jack, Clifford R. et al. (Jan. 2010). “Hypothetical Model of Dynamic Biomarkers of the Alzheimer’s Pathological Cascade”. In: *The Lancet. Neurology* 9.1, pp. 119–128. ISSN: 1474-4465. DOI: [10.1016/S1474-4422\(09\)70299-6](https://doi.org/10.1016/S1474-4422(09)70299-6). pmid: [20083042](https://pubmed.ncbi.nlm.nih.gov/20083042/).
- Marek, Kenneth et al. (Dec. 1, 2011). “The Parkinson Progression Marker Initiative (PPMI)”. In: *Progress in Neurobiology*. Biological Markers for Neurodegenerative Diseases 95.4, pp. 629–635. ISSN: 0301-0082. DOI: [10.1016/j.pneurobio.2011.09.005](https://doi.org/10.1016/j.pneurobio.2011.09.005). URL: <https://www.sciencedirect.com/science/article/pii/S0301008211001651> (visited on 08/17/2023).
- 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).
- Mueller, Susanne G. et al. (July 2005). “Ways toward an Early Diagnosis in Alzheimer’s Disease: The Alzheimer’s Disease Neuroimaging Initiative (ADNI)”. In: *Alzheimer’s & Dementia* 1.1, pp. 55–66. ISSN: 1552-5260, 1552-5279. DOI: [10.1016/j.jalz.2005.06.003](https://doi.org/10.1016/j.jalz.2005.06.003). URL: <https://onlinelibrary.wiley.com/doi/10.1016/j.jalz.2005.06.003> (visited on 08/17/2023).
- 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).

- Schulam, Peter and Suchi Saria (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).
- Venuto, Charles S. et al. (2023). “Predicting Ambulatory Capacity in Parkinson’s Disease to Analyze Progression, Biomarkers, and Trial Design”. In: *Movement Disorders* n/a.n/a (). ISSN: 1531-8257. DOI: [10.1002/mds.29519](https://onlinelibrary.wiley.com/doi/abs/10.1002/mds.29519). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mds.29519> (visited on 08/17/2023).
- 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://onlinelibrary.wiley.com/doi/abs/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).