Estimating the Coefficients of a Linear Differential Operator by Resampling

María Ivette Barraza-Ríos Joan G. Staniswalis
The University of Texas at El Paso, Department of Mathematical Sciences,
500 W. University Ave., El Paso, Texas 79902

Abstract
Principal Differential Analysis (PDA; Ramsay 1996) is used to obtain low-dimensional representations of functional data, where each observation is a curve. PDA seeks to identify a Linear Differential Operator (LDO) of the form $L = \omega_0 I + \omega_1 D + \cdots + \omega_{m-1} D^{m-1} + \omega_m D^m$ that satisfies as closely as possible that $Lx(t) = 0$ for each functional observation $x(t)$. A theorem from analysis establishes that there exists an LDO with coefficients in the Sobolev space, and thus can be approximated by B-splines. Current PDA software used to estimate the LDO assumes that the leading coefficient ω_m is 1, but the Sobolev space is not closed with respect to division. We present a method that eliminates this restriction to ensure that the coefficients of the LDO are in the Sobolev space, and that their approximation by B-splines is mathematically valid. The proposed method is inspired by results in linear regression (Frees 1991; Wu 1986) that show that the weighted average of pairwise slopes between data points is equivalent to the least squares estimator of the regression line slope. By analyzing data, our approach is compared with `pda.fd` (`R` library `fda`).

Key Words: Functional data, principal differential analysis, low-dimensional approximation.

1. Functional Data

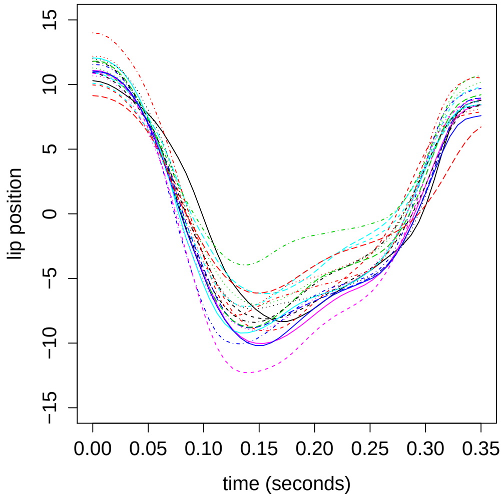


Figure 1: Unregistered lip data, 20 curves

The term "functional data" refers to data in which each response may be represented as a curve. Functional data analysis arose because advances in technology allow for the recording of responses on a finer grid. For example, atmospheric phenomena that used to be monitored once a day are now monitored (almost) continuously.

Figure 1 illustrates an example of functional data considered by Ramsay and Silverman (1997). The position of the center of the bottom lip of a person saying the syllable "bob" was recorded. For each of the 20 vocalizations of the syllable “bob”, the position of the lip was recorded on a regular grid of length 51 over a time interval of .35 seconds. In order to obtain a response that could be considered a curve on the interval [0, 0.35] in time, such as shown in figure 1, a smoothing spline of order 6 and penalty on the fourth derivative with $\lambda = 10^{-12}$ was fit using the R function `smooth.basisPar` in Ramsay’s R library `fda`.

The methodology called Principal Differential Analysis (PDA) was developed by Ramsay (1996). Its objective is to “use a set of n functional observations (curves) $\{x_i\}_{i=1}^{n}$ to define a much smaller set of m functions $\{u_j\}_{j=1}^{m}$ on the basis of which we can obtain efficient approximations of the observed functions” (Ramsay and Silverman 1997, p. 239).

1.1 Principal Differential Analysis

PDA assumes that there exists a linear differential operator (LDO) defined by

$$L = \omega_0 I + \omega_1 D + \cdots + \omega_{m-1} D^{m-1} + \omega_m D^m \tag{1}$$

that satisfies $Lx_i = 0$ for each functional observation $x_i(t) \quad (i = 1, \cdots, n)$. The coefficients of the LDO (the functions $\omega_0, \cdots, \omega_m$, which depend on t) need to be estimated such that the LDO annihilates the collection of curves as closely as possible. Recall, that the null space of the LDO is the collection of all functions annihilated by the LDO.

We know from results in analysis (Coddington and Levinson 1955, Theorem 6.2) that a collection of m functions from a Sobolev space has such an annihilating LDO of order m. Jin, Staniswalis, and Mallawaarachchi (2013, Theorem 1) provide conditions under which the coefficients of an annihilating LDO are also in a Sobolev space, and thus can be approximated well by B-splines.

Note that the coefficients of the LDO are not unique, since the null space is invariant to multiplication of the LDO by any non-zero function from the Sobolev space. Ramsay (1996) avoids this problem by assuming that the coefficient of the leading derivative in the LDO is $\omega_m(t) = 1$. This is equivalent to dividing through by ω_m in the LDO in order to obtain a unique solution for the coefficients. However, the Sobolev space is not closed under division, in which case we can no longer assume that the coefficients $\omega_0, \ldots, \omega_{m-1}$ are in the Sobolev space and approximated well by B-splines. Ramsay leaves this concern aside in the hope of finding a useful working solution. Ramsay represents the smooth coefficients $\omega_0(t), \cdots, \omega_{m-1}(t)$ in a B-spline basis (see Appendix), then estimates the smooth coefficients of the LDO by minimizing $\sum_{i=1}^{n} \|Lx_i\|^2 = \sum_{i=1}^{n} \int_0^T (Lx_i(t))^2 dt$ subject to a penalty term (see `R` package `fda`). Once the coefficients of the LDO are estimated, numerical methods to solve differential equations are used to construct a basis for the null space of the LDO. Ramsay (1996) uses the Runge-Kutta numerical method to find a basis for the null space of the LDO with coefficients estimated from the data when $\omega_m(t) = 1$. Dropping the assumption $\omega_m(t) = 1$ requires the use of a different set of numerical methods, namely, those that solve implicit systems of differential equations. Then, normalized basis functions $u_1(t), \ldots, u_m(t)$ are found by dividing each basis function by the square root of the norm.

Finally, since each curve in the functional data set is believed to be an element of the null space of the LDO, a low-dimensional approximation is obtained by regressing each curve on the normalized basis functions $u_1(t), \ldots, u_m(t)$.

2. Resampling Methods

Suppose we are given data of the form (x_i, y_i) for $i = 1, \ldots, n$ satisfying $y_i = \beta_o + \beta_1 x_i + \epsilon_i$, where $\epsilon_1, \ldots, \epsilon_n$ are iid random variables with mean 0 and common variance σ^2. Frees (1991) investigated robust estimators of β_o, β_1 by pursuing the alternative characterization of the ordinary least squares estimator of the slope parameter. This is described next: compute all pairwise slopes

$$s_{ij} = \frac{y_j - y_i}{x_j - x_i},$$

and set

$$\beta_{Fr} = \frac{\sum w_{ij} s_{ij}}{\sum w_{ij}}.$$

Here, β_{Fr} is a linear combination of all $\binom{n}{2}$ pairwise slopes. It can be shown that choosing $w_{ij} = (x_i - x_j)^2$ yields the least squares estimator

$$\hat{\beta}_1 = \sum_{i=1}^{n} X_i(Y_i - \bar{Y}) / \sum_{i=1}^{n} X_i(X_i - \bar{X}).$$

If instead, we first order the pairwise slopes, then trim unusually high or low slopes, an M-type estimator of the slope parameter is obtained.

The ideas in Frees (1991) inspired the resampling based method for estimation of the coefficients $\omega_0, \ldots, \omega_m$ of the LDO in equation (1) presented here. We compute the annihilating LDO for each collection of m curves sampled from the full collection of n curves. This is computationally intensive because there are $\binom{n}{m}$ subproblems that must be solved for estimation of the LDO. Then a linear combination of the $\binom{n}{m}$ LDO's is used as the final estimate $\hat{L}$ of the LDO in equation (1).

Theorem 2.1 below provides a justification for breaking up the problem into $\binom{n}{m}$ subproblems. In the case of an iid sample of size n, the collection of minimal sufficient statistics for all $\binom{n}{m}$ subproblems is sufficient for the family of joint distributions of the sample. Let $\mathcal{F} = \{f_\theta(x), \theta \in \Omega\}$ denote a familiy of distributions with common support, and $X_1, \ldots, X_n$ a random sample that is iid f_θ. The joint distribution of the sample is given by $f_\theta(\mathbf{x}) = \prod_{i=1}^{n} f_\theta(x_i)$. Let $\mathbf{x}^* = (x_1^*, \ldots, x_m^*)$ denote a sample of size m selected without replacement from $\{x_1, \ldots, x_n\}$.

Theorem 2.1. *Suppose there is a k and $\theta_j \in \Omega$ for $j = 0, \ldots, k$ so that*

$$T(\mathbf{x}) = (T_1(\mathbf{x}), \ldots, T_k(\mathbf{x})) = \left(\frac{f_1(\mathbf{x})}{f_0(\mathbf{x})}, \ldots, \frac{f_k(\mathbf{x})}{f_0(\mathbf{x})}\right) \tag{2}$$

is minimal sufficient for $\mathcal{F}$, where $f_j(\mathbf{x}) = f_{\theta_j}(\mathbf{x})$. Consider $T(\mathbf{x}^*) = (T_1(\mathbf{x}^*), \ldots, T_k(\mathbf{x}^*))$, where

$$T_j(\mathbf{x}^*) = \frac{\prod_{i=1}^{m} f_j(x_i^*)}{\prod_{i=1}^{m} f_0(x_i^*)}.$$

$T(\mathbf{x})$ can be recovered from the collection $\{T(\mathbf{x}^*)\}$ obtained by all possible resamples. Hence $\{T(\mathbf{x}^*)\}$ is sufficient for $\mathcal{F}$.

Proof Set $a_i = \dfrac{f_\theta(x_i)}{f_0(x_i)}, i = 1, \ldots, n$, suppressing the dependence on $\theta \in \{\theta_1, \ldots, \theta_k\}$. Then

$$\frac{\prod_{i=1}^n f_j(x_i)}{\prod_{i=1}^n f_0(x_i)} = a_1 \cdot \ldots \cdot a_n, \quad a_i > 0.$$

The product $a_1 a_2 \cdots a_n$ can be written as a product of terms of the form $a_{i_1} \cdots a_{i_m}$ as follows:

$$a_1 \cdots a_n = \left[\underbrace{(a_1 \cdots a_m)}_{\text{length } m} (a_2 \cdots a_m a_{m+1}) \cdots (a_{n-(m-1)} \cdots a_n) \cdots (a_n a_1 \cdots a_{m-1}) \right]^{(1/m)} .$$

Hence the theorem is proven.

Condition (2) holds for many parametric families by an application of Theorem 6.65 in Casella and Berger (2002, p.309): Gaussian ($k = 2$), Binomial ($k = 1$), Logistic ($k = n$).

2.1 Resampling Solution

For each t, solve for the least squares estimators $\tilde{\omega}_{0,s}, \ldots, \tilde{\omega}_{(m-1),s}$ in

$$\underbrace{\begin{bmatrix} f_1 & Df_1 & \cdots & D^{m-1}f_1 \\ \vdots & \vdots & \vdots & \vdots \\ f_r & Df_r & \cdots & D^{m-1}f_r \end{bmatrix}}_{\mathbf{X}_s} \underbrace{\begin{bmatrix} \tilde{\omega}_{0,s} \\ \vdots \\ \tilde{\omega}_{(m-1),s} \end{bmatrix}}_{\tilde{\boldsymbol{\omega}}_s} = \underbrace{\begin{bmatrix} -D^m f_1 \\ \vdots \\ -D^m f_r \end{bmatrix}}_{\mathbf{z}_s},$$

where $\{f_1, \ldots, f_r\} \subseteq \{x_1, \ldots, x_n\}, m \leq r \leq n$. Using notation in Ramsay (1996, Equation (5), p. 499), the least squares solution is

$$\tilde{\boldsymbol{\omega}}_s(t) = \left[\mathbf{X}_s^T(t)\mathbf{X}_s(t)\right]^{-1} \mathbf{X}_s^T(t)\mathbf{z}_s(t). \tag{3}$$

Note two special cases: (1) $\tilde{\boldsymbol{\omega}}_s(t) = \mathbf{X}_s^{-1}(t)\mathbf{z}_s(t)$, if $r = m$, and (2) the full problem $\tilde{\boldsymbol{\omega}}(t) = \left[\mathbf{X}^T(t)\mathbf{X}(t)\right]^{-1} \mathbf{X}^T(t)\mathbf{z}(t)$, where $\mathbf{X}(t)$ is $\mathbf{X}_s(t)$ with $r = n$.

A “tilde” on the coefficients is used to indicate that the solution is obtained under $\omega_m = 1$. Wu (1986), stated as Theorem 2.2 below, gives us a way to write Ramsay’s pointwise least squares estimates $\tilde{\boldsymbol{\omega}} = (\tilde{\omega}_0, \ldots, \tilde{\omega}_{m-1})^T$ as a weighted average of the coefficients $\tilde{\boldsymbol{\omega}}_{\mathrm{s}} = \left(\tilde{\omega}_{0,s}, \ldots, \tilde{\omega}_{(m-1),s}\right)^T$ obtained by resampling.

Theorem 2.2 stated below tells us how the resampling coefficients $\tilde{\boldsymbol{\omega}}_s(t)$ relate back to $\tilde{\boldsymbol{\omega}}(t)$. Notice that the weights used in the statement of the theorem to combine the $\tilde{\boldsymbol{\omega}}_s(t)$ sum to 1.

Theorem 2.2 (Wu 1986, p. 1267)**.** *The least squares solution $\tilde{\boldsymbol{\omega}}(t)$ of (3) satisfies*

$$\tilde{\boldsymbol{\omega}}(t) = \frac{\sum_r \det \left[\mathbf{X}_s^T(t)\mathbf{X}_s(t)\right] \tilde{\boldsymbol{\omega}}_s(t)}{\sum_r \det \left[\mathbf{X}_s^T(t)\mathbf{X}_s(t)\right]},$$

where $r \geq m$ and $\sum_r$ is the sum over all subsets of size r.

Theorem 2.2 motivates an estimator for the LDO given by equation (1) without the constraint $\omega_m(t) = 1$. Set $r = m$, then the solution is taken to be $\hat{\boldsymbol{\omega}}(t) = \sum_r \boldsymbol{\omega}_s(t)$, where

$$\begin{aligned} \boldsymbol{\omega}_s(t) &= \det\left[\mathbf{X}_s^T(t)\mathbf{X}_s(t)\right]\begin{bmatrix}\tilde{\boldsymbol{\omega}}_s(t)\\ 1\end{bmatrix} \\ &= \left(\det\left[\mathbf{X}_s(t)\right]\right)^2\begin{bmatrix}\mathbf{X}_s^{-1}(t)\mathbf{z}_s(t)\\ 1\end{bmatrix} \\ &= \begin{bmatrix}\det\left[\mathbf{X}_s(t)\right]\text{cofactors}\left[\mathbf{X}_s(t)\right]\mathbf{z}_s(t)\\ \det\left[\mathbf{X}_s^T(t)\mathbf{X}_s(t)\right]\end{bmatrix}. \end{aligned} \tag{4}$$

Note that dividing $\hat{\boldsymbol{\omega}}(t)$ through by by the leading coefficient $\hat{\omega}_m(t)$ recovers $\tilde{\boldsymbol{\omega}}(t)$ with $\tilde{\omega}_m(t) = 1$.

Another solution for $r = m$ instead uses $\boldsymbol{\omega}^\gamma(t) = \sum_r \boldsymbol{\omega}_s^\gamma(t)$, where

$$\begin{aligned} \boldsymbol{\omega}_s^\gamma(t) &= \left(\det\left[\mathbf{X}_s^T(t)\mathbf{X}_s(t)\right]\right)^\gamma\begin{bmatrix}\tilde{\boldsymbol{\omega}}_s(t)\\ 1\end{bmatrix} \\ &= \left(\det\left[\mathbf{X}_s(t)\right]\right)^{2\gamma}\begin{bmatrix}\mathbf{X}_s^{-1}(t)\mathbf{z}_s(t)\\ 1\end{bmatrix} \\ &= \begin{bmatrix}\det\left[\mathbf{X}_s(t)\right]^{2\gamma-1}\text{cofactors}\left[\mathbf{X}_s(t)\right]\mathbf{z}_s(t)\\ \left(\det\left[\mathbf{X}_s^T(t)\mathbf{X}_s(t)\right]\right)^\gamma\end{bmatrix}, \end{aligned} \tag{5}$$

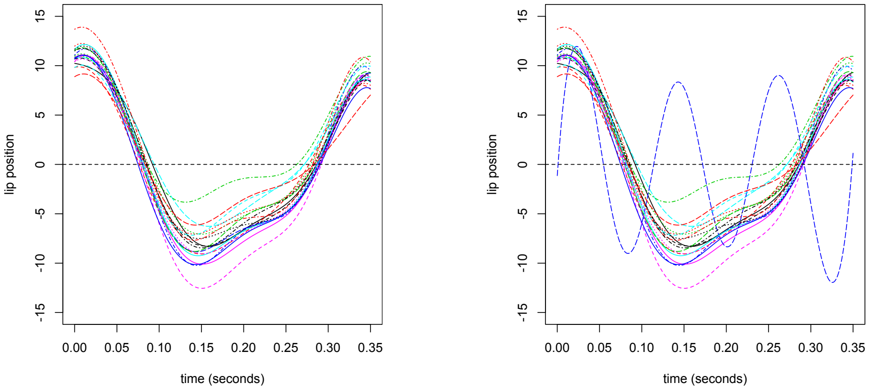
with $1/2 \leq \gamma \leq 1$. This is a variation of an estimator suggested by Wu (1984) to guard against outliers in the multiple regression problem.

The estimator $\boldsymbol{\omega}^\gamma(t)$ implemented here consists of the steps enumerated below.

1. Take m distinct curves $x_{i_1}, \ldots, x_{i_m}$ at a time to compute $K = \binom{n}{m}$ sets of coefficients $\omega_{0k}^\gamma, \omega_{1k}^\gamma, \ldots, \omega_{mk}^\gamma, k = k(i_1, \ldots, i_m) = 1, \ldots, K$ using equation (5).
2. Define the final estimators of the LDO coefficients $\omega_j^\gamma = \sum_{k=1}^K \omega_{jk}^\gamma, j = 0, \ldots, m$. Two cases were considered in the implementation:
   - $\gamma = 1$ to obtain the multiple of the least squares solution given by $\left(\sum_r \det\left[\mathbf{X}_s^T(t)\mathbf{X}_s(t)\right]\right)\tilde{\boldsymbol{\omega}}_s(t)$, and
   - $\gamma = 1/2$, a robust version of the latter solution.
3. Solve the differential equation $\hat{L}x(t) = \omega_0^\gamma x(t) + \omega_1^\gamma Dx(t) + \cdots + \omega_{m-1}^\gamma D^{m-1}x(t) + \omega_m^\gamma D^m x(t) = 0$ to find the basis functions for the null space of the LDO.
4. Use the basis functions to find low-dimensional representations of the curves.
5. Display fits to the data.

3. Test Case: Lip Data, $\mathbf{m = 2}$

The resampling method was applied to the lip data, and to the lip data modified by replacing one curve with an obvious outlier. In this context an outlier is any functional observation that is not in the span of the null space of the LDO. Figure 2 shows both sets of curves registered using the same marks.



(a) (b)

Figure 2: Left, the lip data; right, the lip data with an outlier introduced. Each set of curves is registered using the same marks.

Proposed Method	Ratio of sum of squared norms: PDA/proposed	
	Original lip data	with outlier
$\gamma = 1$	1.00	1.00
$\gamma = 1/2$	0.98	1.63

Table 1: Ratio of sum of squared norms of the forcing functions: PDA/proposed

Ramsay assumes that the lip data is best described by a second-order LDO ($m = 2$), and we adhere to that assumption. All computations were carried out in the R language and environment. We used code from the `lip.R` script provided in Ramsay's R library `fda`. The script used to reproduce Ramsay's results was `lip.R` with a few modifications to adjust to the latest version of the `fda` library.

First, we estimated the coefficients of the LDO. The ratio of the sum of squared norms $\sum_{i=1}^{n} \|Lx_i\|^2$ for Ramsay PDA (using $\lambda = 0$ and 146 knots so that we have 150 elements in the basis) to each of the resampling estimates of the LDO is shown in Table 1. The term contributed by the outlier curve was not included in the sum of squared norms. The PDA estimates of ω_0 and ω_1 were multiplied by ω_2^{γ} to compare the squared norms on the same scale. Focusing on row 1, the fact that the ratio of the sum of squared norms is 1 supports Theorem 2.2 that states that using $\gamma = 1$ with the proposed resampling method yields the pointwise least squares solution $\tilde{\boldsymbol{\omega}}$. Focusing on column 1, an analysis of the original lip data suggests there is a loss in efficiency when using a robust version of resampling ($\gamma = 1/2$). An analysis of the lip data with an outlier suggests that resampling with $\gamma = 1/2$ can lead to substantial gains in efficiency as compared to the current PDA methodology (ratio is 1.63). Simulation studies are needed to see the effect of more than one outlier on the resampling methods illustrated for this test case.

4. Conclusions

The resampling method eliminates the condition that $\omega_m = 1$ in the estimation of the coefficients of an LDO. When applied to the lip data and to the lip data with an outlier introduced, the resampling estimates yielded better sum of squared norms when compared to current PDA methodology. Work in progress is using the DAE solver `daspk` in the R library `deSolve` to obtain the basis for the null space by solving an implicit differential equation.

The resampling approach used in this method has value in other problems where subproblems have exact solutions. Combining $\binom{n}{m}$ subproblems to obtain a final estimator falls under the general heading of Divide-and-Conquer strategies commonly used in Computer Science (Cormen et al. 2010). The Divide-and-Conquer strategy implemented by resampling to solve for the coefficients of the LDO has the following properties:

1. There is no loss of information (Theorem 2.1), and
2. The number of computations when $m \ll n$, and n is large are reduced.

5. Appendix

Regression analysis is used to model and investigate relationships among variables. For example, the relationship between a dependent variable y and independent variable t may be studied through the model

$$y = f(t) + \epsilon, \quad t \in [a, b], \tag{6}$$

where ϵ is a random unobservable error with mean 0 and variance σ^2. Suppose we have data of the form (t_i, y_i) for $i = 1, \dots n$, satisfying equation (6).

In nonparametric regression, we only know that the true regression function $f(t)$ is smooth in t; its shape is estimated from the data, in contrast to parametric regression, in which we assume that the functional form of the regression curve $f(t)$ is known.

Consider the nonparametric estimator of the regression curve obtained as the minimizer of

$$\frac{1}{n}\sum_{i=1}^{n} [y_i - f(t_i)]^2 + \lambda \int_a^b \left[f^{(r)}(t)\right]^2 dt, \tag{7}$$

over f in the Sobolev space of order r denoted by $W_2^r[a, b]$. The Sobolev space of order r is the set of all functions with $r - 1$ continuous derivatives and the integral of the square of the r^{th} derivative exists. The first term of the minimizing criteria (7) is a residual sum of squares, as is used in ordinary least squares regression. The second term involves a penalty, where λ is a positive scalar called the smoothing parameter. The solution f_λ to this minimization problem is called a smoothing spline (Wahba 1990). A spline function $S(t)$ is a piecewise polynomial function defined on the entire real line, subject to a maximum number of continuity constraints.

Definition 5.1. *A* ***spline*** *of order $k \geq 2$ with knots $t_1 < t_2 < \cdots < t_n$ is any function $S(t)$ of the form*

$$S(t) = \sum_{i=0}^{k-1} \alpha_i t^i + \sum_{i=1}^{n} \delta_i (t - t_i)_+^{k-1} \tag{8}$$

for real $\alpha_0, \alpha_1, \cdots, \alpha_{k-1}, \delta_1, \cdots, \delta_n$ *satisfying that* $S(t)$ *has* $k-2$ *continuous derivatives.*

Definition 5.1 describes splines in terms of trucanted polynomial basis: $1, t, \ldots, t^{n-1}$, and

$$(t-t_i)_+ = \begin{cases} (t-t_i), & \text{if } t \geq t_i \\ 0, & \text{if } t < t_i \end{cases} \qquad i = 1, \ldots, n.$$

B-splines can also be used as a basis for computing splines. In particular, the smoothing spline estimator f_λ can be represented with B-splines. Numerical computation with B-splines is more stable than with the truncated power basis. B-splines are usually computed following the Cox-de Boor recursion formula (Eubank 1999). To initialize the recursion, we define an additional k knots and ”stack” them on the endpoints of the interval $[a, b]$ we are considering. Note that if $k = 2r$ is the order of the spline, we stack r knots on each endpoint of the interval. Given knots $t_1, \cdots, t_n$ on $[a, b]$, define $k = 2r$ additional knots as follows:

$$t_{-r+1} = \cdots = t_0 = a$$

$$b = t_{n+1} = \cdots = t_{n+r}.$$

Then the B-spline of order k is defined by the recursion

$$B_{i,1}(t) = \begin{cases} 1, & t_i \leq t < t_{i+1} \\ 0, & \text{otherwise} \end{cases} \qquad i = 0, \ldots, n$$

$$B_{i,k}(t) = \left(\frac{t - t_i}{t_{i+k-1} - t_i}\right) B_{i,k-1}(t) + \left(\frac{t_{i+k} - t}{t_{i+k} - t_{i+1}}\right) B_{i+1,k-1}(t)$$

for $i = -(k-1), \ldots, n$.

6. Acknowledgement

This work was partially supported by grant NIH-SCORE 5SC3GM094073.

References

Casella, A. and R. L. Berger (2002). *Statistical Inference*. Duxbury Advanced Series. Australia; Pacific Grove, California: Thomson Learning.

Coddington, A. and N. Levinson (1955). *Theory of Ordinary Differential Equations*. International series in pure and applied mathematics. New York: Tata McGraw-Hill.

Cormen, T.H et al. (2010). *Introduction of Algorithms*. Cambridge, MA: MIT Press.

Eubank, R.L. (1999). *Nonparametric Regression and Spline Smoothing, Second Edition*. Statistics: A Series of Textbooks and Monographs. New York: Taylor & Francis.

Frees, Edward W. (1991). “Trimmed slope estimates for simple linear regression.” In: *Journal of Statistical Planning and Inference* 27, pp. 203–221.

Jin, S., J.G. Staniswalis, and I. Mallawaarachchi (2013). “Principal Differential Analysis with a Continuous Covariate: Low Dimensional Approximations for Functional Data.” In: *Journal of Statistical Computation and Simulation* 83.10, pp. 1964–1980.

Ramsay, J. (1996). “Principal Differential Analysis: Data Reduction by Differential Operators”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.3, pp. 495–508.

Ramsay, J. and B.W. Silverman (1997). *Functional Data Analysis*. Springer Series in Statistics. New York: Springer.

Wahba, G. (1990). *Spline Models for Observational Data.* CBMS-NSF Regional Conference Series in Applied Mathematics, 59. Philadelphia, PA: SIAM.

Wu, C. F. J. (1984). “Jackknife and bootstrap inference in regression and a class of representations for the LSE”. In: *Wisconsin Univ-Madison, Mathematics Research Center* MSR-TSR-2675.

— (1986). “Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis”. In: *The Annals of Statistics* 14.4, pp. 1261–1295.