Functional data analysis in an operator-based mixed-model framework

BO MARKUSSEN

Department of Basic Sciences and Environment, University of Copenhagen, 1871 Frederiksberg C, Denmark. E-mail: bomar@life.ku.dk

Functional data analysis in a mixed-effects model framework is done using operator calculus. In this approach the functional parameters are treated as serially correlated effects giving an alternative to the penalized likelihood approach, where the functional parameters are treated as fixed effects. Operator approximations for the necessary matrix computations are proposed, and semi-explicit and numerically stable formulae of linear computational complexity are derived for likelihood analysis. The operator approach renders the usage of a functional basis unnecessary and clarifies the role of the boundary conditions.

Keywords: determinant approximation; Gaussian process; Green's function; random effect; serial correlation; operator approximation

1. Introduction

The aim of this paper is to derive operator approximations of the matrix computations used to estimate the fixed and the random effects in a mixed-effects model, where M samples $y_1, \ldots, y_M \in \mathbb{R}^N$ of temporal curves have been observed at N predefined time points $t_1, \ldots, t_N$. The main technical contribution of this paper, making it practically possible to solve the estimation problem as a functional estimation problem, is that the proposed operator approximations have linear computational complexity in the sample length N. Consequently, the mixed-effects inference becomes feasible in the realm of functional data analysis, where N can be large.

Concatenating the samples $y_m = \{y_{mn}\}_{n=1,\ldots,N} \in \mathbb{R}^N$ into an observation vector $y = \{y_m\}_{m=1,\ldots,M} \in \mathbb{R}^{N_{\text{total}}}$ with dimension $N_{\text{total}} = N * M$ the statistical model we use is given by

$$y = \Gamma\beta + Zu + x + \varepsilon. \tag{1}$$

In this linear mixed-effects model the design matrices $\Gamma \in \mathbb{R}^{N_{\text{total}} \times p}$ and $Z \in \mathbb{R}^{N_{\text{total}} \times q}$ are known and assumed to have full ranks p and q, respectively, and the fixed effects $\beta \in \mathbb{R}^p$ and the random effects $u \sim \mathcal{N}_q(0, \sigma^2 G)$ may be shared by the M samples. The random component $x = \{x_m\}_{m=1,\ldots,M} \sim \mathcal{N}_{N_{\text{total}}}(0, \sigma^2 R)$ is partitioned in the same way as the observation vector y and consists of discretized readings $x_m = \{x_m^{\text{fct}}(t_n)\}_{n=1,\ldots,N} \in \mathbb{R}^N$ of unobserved (latent) random functions $x_m^{\text{fct}} : [a, b] \to \mathbb{R}$. We assume that the random functions $x_1^{\text{fct}}, \ldots, x_M^{\text{fct}}$ are independent and identically distributed Gaussian processes with zero mean. The covariance matrix $\sigma^2 R$ will be specified below appealing to the smoothing splines methodology often used in functional data analysis. Due to the i.i.d. assumption there exists a covariance matrix $R_0 \in \mathbb{R}^{N \times N}$ such that

$R = R_0 \otimes \mathbb{I}_M$, where $\otimes$ is the Kronecker tensor product, and $\mathbb{I}_M \in \mathbb{R}^{M \times M}$ is the identity matrix of dimension M. The last component in the mixed-effects model is the measurement noise $\varepsilon \sim \mathcal{N}_{N_{\text{total}}}(0, \sigma^2 \mathbb{I}_{N_{\text{total}}})$.

Our objective is to derive computationally efficient formulae for the maximum likelihood estimate of the fixed effects β, the best linear unbiased predictions of the random effects $u \in \mathbb{R}^q$ and of the latent random functions $x_m^{\text{fct}} : [a, b] \to \mathbb{R}$, and for the restricted likelihood function. The latter allow for restricted likelihood inference on the variance parameters $\sigma^2 > 0$, $G \in \mathbb{R}^{q \times q}$ and $R_0 \in \mathbb{R}^{N \times N}$. The methodology presented in this paper has two notable differences as compared to the penalized likelihood approach to functional data analysis; see, for example, the books by Ramsay and Silverman [11,12]. From the viewpoint of computations we devise methods that work directly on the data vector y and, for example, provide predictions $\partial_t^\mu \mathrm{E}[x_m^{\text{fct}}(t)|y]$ of the temporal derivatives of the latent functional parameters. In particular, there is no basis representation of the functional object $\mathrm{E}[x_m^{\text{fct}}|y]$. This is by contrast with the standard technology used in functional data analysis, where functional parameters are given a finite dimensional representation, for example, in a spline basis, and the sparseness of the associated covariance matrices is invoked to achieve feasible computations. As an alternative to this we use analytically tractable operator approximations of the matrix equations. From the viewpoint of statistical modeling we model the functional parameters x_m^{fct} as random effects. Whether this is preferable over the fixed effect interpretation underlying the penalized likelihood depends on the particular application at hand. The distinction between random and fixed effects is here the same as for classical mixed-effects models; see [13] for a thorough discussion of the issue and [6] for a comparison of the associated inference methodologies.

In the simplified version $y = x + \varepsilon$ of model equation (1), the sample y_m may be understood as a noisy observation of the function $x_m^{\text{fct}} : [a, b] \to \mathbb{R}$ taken at the sample points $t_1, \ldots, t_N$. In the penalized likelihood approach to functional data analysis the functional parameters x_m^{fct} are treated as fixed effects. The penalized negative log likelihood is given by

$$N_{\text{total}} \log \sigma + \frac{1}{2\sigma^2} \sum_{m=1}^{M} \left(\sum_{n=1}^{N} |y_{mn} - x_m^{\text{fct}}(t_n)|^2 + \lambda \int_a^b |\mathscr{K} x_m^{\text{fct}}(t)|^2 \, \mathrm{d}t \right), \tag{2}$$

where $\mathscr{K}$ is a differential operator of some order k measuring the roughness of a function $\theta \in \mathcal{C}^k([a, b]; \mathbb{R})$. The so-called smoothing parameter $\lambda > 0$ quantifies the trade-off between a close fit of the observations and the roughness of the functional parameters. Since the space of functions is infinite-dimensional, such a trade-off is required to avoid overfitting of the finite number of data points.

In this paper we avoid the curse of dimensionality by providing the theoretical solution in the function space before plugging in the observed grid readings to compute the solution. This is done using the operator $\mathscr{L} = \mathscr{K}^\dagger \mathscr{K}$, which is of order $2k$ and defined on $\mathcal{C}^{2k}([a, b]; \mathbb{R})$. To ensure positive definiteness of $\mathscr{L}$ we impose boundary conditions. Let $a_i, b_i \in \{i - 1, 2k - i\}$ for $i = 1, \ldots, k$ be fixed, and let the function space $\mathcal{H}$ be defined by

$$\mathcal{H} = \left\{ \theta \in \mathcal{C}^{2k}([a, b]; \mathbb{R}) | \theta^{(a_i)}(a) = \theta^{(b_i)}(b) = 0 \text{ for } i = 1, \ldots, k \right\}, \tag{3}$$

where $\theta^{(i)}$ denotes the ith order derivative of θ. Applying integration by parts k times the penalty terms in equation (2) may be rewritten via

$$\int_a^b |\mathscr{K}\theta(t)|^2 \, dt = \int_a^b \theta(t)\mathscr{L}\theta(t) \, dt, \qquad \theta \in \mathcal{H}.$$

This identity also implies that $\mathscr{L}$ is a positive semidefinite operator on $\mathcal{H}$. A condition ensuring $\mathscr{L}$ to be invertible is given in Section 3.1. In the affirmative case the inverse operator is given by a so-called Green's function $\mathcal{G}(t, s)$ via $\mathscr{L}^{-1}f(t) = \int_a^b \mathcal{G}(t, s)f(s) \, ds$. Since $\mathscr{L}$ is positive definite it follows that $\mathcal{G}(t, s)$ is positive definite. In particular, the matrix defined by

$$R_0 = \{\mathcal{G}(t_n, t_m)\}_{n,m=1,\ldots,N} \in \mathbb{R}^{N \times N}$$

is positive definite and may be used as the variance of the serially correlated effects x_m. This specification establishes a link between the covariance matrix $\sigma^2 R_0$ of the discretized readings x_m in the model equation (1) and the penalized likelihood equation (2).

The proposed methodology can be slightly generalized taking $\mathscr{L}$ as the sum of squares $\sum_{l=1}^L \mathscr{K}_l^\dagger \mathscr{K}_l$ of operators measuring different aspects of roughness. The operator $\mathscr{L}$ may be interpreted as a precision and used in the parameterization of a statistical model. This is by contrast with standard software for mixed-effects models such as the nlme-package [10] in R or the MIXED procedure in SAS, where the parameterization is done in terms of variances. In [14] a similar approach was taken for the analysis of longitudinal data, and further references may be found in [7], Chapter 8.4.

The remainder of this paper is organized as follows. Section 2 reviews inference techniques for the model equation (1). In particular, we present the matrix formulae that will be approximated by their operator equivalents. Section 3 provides the mathematical contributions of the paper. In this section the operator approximation is introduced and refined for the case of equidistant observations, that is, $t_n = a + \frac{2n-1}{N}(b - a)$. In particular, we derive semi-explicit and numerically stable formulae for the needed computations in the case of equidistant observations. In Section 4 the operator approximation is applied on the matrix formulae from Section 2. This leads to concrete algorithms that have been implemented in an R-package named fdaMixed [8].

2. Inference in the mixed-effects model

This section reviews estimation and inference techniques for the model equation (1). Since the derivations of the matrix formulae stated below are standard (see, e.g., [1–3,13]), no proofs will be given. The dimensions are given by

$$y = \Gamma\beta + Zu + x + \varepsilon \in \mathbb{R}^{N_{\text{total}}}, \qquad \beta \in \mathbb{R}^p, u \in \mathbb{R}^q, x \in \mathbb{R}^{N_{\text{total}}}, \varepsilon \in \mathbb{R}^{N_{\text{total}}},$$

where $N_{\text{total}} = N * M$. Based on the covariance matrices G and $R = R_0 \otimes \mathbb{I}_M$ we define the matrices $A_0 = \mathbb{I}_N + R_0$, $A = \mathbb{I}_{N_{\text{total}}} + R = A_0 \otimes \mathbb{I}_M$, and

$$C_u = (G^{-1} + Z^\top A^{-1} Z)^{-1}, \qquad C_r = A^{-1} - A^{-1} Z C_u Z^\top A^{-1}, \qquad C_\beta = (\Gamma^\top C_r \Gamma)^{-1}.$$

The matrix formulae will be stated such that for moderately sized p and q the computational obstacle of their practical implementation lies in the initialization and inversion of the N-dimensional matrix A_0. The circumvention of this obstacle is the topic of Section 3.

For known variance parameters σ^2, G, R_0, the best unbiased estimate for the fixed effects is given by the maximum likelihood estimate

$$\hat{\beta} = C_\beta \Gamma^\top C_r y = C_\beta \Gamma^\top (A^{-1} y - A^{-1} Z C_u Z^\top A^{-1} y). \tag{4}$$

The best linear unbiased predictions (BLUPs) for the random effects u and the serially correlated effects $x = \{x_m\}_{m=1,\dots,M}$ are given by the conditional means

$$\mathrm{E}[u|y] = C_u Z^\top A^{-1}(y - \Gamma\hat{\beta}), \qquad \mathrm{E}[x|y] = R A^{-1}(y - \Gamma\hat{\beta} - Z\mathrm{E}[u|y]). \tag{5}$$

It is generally agreed (see, e.g., [1] and [7], Chapter 5.3) that the variance parameters may be estimated as the maximizers of the restricted likelihood. One of the factors in the likelihood is the determinant of $A_0 = \mathbb{I}_N + R_0$. To derive the operator approximation of this factor we use the representation

$$\log \det A_0 = \int_0^1 \partial_v \log \det(\mathbb{I} + vR_0)\, dv = \int_0^1 \sum_{j=1}^N e_j^\top (v\mathbb{I}_N + R_0^{-1})^{-1} e_j\, dv, \tag{6}$$

where the vectors $e_j = \{1_{j=n}\}_{n=1,\dots,N} \in \mathbb{R}^N$ for $j = 1, \dots, N$ constitute an orthonormal basis for $\mathbb{R}^N$. Using this representation and introducing the conditional residuals $r = y - \Gamma\hat{\beta} - Z\mathrm{E}[u|y] - \mathrm{E}[x|y]$, the double negative log restricted likelihood is given by

$$(2N_{\mathrm{total}} - 2p)\log\sigma + M \int_0^1 \sum_{j=1}^N e_j^\top (v\mathbb{I}_N + R_0^{-1})^{-1} e_j\, dv$$

$$+ \log\det(\mathbb{I}_q + Z^\top A^{-1} Z G) - \log\det C_\beta \tag{7}$$

$$+ \sigma^{-2}(r^\top r + \mathrm{E}[u|y]^\top G^{-1}\mathrm{E}[u|y] + \mathrm{E}[x|y]^\top R^{-1}\mathrm{E}[x|y]),$$

where it should be kept in mind that C_β, r, $\mathrm{E}[u|y]$, $\mathrm{E}[x|y]$ depend on G and R_0. The profile estimate for the error variance σ^2 has an explicit form,

$$\hat{\sigma}^2 = \frac{1}{N_{\mathrm{total}} - p}(r^\top r + \mathrm{E}[u|y]^\top G^{-1}\mathrm{E}[u|y] + \mathrm{E}[x|y]^\top R^{-1}\mathrm{E}[x|y]).$$

We conclude this section by reviewing some theoretical results on the inference techniques described above. The errors $\hat{\beta} - \beta$, $\mathrm{E}[u|y] - u$, $\mathrm{E}[x|y] - x$ follow a joint Gaussian distribution, and their joint covariance may be derived using [2], Section 2.4. Kackar and Harville [4] show that if the estimators for the variance parameters are translation-invariant and even functions of y, then $\hat{\beta}$, $\mathrm{E}[u|y]$, $\mathrm{E}[x|y]$ remain unbiased when the estimates are inserted in place of the unknown variance parameters. As explained by Welham and Thompson [16] inference on β may be done as χ^2-tests on twice the log ratio between the maximum restricted likelihoods, where the design

matrix under the null hypothesis is used in the definition of the restricted likelihood under the model. Simulation studies done by Morrell [9] suggest that inference on the variance parameters may be done as χ^2-tests on twice the log ratio between the maximum restricted likelihoods, but here the formal asymptotic theory appears to be less developed.

3. Functional embedding of discrete data

Functional data consist of observations of continuous curves at discrete sample points. As an alternative to computations based on spline representations and sparse matrix computations we embed the discrete observations into the continuous setting and approximate the matrix computations by their operator counterparts. In order to describe this operator approximation we first introduce some notation.

By a discretization of size N of the time interval $[a, b]$ we mean a set of points $\mathcal{T} = \{t_1, \ldots, t_N\}$ with $a < t_1 < \cdots < t_N < b$. Such a discretization is said to be equidistant if $t_n = a + \frac{2n-1}{2N}(b-a)$, and in that case we associate the mesh length given by $\Delta = (b - a)/N$. To ease notation we implicitly adjoin the points $t_0 = a$ and $t_{N+1} = b$ to any discretization of size N.

Given a vector $z = \{z_n\}_{n=1,\ldots,N} \in \mathbb{R}^N$ we denote by $\mathscr{E}_z$ the piecewise linear embedding of $z \in \mathbb{R}^N$ into $\mathcal{C}([a, b]; \mathbb{R})$, that is, the function that is linear on the segments $[t_n, t_{n+1}]$ for $n = 0, \ldots, N$ with $\mathscr{E}_z(a) = z_1$, $\mathscr{E}_z(b) = z_N$ and $\mathscr{E}_z(t_n) = z_n$ for $n = 1, \ldots, N$. We also introduce the multiplication operator $\mathscr{M}_\mathcal{T}$ on $\mathcal{C}([a, b]; \mathbb{R})$ defined by

$$\mathscr{M}_\mathcal{T} f(t) = \mathscr{E}_\mu(t) f(t), \qquad f \in \mathcal{C}([a, b]; \mathbb{R}),$$

where $\mu = \{\mu_n\}_{n=1,\ldots,N} \in \mathbb{R}^N$ is given from the discretization $\mathcal{T}$ via

$$\mu_n = \begin{cases} 2/(t_2 + t_1 - 2a), & \text{for } n = 1, \\ 2/(t_{n+1} - t_{n-1}), & \text{for } n = 2, \ldots, N-1, \\ 2/(2b - t_N - t_{N-1}), & \text{for } n = N. \end{cases} \tag{8}$$

In particular, if $\mathcal{T}$ is equidistant, then $\mathscr{M}_\mathcal{T} = \Delta^{-1}\mathbb{I}$.

Proposition 1. *Let a discretization $\mathcal{T}$ of the interval $[a, b]$, $t \in [a, b]$ and $\mathcal{G} \in \mathcal{C}([a, b] \times [a, b]; \mathbb{R})$ be given. Assume that $\mathcal{G}(t, \cdot)$ is twice differentiable on the segments $[t_n, t_{n+1}]$ with continuous derivatives $\mathcal{G}^{(i)}(t, \cdot)$. For $z \in \mathbb{R}^N$ there exists $\xi_n \in (t_n, t_{n+1})$ for $n = 0, \ldots, N$ and $\zeta_1 \in (a, t_1)$, $\zeta_N \in (t_N, b)$ such that $\sum_{n=1}^N \mathcal{G}(t, t_n)z_n - \int_a^b \mathcal{G}(t, s)\mathscr{E}_\mu(s)\mathscr{E}_z(s)\,ds$ equals*

$$\frac{(t_1 - a)^2}{2}\mathcal{G}^{(1)}(t, \zeta_1)\mu_1 z_1 - \frac{(b - t_N)^2}{2}\mathcal{G}^{(1)}(t, \zeta_N)\mu_N z_N$$

$$+ \frac{1}{12}\sum_{n=0}^N (t_{n+1} - t_n)^3 \bigg(\mathcal{G}^{(2)}(t, \xi_n)\mathscr{E}_\mu(\xi_n)\mathscr{E}_z(\xi_n) + 2\mathcal{G}^{(1)}(t, \xi_n)\frac{\mu_{n+1} - \mu_n}{t_{n+1} - t_n}\mathscr{E}_z(\xi_n) \tag{9}$$

$$+ 2\mathcal{G}^{(1)}(t, \xi_n)\mathscr{E}_\mu(\xi_n)\frac{z_{n+1} - z_n}{t_{n+1} - t_n} + 2\mathcal{G}(t, \xi_n)\frac{\mu_{n+1} - \mu_n}{t_{n+1} - t_n}\frac{z_{n+1} - z_n}{t_{n+1} - t_n}\bigg),$$

where μ is given by equation (8).

Proof. The *trapezoidal rule* of integration [5], Section 7.2, gives intermediate points $\xi_n \in (t_n, t_{n+1})$ such that $\int_a^b \mathcal{G}(t,s)\mathscr{E}_\mu(s)\mathscr{E}_z(s)\,\mathrm{d}s$ equals

$$\frac{t_1 - a}{2}\mathcal{G}(t,a)\mathscr{E}_\mu(a)\mathscr{E}_z(a) + \sum_{n=1}^{N}\frac{t_{n+1}-t_{n-1}}{2}\mathcal{G}(t,t_n)\mathscr{E}_\mu(t_n)\mathscr{E}_z(t_n)$$

$$+ \frac{b-t_N}{2}\mathcal{G}(t,b)\mathscr{E}_\mu(b)\mathscr{E}_z(b) - \frac{1}{12}\sum_{n=0}^{N}(t_{n+1}-t_n)^3(\mathcal{G}(t,\cdot)\mathscr{E}_\mu\mathscr{E}_z)^{(2)}(\xi_n).$$

The result follows inserting the piecewise linear functions $\mathscr{E}_\mu$ and $\mathscr{E}_z$, the first-order Taylor expansions at some intermidiate points $\zeta_1 \in (a,t_1)$, $\zeta_N \in (t_N, b)$,

$$\frac{t_1-a}{2}\mathcal{G}(t,a)\mathscr{E}_\mu(a)\mathscr{E}_z(a) = \frac{t_1-a}{2}\mathcal{G}(t,t_1)\mu_1 z_1 - \frac{(t_1-a)^2}{2}\mathcal{G}^{(1)}(t,\zeta_1)\mu_1 z_1,$$

$$\frac{b-t_N}{2}\mathcal{G}(t,b)\mathscr{E}_\mu(b)\mathscr{E}_z(b) = \frac{b-t_N}{2}\mathcal{G}(t,t_N)\mu_N z_N + \frac{(b-t_N)^2}{2}\mathcal{G}^{(1)}(t,\zeta_N)\mu_N z_N,$$

by expanding the second-order derivative and by rearranging the terms. $\square$

Corollary 1. *If the discretization $\mathcal{T}$ is equidistant, then there exists $\widetilde{\xi}_n \in (\xi_{n-1}, \xi_n) \subset (t_{n-1}, t_{n+1})$ for $n = 1, \ldots, N$ such that the approximation error equation* (9) *equals*

$$\frac{b-a}{12N}\sum_{n=1}^{N}\big(t_{n+1}-t_{n-1}-3(\xi_n-\xi_{n-1})\big)\mathcal{G}^{(2)}(t,t_n)z_n$$

$$+ \frac{b-a}{12N}\sum_{n=1}^{N}(t_{n+1}-\xi_n)\big(\mathcal{G}^{(2)}(t,\xi_n)-\mathcal{G}^{(2)}(t,t_n)\big)z_n$$

$$- \frac{b-a}{6N}\sum_{n=1}^{N}(\xi_n-\xi_{n-1})\big(\mathcal{G}^{(2)}(t,\widetilde{\xi}_n)-\mathcal{G}^{(2)}(t,t_n)\big)z_n \tag{10}$$

$$+ \frac{b-a}{12N}\sum_{n=1}^{N}(\xi_{n-1}-t_{n-1})\big(\mathcal{G}^{(2)}(t,\xi_{n-1})-\mathcal{G}^{(2)}(t,t_n)\big)z_n$$

$$+ \frac{(b-a)^2}{8N^2}\mathcal{G}^{(1)}(t,\zeta_1)\mu_1 z_1 - \frac{(b-a)^2}{8N^2}\mathcal{G}^{(1)}(t,\zeta_N)\mu_N z_N.$$

Proof. Equidistant spacing implies that the factors $\mu_n = N/(b-a)$ defined in equation (8) are constant, and the approximation error equation (9) reduces to

$$\frac{(b-a)^2}{8N^2}\mathcal{G}^{(1)}(t,\zeta_1)\mu_1 z_1 - \frac{(b-a)^2}{8N^2}\mathcal{G}^{(1)}(t,\zeta_N)\mu_N z_N$$

$$+ \sum_{n=0}^{N}\frac{b-a}{12N}\left(\frac{b-a}{N}\mathcal{G}^{(2)}(t,\xi_n)\mathscr{E}_z(\xi_n)+2\mathcal{G}^{(1)}(t,\xi_n)(z_{n+1}-z_n)\right).$$

The last sum equals

$$\sum_{n=0}^{N} \frac{b-a}{12N} \mathcal{G}^{(2)}(t, \xi_n)\big((\xi_n - t_n)z_{n+1} + (t_{n+1} - \xi_n)z_n\big)$$

$$+ \sum_{n=1}^{N} \frac{b-a}{6N} \big(\mathcal{G}^{(1)}(t, \xi_{n-1}) - \mathcal{G}^{(1)}(t, \xi_n)\big) z_n.$$

By Taylor's theorem there exists $\widetilde{\xi}_n \in (\xi_{n-1}, \xi_n)$ such that this equals

$$\sum_{n=1}^{N} \frac{b-a}{12N} \big((t_{n+1} - \xi_n)\mathcal{G}^{(2)}(t, \xi_n) + (\xi_{n-1} - t_{n-1})\mathcal{G}^{(2)}(t, \xi_{n-1})$$

$$- 2(\xi_n - \xi_{n-1})\mathcal{G}^{(2)}(t, \widetilde{\xi}_n)\big) z_n.$$

The corollary follows centering the terms $\mathcal{G}^{(2)}(t, \cdot)$ around $\mathcal{G}^{(2)}(t, t_n)$. □

If the matrix $D \in \mathbb{R}^{N \times N}$ and the integral operator $\mathcal{G}$ on $\mathcal{C}([a, b]; \mathbb{R})$ are defined by $D = \{\mathcal{G}(t_n, t_m)\}_{n,m=1,...,N}$ and $\mathcal{G}f(t) = \int_a^b \mathcal{G}(t, s)f(s)\,ds$, then the preceding results suggest the approximation

$$Dz \approx \{\mathcal{G}\mathcal{M}_{\mathcal{T}}\mathcal{E}_z(t_n)\}_{n=1,...,N} \in \mathbb{R}^N, \qquad z \in \mathbb{R}^N. \tag{11}$$

Green's functions usually possess sufficient smoothness for Proposition 1 to apply (see, e.g., [15]), and hence the approximation error in equation (11) vanishes as $\max_{n=0,...,N} |t_{n+1} - t_n|$ goes to zero. In case of equidistant discretizations this property is refined in Corollary 1. The first term in equation (10) is of size $O(N^{-1})$ and the other terms are of size $O(N^{-2})$. Perhaps the first term can be used to derive and correct a bias arising from the proposed operator approximation, but we will leave this to be studied in future work.

3.1. Explicit operator computations

To motivate the derivations done in this section we may consider the model equation (1) without the fixed and the random effects, that is, $y = x + \varepsilon$. In this case equation (11) implies the approximation of the prediction equation (5) of the mth serially correlated effect given by

$$\mathrm{E}[x_m|y] = R_0 A_0^{-1} y_m = (\mathbb{I}_N + R_0^{-1})^{-1} y_m \approx \{(\mathbb{I} + \mathcal{M}_{\mathcal{T}}^{-1}\mathcal{L})^{-1}\mathcal{E}_{y_m}(t_n)\}_{n=1,...,N}$$

and the approximation of the logarithmic determinant equation (6) given by

$$\log \det A_0 = \int_0^1 \sum_{j=1}^{N} e_j^\top (v\mathbb{I}_N + R_0^{-1})^{-1} e_j \, dv$$

$$\approx \int_0^1 \sum_{j=1}^{N} (v\mathbb{I} + \mathcal{M}_{\mathcal{T}}^{-1}\mathcal{L})^{-1}\mathcal{E}_{e_j}(t_j) \, dv \approx \int_0^1 \int_a^b \mathcal{G}_v(t, t) \, dt \, dv, \tag{12}$$

where $\mathcal{G}_v$ is the Green's function for $v\mathbb{I} + \mathcal{M}_{\mathcal{T}}^{-1}\mathcal{L}$. As shown in Section 4 the matrix formulae used for inference in the full mixed-effects model equation (1) may be similarly approximated. In order to develop our computational methodology we derive semi-explicit and numerically stable inversion formulae for differential operators of the type $\mathcal{L}_* = \mathbb{I} + \mathcal{M}_{\mathcal{T}}^{-1}\mathcal{L}$. If the discretization $\mathcal{T}$ is equidistant with mesh length Δ, and the differential operator $\mathcal{L}$ has constant coefficients, then $\mathcal{L}_* = \mathbb{I} + \Delta\mathcal{L}$ may be inverted using Theorem 1 stated below. Boundary conditions play an essential role in this theorem, and the reader may want to refresh the definition of the space $\mathcal{H}$ given in equation (3).

Theorem 1. *Consider a differential operator $\mathcal{L}_*$ on $\mathcal{H}$ given by*

$$\mathcal{L}_*\theta(t) = \alpha_{2k}\theta^{(2k)}(t) + \alpha_{2k-1}\theta^{(2k-1)}(t) + \cdots + \alpha_1\theta^{(1)}(t) + \alpha_0\theta(t) \tag{13}$$

with $\alpha_{2k} \neq 0$. Let $J = \mathrm{diag}(J_1, \ldots, J_p) \in \mathbb{C}^{2k \times 2k}$, with $J_j \in \mathbb{C}^{k_j \times k_j}$, be the Jordan canonical form of the companion matrix

$$C = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \\ -\dfrac{\alpha_0}{\alpha_{2k}} & -\dfrac{\alpha_1}{\alpha_{2k}} & \cdots & \cdots & -\dfrac{\alpha_{2k-1}}{\alpha_{2k}} \end{pmatrix} \in \mathbb{R}^{2k \times 2k}. \tag{14}$$

Let $M \in \mathbb{C}^{2k \times 2k}$ be a non-trivial solution of the matrix equation $CM = MJ$, and let $M_{1j} \in \mathbb{C}^{1 \times k_j}$ be the decomposition of the first row of M along the Jordan blocks J_j. Let $v_1 = (1 \ \cdots \ 1) \in \mathbb{R}^{1 \times k}$, $v_2 = (0 \ \cdots \ 0 \ 1)^\top \in \mathbb{R}^{2k \times 1}$, and let F_a, $F_b \in \mathbb{R}^{k \times 2k}$ be given by

$$F_a = \{1_{j-1=a_i}\}_{\substack{i=1,\ldots,k \\ j=1,\ldots,2k}}, \qquad F_b = \{1_{j-1=b_i}\}_{\substack{i=1,\ldots,k \\ j=1,\ldots,2k}}.$$

Let $\bar{v}_1 = (v_1 \ v_1) \in \mathbb{R}^{1 \times 2k}$, and let $\bar{F}_a$, $\bar{F}_b$, $W \in \mathbb{R}^{2k \times 2k}$ be defined by

$$\bar{F}_a = \begin{pmatrix} F_a \\ 0_{k \times 2k} \end{pmatrix}, \qquad \bar{F}_b = \begin{pmatrix} 0_{k \times 2k} \\ F_b \end{pmatrix}, \qquad W = \begin{pmatrix} M_{11} & \cdots & M_{1p} \\ M_{11}J_1 & \cdots & M_{1p}J_p \\ \vdots & & \vdots \\ M_{11}J_1^{2k-1} & \cdots & M_{1p}J_p^{2k-1} \end{pmatrix}.$$

If the matrix $H = \bar{F}_a W \exp(aJ) + \bar{F}_b W \exp(bJ)$ is invertible, then $\mathcal{L}_*$ is invertible. In the affirmative case the inverse operator is an integral operator $\mathcal{L}_*^{-1}f(t) = \int_a^b \mathcal{G}_*(t,s)f(s)\,\mathrm{d}s$, where the Green's function $\mathcal{G}_*$ is given by

$$\mathcal{G}_*(t,s) = \begin{cases} \alpha_{2k}^{-1}\bar{v}_1\exp(tJ)H^{-1}\bar{F}_a W\exp\big((a-s)J\big)W^{-1}v_2, & \text{for } s \leq t, \\ -\alpha_{2k}^{-1}\bar{v}_1\exp(tJ)H^{-1}\bar{F}_b W\exp\big((b-s)J\big)W^{-1}v_2, & \text{for } t \leq s. \end{cases} \tag{15}$$

Proof. The proof follows specializing and condensing [15], Theorem 3. The signs of [15], equation (3.15), equation (3.24), should be changed due to a mistake of sign in [15], equation (3.9).

We allow for leading coefficient $\alpha_{2k} \neq 1$ and have interchanged the indices k and p to align with the notation used in the present paper. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Formula (15) is explicit and most satisfactory from a theoretical point of view. But from a practical point of view the formula can be numerically unstable since the exponentials $\exp(tJ)$, $\exp((a-s)J)$ and $\exp((b-s)J)$ are weighted against similar exponentials in the definition of the matrix H. Imposing symmetry of the Jordan matrix it is, however, possible to remove the potential numerical instabilities.

Proposition 2. *Suppose that the characteristic polynomial*

$$\alpha_{2k}z^{2k} + \alpha_{2k-1}z^{2k-1} + \cdots + \alpha_1 z + \alpha_0 = 0 \tag{16}$$

for the differential operator (13) has $2k$ distinct roots $\eta_1^-, \eta_1^+, \ldots, \eta_k^-, \eta_k^+ \in \mathbb{C}$ such that the real values of the k eigenvalues $\eta_1^-, \ldots, \eta_k^-$ are non-positive and the real values of the k eigenvalues $\eta_1^+, \ldots, \eta_k^+$ are non-negative. Then the Jordan canonical form of the companion matrix equation (14) is diagonal with block diagonals consisting of eigenvalues with non-positive and non-negative real values, respectively,

$$J = \begin{pmatrix} J_- & 0_{k\times k} \\ 0_{k\times k} & J_+ \end{pmatrix}, \qquad J_- = \mathrm{diag}(\eta_1^-, \ldots, \eta_k^-), \qquad J_+ = \mathrm{diag}(\eta_1^+, \ldots, \eta_k^+),$$

and the matrix $W = \binom{W_-}{W_+} \in \mathbb{C}^{2k\times 2k}$ may be decomposed via $W_-, W_+ \in \mathbb{C}^{2k\times k}$ defined by

$$W_- = \begin{pmatrix} 1 & \cdots & 1 \\ \eta_1^- & \cdots & \eta_k^- \\ \vdots & \ddots & \vdots \\ (\eta_1^-)^{2k-1} & \cdots & (\eta_k^-)^{2k-1} \end{pmatrix}, \qquad W_+ = \begin{pmatrix} 1 & \cdots & 1 \\ \eta_1^+ & \cdots & \eta_k^+ \\ \vdots & \ddots & \vdots \\ (\eta_1^+)^{2k-1} & \cdots & (\eta_k^+)^{2k-1} \end{pmatrix}.$$

Furthermore, define $v_1 = (1 \;\; \cdots \;\; 1) \in \mathbb{R}^{1\times k}$, $v_2 = (0 \;\; \cdots \;\; 0 \;\; 1)^\top \in \mathbb{R}^{2k\times 1}$, $v_-, v_+ \in \mathbb{R}^{k\times 1}$ via $W^{-1}v_2 = \binom{v_-}{v_+}$, and the vectors $\phi_\mu(t), \psi_\mu(t) \in \mathbb{R}^{1\times k}$ for $t \in [a,b]$ and $\mu \in \mathbb{N}_0$ by

$$\phi_\mu(t) = \alpha_{2k}^{-1}\big(v_1 J_-^\mu - v_1 J_+^\mu e^{-(b-t)J_+}(F_b W_+)^{-1}F_b W_- e^{(b-t)J_-}\big)$$

$$\cdot \big(I_{k\times k} - e^{(t-a)J_-}(F_a W_-)^{-1}F_a W_+ e^{-(b-a)J_+}(F_b W_+)^{-1}F_b W_- e^{(b-t)J_-}\big)^{-1}$$

and

$$\psi_\mu(t) = \alpha_{2k}^{-1}\big(v_1 J_+^\mu - v_1 J_-^\mu e^{(t-a)J_-}(F_a W_-)^{-1}F_a W_+ e^{-(t-a)J_+}\big)$$

$$\cdot \big(I_{k\times k} - e^{-(b-t)J_+}(F_b W_+)^{-1}F_b W_- e^{(b-a)J_-}(F_a W_-)^{-1}F_a W_+ e^{-(t-a)J_+}\big)^{-1}.$$

Then the μth partial derivative $\partial_t^\mu \mathcal{G}_*(t,s)$ of the Green's function defined in equation (15) may be rewritten as the numerically stable expression

$$\begin{cases} \phi_\mu(t)e^{(t-s)J_-}\big(v_- + e^{(s-a)J_-}(F_a W_-)^{-1}F_a W_+ e^{-(s-a)J_+}v_+\big), & \text{for } s \leq t, \\ -\psi_\mu(t)e^{-(s-t)J_+}\big(v_+ + e^{-(b-s)J_+}(F_b W_+)^{-1}F_b W_- e^{(b-s)J_-}v_-\big), & \text{for } t \leq s. \end{cases} \tag{17}$$

Proof. From equation (15) we have that $\partial_t^\mu \mathcal{G}_*(t, s)$ equals

$$\begin{cases} \alpha_{2k}^{-1} \bar{v}_1 J^\mu \exp(tJ) H^{-1} \bar{F}_a W \exp\big((a-s)J\big) W^{-1} v_2, & \text{for } s \leq t, \\ -\alpha_{2k}^{-1} \bar{v}_1 J^\mu \exp(tJ) H^{-1} \bar{F}_b W \exp\big((b-s)J\big) W^{-1} v_2, & \text{for } t \leq s. \end{cases}$$

The crux of the reformulation of this representation lies in the inversion of the matrix $H = \bar{F}_a W \exp(aJ) + \bar{F}_b W \exp(bJ)$. To this end, we write $H e^{-tJ}$ and $e^{tJ} H^{-1}$ as block matrices with $k \times k$-blocks,

$$H e^{-tJ} = \begin{pmatrix} F_a W_- e^{(a-t)J_-} & F_a W_+ e^{(a-t)J_+} \\ F_b W_- e^{(b-t)J_-} & F_b W_+ e^{(b-t)J_+} \end{pmatrix}, \quad e^{tJ} H^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

Using elementary matrix algebra we find that $w_1 A_{11} + w_2 A_{21}$ for general $w_1, w_2 \in \mathbb{R}^{1 \times k}$ equals

$$\big(w_1 - w_2 e^{-(b-t)J_+} (F_b W_+)^{-1} F_b W_- e^{(b-t)J_-}\big)$$
$$\big(F_a W_- e^{(a-t)J_-} - F_a W_+ e^{-(b-a)J_+} (F_b W_+)^{-1} F_b W_- e^{(b-t)J_-}\big)^{-1}. \tag{18}$$

Inserting this above we have that $\partial_t^\mu \mathcal{G}_*(t, s)$ for $s \leq t$ equals

$$\alpha_{2k}^{-1} \begin{pmatrix} v_1 J_-^\mu & v_1 J_+^\mu \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} F_a W_- e^{(a-s)J_-} & F_a W_+ e^{(a-s)J_+} \\ 0_{k \times k} & 0_{k \times k} \end{pmatrix} \begin{pmatrix} v_- \\ v_+ \end{pmatrix}$$

which equals

$$\alpha_{2k}^{-1} (v_1 J_-^\mu A_{11} + v_1 J_+^\mu A_{21}) \big(F_a W_- e^{(a-s)J_-} v_- + F_a W_+ e^{(a-s)J_+} v_+\big). \tag{19}$$

Combining equations (18) and (19) and rearranging the exponential factors we arrive at equation (17) for $s \leq t$. The reformulation is done similarly for $t \leq s$. □

Remark. From the viewpoint of statistical modeling, the results in [15] are more general in two valuable ways. Firstly, the boundary conditions separately given at the end-points of the sample interval via the matrices F_a, F_b in Theorem 1 may be given in form of linear combinations of the curve and its derivatives at a and b via general $\bar{F}_a$ and $\bar{F}_b$. In particular, boundary conditions enforcing periodicity may be stated. But to derive the numerically stable formulae stated in Proposition 2, we have refrained from this possibility. Secondly, the results in [15] are given for matrix-valued functions. This generalization allows our methods to be extended to multivariate functional data analysis.

In the following theorem the explicit inversion formula is applied to derive a simultaneous computation of $\partial_t^\mu (\mathbb{I} + \Delta \mathcal{L})^{-1} \mathcal{E}_z(t_n)$ for $n = 1, \ldots, N$, where $z \in \mathbb{R}^N$, that easily may be implemented with computational complexity $O(N)$. Furthermore, the inner integral in the approximation equation (12) of the logarithmic determinant may be explicitly computed for Lebesgue almost all $v \in [0, 1]$. In the statement of the theorem we denote by $\odot$ the element-wise multiplication of matrices or vectors of the same dimension. Unless specified otherwise the $\odot$ operation is performed after ordinary matrix multiplications.

Theorem 2. *Suppose the discretization $\mathcal{T}$ is equidistant with mesh length $\Delta = (b-a)/N$, and assume that the operator in equation (13) given by $\mathscr{L}_* = \mathbb{I} + \Delta\mathscr{L}$ satisfies the conditions of Proposition 2. Denote by $\mathcal{G}_*$ the Green's function for $\mathscr{L}_*$, let $J_-, J_+, W_-, W_+, v_-, v_+, \phi_\mu(t), \psi_\mu(t)$ be as defined in Proposition 2, and let $\xi_-, \xi_-^0, \xi_-^1, \xi_+, \xi_+^0, \xi_+^1 \in \mathbb{R}^{k\times 1}$ be defined by*

$$\xi_- = \left\{\frac{\exp(\Delta\eta_i^-/2) - 1}{\eta_i^-}\right\}_{i=1,\dots,k}, \qquad \xi_+ = \left\{\frac{1 - \exp(-\Delta\eta_i^+/2)}{\eta_i^+}\right\}_{i=1,\dots,k},$$

$$\xi_-^0 = \left\{\frac{1 - (1 - \Delta\eta_i^-)\exp(\Delta\eta_i^-)}{\Delta(\eta_i^-)^2}\right\}_{i=1,\dots,k}, \qquad \xi_+^0 = \left\{\frac{\exp(-\Delta\eta_i^+) - 1 + \Delta\eta_i^+}{\Delta(\eta_i^+)^2}\right\}_{i=1,\dots,k},$$

$$\xi_-^1 = \left\{\frac{\exp(\Delta\eta_i^-) - 1 - \Delta\eta_i^-}{\Delta(\eta_i^-)^2}\right\}_{i=1,\dots,k}, \qquad \xi_+^1 = \left\{\frac{1 - (1 + \Delta\eta_i^+)\exp(-\Delta\eta_i^+)}{\Delta(\eta_i^+)^2}\right\}_{i=1,\dots,k}.$$

For $z = \{z_j\}_{j=1,\dots,N} \in \mathbb{R}^N$ the μth derivative $\partial_t^\mu(\mathbb{I} + \Delta\mathscr{L})^{-1}\mathscr{E}_z(t_n)$ taken at the sample point t_n is given by

$$\phi_\mu(t_n)1_{n>1}\sum_{j=1}^{n-1}e^{(t_n-t_{j+1})J_-}(v_-\odot\xi_-^0)z_j$$

$$+ \phi_\mu(t_n)\sum_{j=1}^{n}e^{(t_n-t_j)J_-}\big(v_-\odot(1_{j=1}\xi_- + 1_{j>1}\xi_-^1)\big)z_j$$

$$- \psi_\mu(t_n)\sum_{j=n}^{N}e^{-(t_j-t_n)J_+}\big(v_+\odot(1_{j<N}\xi_+^0 + 1_{j=N}\xi_+)\big)z_j$$

$$- \psi_\mu(t_n)1_{n<N}\sum_{j=n+1}^{N}e^{-(t_{j-1}-t_n)J_+}(v_+\odot\xi_+^1)z_j$$

$$+ \phi_\mu(t_n)e^{(t_n-a)J_-}(F_aW_-)^{-1}F_aW_+1_{n>1}\sum_{j=1}^{n-1}e^{-(t_j-a)J_+}(v_+\odot\xi_+^0)z_j$$

$$+ \phi_\mu(t_n)e^{(t_n-a)J_-}(F_aW_-)^{-1}F_aW_+\sum_{j=1}^{n}e^{-(t_{j-1}-a)J_+}\big(v_+\odot(1_{j=1}\xi_+ + 1_{j>1}\xi_+^1)\big)z_j$$

$$- \psi_\mu(t_n)e^{-(b-t_n)J_+}(F_bW_+)^{-1}F_bW_-\sum_{j=n}^{N}e^{(b-t_{j+1})J_-}\big(v_-\odot(1_{j<N}\xi_-^0 + 1_{j=N}\xi_-)\big)z_j$$

$$- \psi_\mu(t_n)e^{-(b-t_n)J_+}(F_bW_+)^{-1}F_bW_-1_{n<N}\sum_{j=n+1}^{N}e^{(b-t_j)J_-}(v_-\odot\xi_-^1)z_j.$$

Concerning the log determinant assume that the operator in equation (13) given by $\mathscr{L}_* = v\mathbb{I} + \Delta\mathscr{L}$ for fixed $v \in [0,1]$ satisfies the conditions of Proposition 2. Let the matrices

$A_{--}, A_{++}, A_{-+}, A_{+-} \in \mathbb{R}^{k \times k}$ *be defined by*

$$A_{--} = \left\{ 1_{i=j} N e^{(b-a)\eta_i^-} + 1_{i \neq j} \frac{e^{(b-a)\eta_i^-} - e^{(b-a)\eta_j^-}}{\Delta(\eta_i^- - \eta_j^-)} \right\}_{i,j=1,\ldots,k},$$

$$A_{++} = \left\{ 1_{i=j} N e^{-(b-a)\eta_i^+} + 1_{i \neq j} \frac{e^{-(b-a)\eta_i^+} - e^{-(b-a)\eta_j^+}}{\Delta(-\eta_i^+ + \eta_j^+)} \right\}_{i,j=1,\ldots,k},$$

$$A_{-+} = \left\{ \frac{1 - e^{(b-a)(\eta_i^- - \eta_j^+)}}{\Delta(-\eta_i^- + \eta_j^+)} \right\}_{i,j=1,\ldots,k},$$

$$A_{+-} = \left\{ \frac{1 - e^{-(b-a)(\eta_i^+ - \eta_j^-)}}{\Delta(\eta_i^+ - \eta_j^-)} \right\}_{i,j=1,\ldots,k},$$

and let the matrix $B \in \mathbb{R}^{k \times k}$ *be defined by*

$$(F_a W_-)^{-1} F_a W_+ e^{-(b-a)J_+} (F_b W_+)^{-1} F_b W_-$$

$$\left(I_{k \times k} - e^{(b-a)J_-} (F_a W_-)^{-1} F_a W_+ e^{-(b-a)J_+} (F_b W_+)^{-1} F_b W_- \right)^{-1}.$$

Denoting by $\tau > 0$ *the leading coefficient of* $\mathscr{L}$, *then the integral* $\int_a^b \mathcal{G}_*(t,t)\,\mathrm{d}t$ *equals the sum of the following 8 terms*:

$$I = N \tau^{-1} v_1 v_-,$$

$$II = \tau^{-1} v_1 \big((F_a W_-)^{-1} F_a W_+ \odot A_{-+} \big) v_+,$$

$$III = -\tau^{-1} v_1 \big((F_b W_+)^{-1} F_b W_- \odot A_{+-} \big) v_-,$$

$$IV = -\tau^{-1} v_1 \big((F_b W_+)^{-1} F_b W_- e^{(b-a)J_-} (F_a W_-)^{-1} F_a W_+ \odot A_{++} \big) v_+,$$

$$V = \tau^{-1} v_1 (B \odot A_{--}) v_-,$$

$$VI = \tau^{-1} v_1 \big(B e^{(b-a)J_-} (F_a W_-)^{-1} F_a W_+ \odot A_{-+} \big) v_+,$$

$$VII = -\tau^{-1} v_1 \big((F_b W_+)^{-1} F_b W_- e^{(b-a)J_-} B \odot A_{+-} \big) v_-,$$

$$VIII = -\tau^{-1} v_1 \big((F_b W_+)^{-1} F_b W_- e^{(b-a)J_-} B e^{(b-a)J_-} (F_a W_-)^{-1} F_a W_+ \odot A_{++} \big) v_+.$$

Proof. Since the characteristic polynomial has distinct roots $\eta_1, \ldots, \eta_{2k}$, the Jordan canonical form of the companion matrix is diagonal, and equation (15) implies that $\partial_t^\mu (\mathbb{I} + \Delta \mathscr{L})^{-1} \mathscr{E}_z(t)$ equals

$$\alpha_{2k}^{-1} \bar{v}_1 J^\mu \exp(tJ) H^{-1} \bar{F}_a W \exp(aJ) \left\{ \int_a^t e^{-s\eta_i} \mathscr{E}_z(s)\,\mathrm{d}s \cdot (W^{-1} v_2)_i \right\}_{i=1,\ldots,2k}$$

$$- \alpha_{2k}^{-1} \bar{v}_1 J^\mu \exp(tJ) H^{-1} \bar{F}_b W \exp(bJ) \left\{ \int_t^b e^{-s\eta_i} \mathscr{E}_z(s)\,\mathrm{d}s \cdot (W^{-1} v_2)_i \right\}_{i=1,\ldots,2k}.$$

Since the function $\mathscr{E}_z$ is piecewise linear, the above integrals can be explicitly evaluated over the intervals $[t_j, t_{j+1}]$. For $j = 0, N$, we have

$$\int_a^{t_1} e^{-s\eta_i} \mathscr{E}_z(s) \, ds = e^{-a\eta_i} \frac{1 - \exp(-\Delta\eta_i/2)}{\eta_i} z_1,$$

$$\int_{t_N}^b e^{-s\eta_i} \mathscr{E}_z(s) \, ds = e^{-t_N\eta_i} \frac{1 - \exp(-\Delta\eta_i/2)}{\eta_i} z_N,$$

and for $j = 1, \ldots, N-1$, we have

$$\int_{t_j}^{t_{j+1}} e^{-s\eta_i} \mathscr{E}_z(s) \, ds$$

$$= \int_0^\Delta e^{-t_j\eta_i - \eta_i s} \left((1 - s\Delta^{-1})z_j + s\Delta^{-1}z_{j+1} \right) ds$$

$$= e^{-t_j\eta_i} \int_0^\Delta e^{-\eta_i s}(1 - \Delta^{-1}s) \, ds \, z_j + e^{-t_j\eta_i} \int_0^\Delta e^{-\eta_i s} \Delta^{-1}s \, ds \, z_{j+1}$$

$$= e^{-t_j\eta_i} \frac{\exp(-\Delta\eta_i) - 1 + \Delta\eta_i}{\Delta(\eta_i)^2} z_j + e^{-t_j\eta_i} \frac{1 - (1 + \Delta\eta_i)\exp(-\Delta\eta_i)}{\Delta(\eta_i)^2} z_{j+1}.$$

Arranging the eigenvalues as $\eta_1^-, \ldots, \eta_k^-, \eta_1^+, \ldots, \eta_k^+$ and inserting the definition of ξ_-, ξ_+, ξ_-^0, $\xi_+^0, \xi_-^1, \xi_+^1$, we have that $\partial_t^\mu (\mathbb{I} + \Delta\mathscr{L})^{-1}\mathscr{E}_z(t_n)$ equals

$$1_{n>1} \sum_{j=1}^{n-1} \alpha_{2k}^{-1} \bar{v}_1 J^\mu e^{t_n J} H^{-1} \bar{F}_a W e^{(a-t_j)J} \begin{pmatrix} v_- \odot e^{-\Delta J_-} \xi_-^0 \\ v_+ \odot \xi_+^0 \end{pmatrix} z_j$$

$$+ \sum_{j=1}^n \alpha_{2k}^{-1} \bar{v}_1 J^\mu e^{t_n J} H^{-1} \bar{F}_a W e^{(a-t_{j-1})J} \begin{pmatrix} v_- \odot (1_{j=1}e^{-\Delta J_-/2}\xi_- + 1_{j>1}e^{-\Delta J_-}\xi_-^1) \\ v_+ \odot (1_{j=1}\xi_+ + 1_{j>1}\xi_+^1) \end{pmatrix} z_j$$

$$- \sum_{j=n}^N \alpha_{2k}^{-1} \bar{v}_1 J^\mu e^{t_n J} H^{-1} \bar{F}_b W e^{(b-t_j)J} \begin{pmatrix} v_- \odot (1_{j<N}e^{-\Delta J_-}\xi_-^0 + 1_{j=N}e^{-\Delta J_-/2}\xi_-) \\ v_+ \odot (1_{j<N}\xi_+^0 + 1_{j=N}\xi_+) \end{pmatrix} z_j$$

$$- 1_{n<N} \sum_{j=n+1}^N \alpha_{2k}^{-1} \bar{v}_1 J^\mu e^{t_n J} H^{-1} \bar{F}_b W e^{(b-t_{j-1})J} \begin{pmatrix} v_- \odot e^{-\Delta J_-}\xi_-^1 \\ v_+ \odot \xi_+^1 \end{pmatrix} z_j.$$

The exponential factors on the terms ξ_-, ξ_-^0, ξ_-^1 may be assimilated in the exponential factors before the large parenthesis using $t_{j+1} - t_j = \Delta$ for $j = 1, \ldots, N$ and $t_2 - t_1 = t_{N+1} - t_N = \Delta/2$. Thereafter the terms in these sums are of the same type as in equation (19) with v_-, v_+ replaced by $v_- \odot \xi_-^0$, $v_+ \odot \xi_+^0$ etc., and the formula for $\partial_t^\mu (\mathbb{I} + \Delta\mathscr{L})^{-1}\mathscr{E}_z(t_n)$ follows by invoking the same reformulations as used in the proof of Proposition 2.

Finally, we consider the Green's function $\mathcal{G}_*$ for $\mathcal{L}_* = v\mathbb{I} + \Delta\mathcal{L}$. The differential operator $\mathcal{L}_*$ has leading coefficient $\alpha_{2k} = \Delta\tau$, and inserting $s = t$ in the first part of equation (17), we find that $\mathcal{G}_*(t, t)$ equals

$$\Delta^{-1}\tau^{-1}\left(v_1 e^{-(b-t)J_-} - v_1 e^{-(b-t)J_+}(F_b W_+)^{-1} F_b W_-\right)$$

$$\left(I_{k\times k} - e^{(b-a)J_-}(F_a W_-)^{-1} F_a W_+ e^{-(b-a)J_+}(F_b W_+)^{-1} F_b W_-\right)^{-1}$$

$$\left(e^{(b-t)J_-} v_- + e^{(b-a)J_-}(F_a W_-)^{-1} F_a W_+ e^{-(t-a)J_+} v_+\right).$$

To remove the possibly exploding exponential factor $e^{-(b-t)J_-}$ in the first factor, we invoke the matrix formula $(I - X)^{-1} = I + X(I - X)^{-1}$ on the second factor and rearranging the exponential factors. Doing this $\mathcal{G}_*(t, t)$ is rewritten as the numerically stable expression

$$\Delta^{-1}\tau^{-1}\left(v_1 - v_1 e^{-(b-t)J_+}(F_b W_+)^{-1} F_b W_- e^{(b-t)J_-}\right)$$

$$\left(v_- + e^{(t-a)J_-}(F_a W_-)^{-1} F_a W_+ e^{-(t-a)J_+} v_+\right)$$

$$+ \Delta^{-1}\tau^{-1}\left(v_1 e^{(t-a)J_-} - v_1 e^{-(b-t)J_+}(F_b W_+)^{-1} F_b W_- e^{(b-a)J_-}\right)$$

$$(F_a W_-)^{-1} F_a W_+ e^{-(b-a)J_+}(F_b W_+)^{-1} F_b W_-$$

$$\left(I_{k\times k} - e^{(b-a)J_-}(F_a W_-)^{-1} F_a W_+ e^{-(b-a)J_+}(F_b W_+)^{-1} F_b W_-\right)^{-1}$$

$$\left(e^{(b-t)J_-} v_- + e^{(b-a)J_-}(F_a W_-)^{-1} F_a W_+ e^{-(t-a)J_+} v_+\right).$$

This expression is expanded into the sum of 8 terms, which all may be explicitly integrated over the interval $[a, b]$. For instance is the integral over the second term given by

$$\int_a^b \Delta^{-1}\tau^{-1} v_1 e^{(t-a)J_-}(F_a W_-)^{-1} F_a W_+ e^{-(t-a)J_+} v_+ \, dt,$$

which equals $\tau^{-1} v_1((F_a W_-)^{-1} F_a W_+ \odot A_{-+}) v_+$. $\square$

Remark. The predictors $\mathrm{E}[x_m | y]$ may be seen as the predictors $\mathrm{E}[x_m^{\mathrm{fct}} | y]$ for the functional parameters x_m^{fct} evaluated at the sample points t_n. The formulae stated in Theorem 2 may be extended to functional representations for $\mathrm{E}[x_m^{\mathrm{fct}} | y]$. Doing this the predictions between sample points will be given as linear combinations of exponential functions.

Remark. If the kernel $\mathcal{G}(t, s)$ is constant, say $\mathcal{G}(t, s) = \lambda$, then the operator approximation

$$\int_0^1 \sum_{j=1}^N (v\mathbb{I} + \mathcal{M}_{\mathcal{T}}^{-1}\mathcal{L})^{-1}\mathcal{E}_{e_j}(t_j) \, dv = \int_0^1 \int_a^b \frac{\lambda N/(b-a)}{1 + Nv\lambda} \, dt \, dv = \log(1 + N\lambda)$$

gives the exact log determinant of $\{1_{n=m} + \mathcal{G}(t_n, t_m)\}_{n,m} = \mathbb{I}_N + \{\lambda\}_{n,m}$. The particular construction of the embedding operator $\mathcal{E}_z$ was chosen to achieve this property.

A fundamental difference between our operator methods and the smoothing spline technology lies in our dependence on boundary conditions. Whether boundary conditions are desirable in statistical modeling depends on the data situation at hand. If we have additional knowledge implying particular boundary conditions, then this may be used in the statistical model. However, in many data situations such additional knowledge is not available, and the requirement to specify boundary conditions may be disturbing. Here our advice is to use Neumann-type conditions. Although the covariance function $\mathcal{G}(t, s)$ is not defined for Neumann conditions as noted in the following example, this is possible due to the regularization induced by the measurement noise; that is, $\mathbb{I} + \mathcal{M}_T^{-1}\mathcal{L}$ is non-singular by construction.

Example. For $\mathcal{K} = \lambda \partial_t$ we have $\mathcal{L} = \mathcal{K}^\dagger \mathcal{K} = -\lambda^2 \partial_t^2$. Consider the following two sets of boundary conditions:

$$(B1):\ \theta(a) = \theta^{(1)}(b) = 0, \qquad (B2):\ \theta(a) = \theta(b) = 0.$$

We have $\mathcal{L}^{-1}\theta(t) = \int_a^b \mathcal{G}(t, s)\theta(s)\,\mathrm{d}s$ with

$$\mathcal{G}(t, s) = \begin{cases} \lambda^{-2}\big((t \wedge s) - a\big), & \text{for boundary conditions (B1)}, \\ \lambda^{-2}\dfrac{((t \wedge s) - a)(b - (t \vee s))}{b - a}, & \text{for boundary conditions (B2)}. \end{cases}$$

Thus, the Laplace operator with boundary conditions (B1) leads to the Brownian motion, and the Laplace operator with boundary conditions (B2) leads to the Brownian bridge. The Laplace operator with Neumann boundary conditions $\theta^{(1)}(a) = \theta^{(1)}(b) = 0$ is not positive definite. Even so, this operator can be used in a statistical model, where it implies an improper prior for the serially correlated effects in terms of a Brownian motion with a free level.

To compute the approximative log likelihood we find the Green's function $\mathcal{G}_v$ for $v\mathbb{I} + \frac{b-a}{N}\mathcal{L}$. In case of the Brownian, motion equation (17) gives

$$\mathcal{G}_v(t, s) = \frac{1}{\lambda}\sqrt{\frac{b-a}{Nv}}$$

$$\times \frac{\sinh((((t \wedge s) - a)/(\lambda\sqrt{b-a}))\sqrt{Nv})\cosh(((b - (t \vee s))/(\lambda\sqrt{b-a}))\sqrt{Nv})}{\cosh(\lambda^{-1}\sqrt{b-a}\sqrt{Nv})},$$

$$\int_a^b \mathcal{G}_v(t, t)\,\mathrm{d}t = \frac{\sqrt{b-a}}{2\lambda}\sqrt{\frac{N}{v}} \cdot \frac{\sinh(\lambda^{-1}\sqrt{b-a}\sqrt{Nv})}{\cosh(\lambda^{-1}\sqrt{b-a}\sqrt{Nv})}.$$

In case of the Brownian bridge, equation (17) gives

$$\mathcal{G}_v(t, s) = \frac{1}{\lambda}\sqrt{\frac{b-a}{Nv}}$$

$$\times \frac{\sinh((((t \wedge s) - a)/(\lambda\sqrt{b-a}))\sqrt{Nv})\sinh(((b - (t \vee s))/(\lambda\sqrt{b-a}))\sqrt{Nv})}{\sinh(\lambda^{-1}\sqrt{b-a}\sqrt{Nv})},$$

$$\int_a^b \mathcal{G}_v(t,t)\,dt = \frac{\sqrt{b-a}}{2\lambda}\sqrt{\frac{N}{v}} \cdot \frac{\cosh(\lambda^{-1}\sqrt{b-a}\sqrt{Nv})}{\sinh(\lambda^{-1}\sqrt{b-a}\sqrt{Nv})} - \frac{1}{2v}.$$

In both cases the double integrals $\int_0^1 \int_a^b \mathcal{G}_v(t,t)\,dt\,dv$ can be computed giving explicit formulae for the operator approximation of the matrix determinants. In case of an equidistantly sampled Brownian motion, we have

$$\log\det\{1_{n=m} + \mathcal{G}(t_n,t_m)\}_{n,m=1,\dots,N} \approx \log\bigl(\cosh(\lambda^{-1}\sqrt{b-a}\sqrt{N})\bigr),$$

and in case of an equidistantly sampled Brownian bridge, we have

$$\log\det\{1_{n=m} + \mathcal{G}(t_n,t_m)\}_{n,m=1,\dots,N} \approx \log\left(\frac{\sinh(\lambda^{-1}\sqrt{b-a}\sqrt{N})}{\lambda^{-1}\sqrt{b-a}\sqrt{N}}\right).$$

4. Approximative inference

In this section we combine the matrix formulae listed in Section 2 with the operator approximation developed in Section 3. The obstacle in the matrix computations is the inversion of the matrix $A_0 = \mathbb{I}_N + R_0 \in \mathbb{R}^{N\times N}$. Here $R_0 = \{\mathcal{G}(t_n,t_m)\}_{n,m=1,\dots,N}$ is defined via a discretization $\mathcal{T} = \{t_1,\dots,t_N\}$ and the Green's function $\mathcal{G}$ for a differential operator $\mathcal{L} = \sum_{l=1}^L \mathcal{K}_l^\dagger \mathcal{K}_l$.

The maximum likelihood estimator and the BLUPs given in equations (4) and (5) are approximated using the block structure $A = A_0 \otimes \mathbb{I}_M$, the identity $A_0^{-1} z = z - R_0 A_0^{-1} z$ for $z \in \mathbb{R}^N$ and the approximation

$$R_0 A_0^{-1} z = A_0^{-1} R_0 z = (\mathbb{I} + R_0^{-1})^{-1} z \approx \{(\mathbb{I} + \mathcal{M}_\mathcal{T}^{-1}\mathcal{L})^{-1}\mathcal{E}_z(t_n)\}_{n=1,\dots,N}.$$

Note that this approximation is applied both on the individual sample vectors $y_m \in \mathbb{R}^N$ and on the sections of the columns of the design matrices Γ and Z. The approximation of the logarithmic determinant equation (6) in the restricted likelihood equation (7) has already been stated in equation (12), and the quadratic form of the serially correlated effects is approximated by

$$\mathrm{E}[x|y]^\top R^{-1}\mathrm{E}[x|y] \approx \sum_{m=1}^M \sum_{n=1}^N \mathrm{E}[x_m(t_n)|y]^\top \mathcal{M}_\mathcal{T}^{-1}\mathcal{L}\mathrm{E}[x_m(t_n)|y].$$

Furthermore, for an equidistant discretization with mesh length Δ, we have

$$\mathrm{E}[x|y]^\top R^{-1}\mathrm{E}[x|y] \approx \Delta \sum_{l=1}^L \sum_{m=1}^M \sum_{n=1}^N (\mathcal{K}_l\mathrm{E}[x_m(t_n)|y])^\top (\mathcal{K}_l\mathrm{E}[x_m(t_n)|y]).$$

If the discretization $\mathcal{T}$ is equidistant, then semi-explicit and numerically stable formulae for the above approximations are given in Section 3.1. For general discretizations the operator approximations may be found as numerical solutions to ordinary differential equations; for example, the function $f = (\mathbb{I} + \mathcal{M}_\mathcal{T}^{-1}\mathcal{L})^{-1}\mathcal{E}_z \in \mathcal{H}$ obeys to the differential equation $f + \mathcal{M}_\mathcal{T}^{-1}\mathcal{L}f = \mathcal{E}_z$.

Acknowledgements

References

[1] Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* **72** 320–340. MR0451550

[2] Henderson, H.V. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31** 423–447.

[3] Henderson, H.V. and Searle, S.R. (1981). On deriving the inverse of a sum of matrices. *SIAM Rev.* **23** 53–60. MR0605440

[4] Kackar, R.N. and Harville, D.A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Comm. Statist. A—Theory Methods* **10** 1249–1261. MR0625025

[5] Kincaid, D. and Cheney, W. (1991). *Numerical Analysis*. Pacific Grove, CA: Brooks/Cole Publishing Co. MR1099375

[6] Kou, S.C. and Efron, B. (2002). Smoothers and the C_p, generalized maximum likelihood, and extended exponential criteria: A geometric approach. *J. Amer. Statist. Assoc.* **97** 766–782. MR1941408

[7] Lee, Y., Nelder, J.A. and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects. Monographs on Statistics and Applied Probability* **106**. Boca Raton, FL: Chapman & Hall/CRC. MR2259540

[8] Markussen, B. (2011). fdaMixed: Functional data analysis in a mixed model framework. R package version 0.1. Available at http://CRAN.R-project.org/package=fdaMixed.

[9] Morrell, C.H. (1998). Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics* **54** 1560–1568.

[10] Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-PLUS. Springer Series in Statistics and Computing*. New York: Springer.

[11] Ramsay, J.O. and Silverman, B.W. (2002). *Applied Functional Data Analysis. Springer Series in Statistics*. New York: Springer. MR1910407

[12] Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2168993

[13] Robinson, G.K. (1991). That BLUP is a good thing: The estimation of random effects. *Statist. Sci.* **6** 15–51. MR1108815

[14] Verbyla, A.P., Cullis, B.R., Kenward, M.G. and Welham, S.J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines. *Appl. Statist.* **48** 269–311.

[15] Villanueva, R.J. and Jódar, L. (1992). Generalized Green's functions for higher order boundary value matrix differential systems. *Internat. J. Math. Math. Sci.* **15** 523–535. MR1169817

[16] Welham, S.J. and Thompson, R. (1997). Likelihood ratio tests for fixed model terms using residual maximum likelihood. *J. Roy. Statist. Soc. Ser. B* **59** 701–714. MR1452034