Computational Statistics and Data Analysis 72 (2014) 227–240

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Approximate inference for spatial functional data on massively parallel processors

Lars Lau Rakêt[a,*], Bo Markussen[b]

[a] *Department of Computer Science, University of Copenhagen, Denmark*

[b] *Department of Mathematical Sciences, University of Copenhagen, Denmark*

ARTICLE INFO

Article history:
Received 17 June 2013
Received in revised form 11 October 2013
Accepted 12 October 2013
Available online 23 October 2013

Keywords:
Functional data analysis
Functional mixed-effects model
Gaussian processes
GPU
Likelihood analysis
Operator approximations

ABSTRACT

With continually increasing data sizes, the relevance of the big n problem of classical likelihood approaches is greater than ever. The functional mixed-effects model is a well established class of models for analyzing functional data. Spatial functional data in a mixed-effects setting is considered, and so-called operator approximations for doing inference in the resulting models are presented. These approximations embed observations in function space, transferring likelihood calculations to the functional domain. The resulting approximated problems are naturally parallel and can be solved in linear time. An extremely efficient GPU implementation is presented, and the proposed methods are illustrated by conducting a classical statistical analysis of 2D chromatography data consisting of more than 140 million spatially correlated observation points.[1]

1. Introduction

During the last half century, functional data analysis has become a well-established statistical discipline (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Horváth and Kokoszka, 2012). The continuous sophistication of instruments gives rise to an increasing number of problems where functional aspects have to be taken into account. Statistical analysis of functional data generally involves the ill-posed problem of inferring an infinite-dimensional function from discrete data points. This requires some sort of regularization, and the type of regularization is often chosen in terms of roughness penalties that lead to sparse representations of the inferred function in terms of simple basis functions (Wahba, 1990), thus reducing the computational complexity. The most typical specification, however, considers the inverse regularization process where a sparse basis is chosen explicitly for the given problem, which may then be further regularized through a roughness penalty (Ramsay and Silverman, 2005).

This paper takes a different path for model specification; we consider functional mixed-effects models with random effects generated by Gaussian processes, and present a framework that moves the calculations needed in such analyses from the discrete domain induced by the observations to the underlying functional domain. As a consequence it is possible to efficiently compute the functions in question, even if the regularization does not lead to sparse representations. The methods are based on the one-dimensional operator approximations of Markussen (2013), and here new results and resolution strategies are presented for high-dimensional domains.

The functional viewpoint sheds new light on some of the current challenges in statistics (Jordan, 2011), by both reducing the computational complexity of a large class of statistical problems dramatically, and at the same time revealing a natural

* Corresponding author. Tel.: +45 50540319.
E-mail address: larslau@diku.dk (L.L. Rakêt).

1 Code for analyzing spatial functional data on graphics processing units is available as Supplementary material.

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.10.016

link between partial differential equations and a large number of statistical models, including functional mixed-effects models, penalized likelihood, and Bayesian models.

In addition to reducing the computational complexity, the proposed resolution strategies are highly parallel, and naturally suited for implementation on massively parallel processors like graphics processing units (GPUs). While parallelization and GPUs have received some attention in the statistical community in recent years, the main focus has been on parallelizing matrix operations and sampling techniques (Suchard et al., 2010; da Silva, 2010). To our knowledge, this work marks the first attempt of actively formulating solutions for classical statistical problems in a way that is particularly beneficial for implementation on massively parallel hardware.

The proposed methods are illustrated by conducting a classical statistical analysis of a dataset of 2D chromatograms with more than 140 million spatially correlated observations on a GPU.

2. Model and estimation

We consider spatial functional data on a domain $\mathscr{T} \subseteq \mathbb{R}^d$. Suppose we are given k noisy vectorized functional samples $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ each consisting of n observation points. We assume that the observations are generated from the following functional mixed-effect model:

$$y_i(\boldsymbol{t}) = \theta_{e(i)}(\boldsymbol{t}) + x_i(\boldsymbol{t}) + \varepsilon_i(\boldsymbol{t}) \tag{1}$$

where $e : \{1, \ldots, k\} \to \{1, \ldots, p\}$ is a factor, $\theta_{e(i)}$ is the fixed functional mean for group $e(i)$, x_i is a zero-mean Gaussian process with covariance function $\tau^2 \mathcal{G}$, and ε_i is a Gaussian white noise process with variance σ^2.

A wide variety of functional mixed-effects models have previously been considered. One of the dominant approaches is to model functional effects using smoothing splines (Wahba, 1990). Such constructions are considered by Wang (1998) and Guo (2002). Modeling of mixed effects in terms of penalized splines is considered by Chen and Wang (2011), and Lee et al. (2013) propose a related method based on nested basis functions for spatial mixed-effects models. An alternative approach to functional mixed-effect models considers the problem in a nonparametric setting, where no distributional or parametric assumptions are made on the random effects. Boularan et al. (1994) considered modeling of growth curves, assuming only that population and individual effects were twice differentiable, and proposed kernel smoothing estimates for the effects. On a similar note, Núñez-Antón et al. (1999) considered a nonparametric three-level model and applied it to speech recognition data. For the use of nonparametric statistical modeling techniques for functional data we refer to the monograph by Ferraty and Vieu (2006), and for a review on functional mixed-effects models we refer to Liu and Guo (2012).

Now, let $\boldsymbol{y}$ be the concatenation of all the vectorized observations of length $N = kn$. The discrete observation $\boldsymbol{y}$ generated by function evaluation at the points $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$ in the model (1) may be modeled by a conventional linear mixed-effects model

$$\boldsymbol{y} = \boldsymbol{\Gamma}\boldsymbol{\theta} + \boldsymbol{x} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\boldsymbol{\Gamma} = \mathbb{I}_n \otimes \boldsymbol{\Gamma}_0$ is the design matrix corresponding to the factor e and $\boldsymbol{\theta} \in \mathbb{R}^{np}$ is a vector of parameters describing the group mean functions pointwise, $\boldsymbol{x}$ consists of the spatially correlated effects, $\boldsymbol{x} \sim \mathcal{N}(0, \mathbb{I}_k \otimes \tau^2 \boldsymbol{\Sigma})$ with covariance matrix $\boldsymbol{\Sigma} = \left\{\mathcal{G}(\boldsymbol{t}_i, \boldsymbol{t}_j)\right\}_{i,j}$, and $\boldsymbol{\varepsilon}$ is independent, identically distributed Gaussian noise $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_N)$. Since the design is constant across all observations, i.e. given by $\boldsymbol{\Gamma}_0$, the fixed effect $\boldsymbol{\theta}$ can be estimated pointwise. The solution strategy presented below may also be adapted to the situation with a low rank design matrix following Markussen (2013).

Functional mixed-effect models are typically modeled with fixed effects of a functional nature. For simplicity, we parametrize the fixed effect with one parameter per observation point, mimicking classical mixed-effects models. The adaption to functional fixed effects given by a limited number of basis functions can be done following the previously mentioned references. In particular, the computations needed for fixed effects parametrized in terms of smoothing splines closely follow the computations related to the spatially correlated effect $\boldsymbol{x}$, and the presented methods naturally extend to such parametrizations.

The best linear unbiased prediction for the spatially correlated effects in the model (2) is done by means of the conditional expectation (Robinson, 1991)

$$E[\boldsymbol{x} \mid \boldsymbol{y}] = \left(\mathbb{I}_k \otimes \tau^2 \boldsymbol{\Sigma}\right) \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\Gamma}\hat{\boldsymbol{\theta}}), \tag{3}$$

where $\boldsymbol{V} = \sigma^2 \mathbb{I}_N + \mathbb{I}_k \otimes \tau^2 \boldsymbol{\Sigma}$. The variance parameters are typically estimated by minimizing the negative log restricted likelihood (Harville, 1977; Lee et al., 2006)

$$\ell_{\boldsymbol{y}}(\sigma, \tau) = \log\det \boldsymbol{V} + \log\det[\boldsymbol{\Gamma}^\top \boldsymbol{V}^{-1} \boldsymbol{\Gamma}] + (\boldsymbol{y} - \boldsymbol{\Gamma}\hat{\boldsymbol{\theta}})^\top \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\Gamma}\hat{\boldsymbol{\theta}}). \tag{4}$$

For later use it is noted that the last term in the likelihood function can be written as

$$\begin{aligned}(\boldsymbol{y} - \boldsymbol{\Gamma}\hat{\boldsymbol{\theta}})^\top \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\Gamma}\hat{\boldsymbol{\theta}}) &= \frac{1}{\sigma^2}(\boldsymbol{y} - \boldsymbol{\Gamma}\hat{\boldsymbol{\theta}})^\top(\boldsymbol{y} - \boldsymbol{\Gamma}\hat{\boldsymbol{\theta}} - E[\boldsymbol{x} \mid \boldsymbol{y}]) \\ &= \frac{1}{\sigma^2}(\boldsymbol{y} - \boldsymbol{\Gamma}\hat{\boldsymbol{\theta}} - E[\boldsymbol{x} \mid \boldsymbol{y}])^\top(\boldsymbol{y} - \boldsymbol{\Gamma}\hat{\boldsymbol{\theta}} - E[\boldsymbol{x} \mid \boldsymbol{y}]) + \frac{1}{\sigma^2}E[\boldsymbol{x} \mid \boldsymbol{y}]^\top(\boldsymbol{y} - \boldsymbol{\Gamma}\hat{\boldsymbol{\theta}} - E[\boldsymbol{x} \mid \boldsymbol{y}]).\end{aligned} \tag{5}$$