Journal of Statistical Planning and Inference 206 (2020) 263–277

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Optimal design for adaptive smoothing splines

Jiali Wang [a,*], Arūnas P. Verbyla [a], Bomin Jiang [b], Alexander B. Zwart [a], Cheng Soon Ong [a,c], Xavier R.R. Sirault [d,e], Klara L. Verbyla [a]

[a] Data61, Commonwealth Scientific and Industrial Research Organisation, Australia
[b] MIT Institute for Data, Systems, and Society, USA
[c] Department of Computer Science, Australian National University, Australia
[d] Agriculture and Food, Commonwealth Scientific and Industrial Research Organisation, Australia
[e] The High Resolution Plant Phenomics Centre, Australian Plant Phenomics Facility, Commonwealth Scientific and Industrial Research Organisation, Australia

ARTICLE INFO

Article history:
Received 1 August 2019
Received in revised form 26 September 2019
Accepted 9 October 2019
Available online 16 October 2019

Keywords:
Adaptive smoothing splines
Growth curve modelling
Linear mixed models
Optimal design

ABSTRACT

We consider the design problem of collecting temporal/longitudinal data. The adaptive smoothing spline is used as the analysis model where the prior curvature information can be naturally incorporated as a weighted smoothness penalty. The estimator of the curve is expressed in linear mixed model form, and the information matrix of the parameters is derived. The D-optimality criterion is then used to compute the optimal design points. An extension is considered, for the case where subpopulations exert different prior curvature patterns. We compare properties of the optimal designs with the uniform design using simulated data and apply our method to the Berkeley growth data to estimate the optimal ages to measure heights for males and females. The approach is implemented in an R package called "ODsplines", which is available from github.com/jialiwang1211/ODsplines.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Data collection can be costly preventing researchers from performing measurements on the continuum. Optimal design concerns the matter of planning for the collection of data from experiments in an efficient manner. In a broad sense, randomization, blocking and replication are the fundamental principles when designing an experiment and there is a vast literature on these topics (Atkinson et al., 2007; Montgomery, 2017). In biological and agricultural studies, researchers often monitor the growth of plants, for example, responses to stress conditions over time and different growing trajectories with different genotypes. Even with modern high-throughput automatic scanning platforms, cost and time constraints may still make collecting high-frequency data from a large number of biological replicates unrealistic. Therefore, principled guidance for collecting data is required to maximize the information gain from limited data and reduce bias when making an inference. In this article, we limit our attention to the design problem when collecting temporal/longitudinal data, that is, when to perform measurements of the experimental units over time in order to best capture the true behaviour of the system being observed.

The optimality of a design depends upon the statistical model that will be used to analyse the data. Consider the model $y = \eta(t, \boldsymbol{\beta}) + \epsilon$, where t denotes time, y denotes the response and $\boldsymbol{\beta}$ is the vector of model parameters. Note that while we use time t as the time ordinate in this paper, more generally the ordinate can be of other types, such as dose concentration

* Correspondence to: Data61, Commonwealth Scientific and Industrial Research Organisation, ACT 2601, Australia.
E-mail address: wangjiali0320@gmail.com (J. Wang).

https://doi.org/10.1016/j.jspi.2019.10.002
0378-3758/© 2019 Elsevier B.V. All rights reserved.

in a dose–response curve. The design objective with respect to this model is to find the vector $\boldsymbol{T}^* = (t_1^*, \ldots, t_n^*)^T$ (which we refer to as the 'optimal design' or the set of 'optimal design points').

Various choices of linear model can be applied to time dependent data, for example, a polynomial regression model with Gaussian noise. The information matrix of the parameters can be estimated using the method of least squares and does not depend on $\boldsymbol{\beta}$, hence the solution $\boldsymbol{T}^*$ is independent of the parameter values of the fitted curve. However, linear models can be too restrictive for describing a dynamic system and are prone to under-/over-fitting. Parametric nonlinear models can be fitted instead, for example, three-parameter logistic model and Gompertz model (Paine et al., 2012). Linearization by Taylor expansion about a parameter value $\boldsymbol{\beta}^*$ can be used to convert the nonlinear model into linear model form, however a prior knowledge of $\boldsymbol{\beta}^*$ is then required (Atkinson et al., 2007, Chapter 17) to derive the optimal design. Optimal design for the logistic model has been studied in Li and Majumdar (2008) and for the Gompertz model in Li (2012). As an alternative, the Bayesian approach to optimal design incorporates uncertainty in the parameter values when constructing the objective function by averaging the information matrix over the prior distribution of the parameters; see Chaloner and Verdinelli (1995) for a review, and Donev et al. (2008) for an application.

Nonlinear models with few parameters may remain insufficiently flexible in many applications, and optimal designs derived from these models can be very sensitive to the choice of $\boldsymbol{\beta}^*$. Instead we consider nonparametric models, specifically the smoothing spline as the analysis model from which optimal designs are derived. Due to its nonlinear nature, solving the design problem for smoothing spline also requires prior knowledge about the parameters (potentially in a high dimensional space), and choosing a good prior is challenging. There has been extensive research on knot selection for fitting spline models after the data have been collected (Miyata and Shen, 2003; DiMatteo et al., 2001), but there is a relative sparsity of literature on the determination of design points prior to conducting an experiment. Park (1978) derived the D-optimal design for segmented polynomial regression with a single knot. Some extensions were made by Kaishev (1989) and Heiligers (1998), to consider an arbitrary number of knots and multiplicities in polynomial spline regression. In these cases, knots did not need to coincide with the design points and did need to be determined as the prior input. Dette et al. (2008) indicated that the optimal designs were not necessarily robust with respect to the prior guess for the vector of knots, and they proposed a standardized maximin D-optimal design for free knot least square splines which they found was less sensitive to the specification of the unknown knots. Instead of working with polynomial splines via least squares, Dette et al. (2011) assumed the curve was estimated from a smoothing spline where the smoothness was controlled by the smoothing parameter λ, and they derived the information matrix via a system of new basis functions. The prior knowledge that was required to optimize the design was the level of smoothness, and they showed through simulations that as the smoothing parameter increased, G-optimal design points became more concentrated at the boundaries of the design region, but that D-optimal design points were less affected. Notice that in Dette et al. (2011), the design points were distributed symmetrically across the design interval due to the fact that the only prior knowledge used in optimization was the global smoothness parameter λ, so that no local properties could be incorporated into the design. It is worth noting that there are some recent works on D-optimal designs for active learning in the discipline of computer vision and they were primarily applied to determine the unlabelled data to better the separation of images (He, 2009; Gu and Jin, 2013). The neighbourhood structure of the data was preserved by imposing a similarity based locality preserving regularizer, based on the prior belief that if two points are close to each other, their measurements should be close as well.

In this paper, we propose incorporating the curvature (or second derivative/acceleration) of the curve as the prior information for optimization. The prior curvature knowledge is particularly informative when determining optimal sampling points for longitudinal data, because intuitively more observations should be placed at the locations where the shape of the curve is changing rapidly. To the best of our knowledge, including curvature as the prior information in a smoothing spline optimal design problem has not been explored in the literature.

The paper is structured as follows. In Section 2, we review the adaptive smoothing spline model and show how curvature information can be incorporated naturally into the design problem. Under some mild conditions, the estimated curve can be represented in matrix form similar to the natural cubic spline, which has the equivalent linear mixed model formulation. The D-optimality criterion is used to define the optimization problem. A numerical approach to finding the optimal design points is then outlined. In Section 3, we consider the issues involved in obtaining prior curvature information from historical data, as well as choosing the smoothing parameter and the number of design points. In Section 4, two simulation studies are performed for growth curves that are assumed to follow a logistic model, and a mixture parametric model. We compare the optimal design with the uniform design with respect to the distribution of the design points and goodness-of-fit. The female height data set from the Berkeley Growth study (Tuddenham, 1954) is used as a real data example. In Section 5 we consider the determination of optimal design points where subpopulations with different curvature patterns exist. Results from a third simulation for two logistic curves are presented and the Berkeley Growth study data set with both males and females is used to illustrate this scenario. Section 6 concludes the paper and presents some future research directions.

2. Optimal designs for adaptive smoothing splines

Consider the model fitted to the data

$$y_i = g(t_i) + \epsilon_i, \ \ i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T \sim N(0, \sigma^2 \boldsymbol{R})$. Without loss of generality, we assume the residuals are independent and identically distributed, such that $\boldsymbol{R} = \boldsymbol{I_n}$, and the design points $t_1, \ldots, t_n$ are distinct and bounded within the design region $\mathcal{T}$: $t_1 < t_2 < \cdots < t_{n-1} < t_n \in \mathcal{T} = [0, 1]$. To estimate a smooth curve, we minimize the residual sum of squares

$$\sum_{i=1}^{n} (y_i - g(t_i))^2,$$

for all possible $\boldsymbol{g}$ from the set

$$\mathscr{F} = \{\boldsymbol{g} : \boldsymbol{g}^{(m)} \text{ is continuous, m} = 0,1,2 \text{ and } \int_0^1 [g''(t)]^2 dt < \infty \mid \int_0^1 \left[\frac{g''(t)}{f''(t)}\right]^2 dt < S\},$$

for some positive number S. The function $f''(t)$ is the curvature, which must be specified as the prior input. Denote $\lambda(t) = 1/[f''(t)]^2$. We assume that $\lambda(t)$ is bounded integrable on [0,1] and $1/\lambda(t)$ is integrable on [0,1]. An estimator of $\boldsymbol{g}$ can be found as the solution to the minimization problem

$$\min_{\boldsymbol{g} \in \mathscr{F}} \left\{ \sum_{i=1}^{n} (y_i - g(t_i))^2 + \rho \int_0^1 \lambda(t)[g''(t)]^2 dt \right\}, \tag{2}$$

where ρ is the smoothing parameter.

Model (2) has been studied in Pintore et al. (2006) where the solution is given by reproducing kernels within a reproducing kernel Hilbert space (Wahba, 1990). The time dependence of $\lambda(t)$ makes the spline model adaptive, in that different smoothness penalties are applied at different times t. In Pintore et al. (2006), $\lambda(t)$ is estimated from the data through generalized cross validation, whereas in our design problem, we treat $\lambda(t)$ as a prior function based on the prior curvature $f''(t)$. The term $\left[g''(t)/f''(t)\right]^2$ can be interpreted as a weighted roughness penalty, so that the smoothness of the curve is not globally uniform but is scaled by the prior knowledge of the curvature. The optimization problem (2) thus achieves a trade-off between goodness-of-fit and the weighted smoothness of the fitted curve. The tuning parameter ρ is subject to selection, and we will discuss the choice of ρ in Section 3.2.

2.1. Constructing adaptive smoothing splines

When constructing the spline model, we assume a polynomial of order 4 between any two interior knots and continuity between segments up to order 2, as well as linear functions at the two endpoints of the fitting interval. Given these conditions, we can now derive the estimated curve $\hat{\boldsymbol{g}}$ as the solution to (2) at the knots in matrix form.

Theorem 1. *Denote the knot–response data pairs as $(t_1, y_1), \ldots, (t_n, y_n)$, the knots vector $\boldsymbol{T} = (t_1, \ldots, t_n)^T$ and the response vector $\boldsymbol{y} = (y_1, \ldots, y_n)^T$. Assume for each interval $[t_i, t_{i+1}]$, $i = 1, \ldots, n-1$, there exists λ_i, such that*

$$\int_{t_i}^{t_{i+1}} \lambda(t)[g''(t)]^2 dt = \lambda_i \int_{t_i}^{t_{i+1}} [g''(t)]^2 dt. \tag{3}$$

At the knots $\boldsymbol{T}$, an estimator of $\boldsymbol{g}$ that solves the minimization problem in Eq. (2) is

$$\hat{\boldsymbol{g}} = (\boldsymbol{I_n} + \eta \boldsymbol{\Delta} \boldsymbol{G}^{-1} \boldsymbol{G}^* \boldsymbol{G}^{-1} \boldsymbol{\Delta}^T)^{-1} \boldsymbol{y}, \tag{4}$$

where $\eta = \frac{4}{3}\rho$, $h_i = t_{i+1} - t_i$, $\boldsymbol{\Delta}$ is a matrix of dimension $n \times (n-2)$ with non-zero elements

$$\delta_{i,i} = \frac{1}{h_i}, \ \delta_{i+1,i} = -\left(\frac{1}{h_i} + \frac{1}{h_{i+1}}\right), \ \delta_{i+2,i} = \frac{1}{h_{i+1}};$$

$\boldsymbol{G}$ is a matrix of dimension $(n-2) \times (n-2)$ with non-zero elements

$$g_{i,i} = \frac{1}{3}(h_i + h_{i+1}), \ g_{i,i+1} = g_{i+1,i} = \frac{1}{6} h_{i+1};$$

$\boldsymbol{G}^*$ is a matrix of dimension $(n-2) \times (n-2)$ with non-zero elements

$$g^*_{i,i} = \frac{1}{3}(h_i \lambda_i + h_{i+1} \lambda_{i+1}), \ g^*_{i,i+1} = g^*_{i+1,i} = \frac{1}{6} h_{i+1} \lambda_{i+1}.$$

Proof. See Appendix A. □

Corollary 1. *When the prior curvature function is constant, Eq. (4) simplifies to the solution for natural cubic spline (Green and Silverman, 1993, Chapter 2).*

Proof. Since the prior curvature function is constant, $\lambda(t) = \lambda_i = c,\ t \in \mathcal{T},\ i = 1, \ldots, n-1$, where c denotes a constant value. Then $\boldsymbol{G}^* = c\boldsymbol{G}$. The estimator in Eq. (4) is

$$\begin{aligned}\hat{\boldsymbol{g}} &= (\boldsymbol{I_n} + \eta\boldsymbol{\Delta}\boldsymbol{G}^{-1}\boldsymbol{G}^*\boldsymbol{G}^{-1}\boldsymbol{\Delta}^T)^{-1}\boldsymbol{y} \\ &= (\boldsymbol{I_n} + \eta c\boldsymbol{\Delta}\boldsymbol{G}^{-1}\boldsymbol{G}\boldsymbol{G}^{-1}\boldsymbol{\Delta}^T)^{-1}\boldsymbol{y} \\ &= (\boldsymbol{I_n} + \eta'\boldsymbol{\Delta}\boldsymbol{G}^{-1}\boldsymbol{\Delta}^T)^{-1}\boldsymbol{y},\end{aligned}$$

where $\eta' = \eta c$ is a smoothing parameter. □

There are a few points to note here. Firstly, in the optimal design problem, we assume that the locations of the design points coincide with the knot locations. Secondly, Pintore et al. (2006) assumed that $\lambda(t)$ was piecewise-constant with jumps at the knots, but we replaced this assumption with (3), hence for each interval $[t_i, t_{i+1}]$ we assume that there exists a 'representative' constant λ_i. This assumption enables us to remove the dependence on λ_i from the integrand, and hence represent $\hat{\boldsymbol{g}}$ in an elegant form. We will discuss how to approximate λ_i from prior knowledge using the curvature $f''(t)$ in Section 3.1.

2.2. Linear mixed model representation

We represent $\boldsymbol{y} = \hat{\boldsymbol{g}} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I_n})$ in a linear mixed model form, similar to Wang (1998) and Verbyla et al. (1999).

Theorem 2. *Using the same definitions of the matrices as in Theorem 1, a re-formulation of (4) is*

$$\begin{aligned}\hat{\boldsymbol{g}} &= \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{Z}\tilde{\boldsymbol{u}}, \\ \tilde{\boldsymbol{u}} &\sim N_{n-2}(\boldsymbol{0}, \gamma\tilde{\boldsymbol{G}}),\end{aligned} \tag{5}$$

where $\gamma = \sigma^2/\eta$, $\tilde{\boldsymbol{G}} = \boldsymbol{G}(\boldsymbol{G}^*)^{-1}\boldsymbol{G}$, $\boldsymbol{X} = (\mathbb{1}, \boldsymbol{T})$ is a matrix of dimension $n \times 2$ whose first column $\mathbb{1}$ is a vector of 1, and the second column $\boldsymbol{T} = (t_1, \ldots, t_n)^T$ is a vector of the design points, and $\boldsymbol{Z} = \boldsymbol{\Delta}(\boldsymbol{\Delta}^T\boldsymbol{\Delta})^{-1}$.

The proof of this result follows the same reasoning as in Verbyla et al. (1999), Appendix A.

In the term $\boldsymbol{Z}\tilde{\boldsymbol{u}}$ of Eq. (5), both the design matrix $\boldsymbol{Z}$ and the random effects $\tilde{\boldsymbol{u}}$ are functions of $\boldsymbol{T}$. Recall that in the design problem, the aim is to minimize the sampling variance of the parameters by choosing the values of the input variables $\boldsymbol{T}$ from the design region, so it is desirable to remove the dependence of $\tilde{\boldsymbol{u}}$ in Eq. (5) on t, by applying the transformation defined in the following corollary.

Corollary 2. *Let $\underline{\boldsymbol{Z}} = \boldsymbol{Z}\tilde{\boldsymbol{G}}^{\frac{1}{2}}$, $\underline{\tilde{\boldsymbol{u}}} = \tilde{\boldsymbol{G}}^{-\frac{1}{2}}\tilde{\boldsymbol{u}}$, where $\tilde{\boldsymbol{G}}^{\frac{1}{2}}$ is the square root of matrix $\tilde{\boldsymbol{G}}$ such that $\tilde{\boldsymbol{G}}^{\frac{1}{2}}\tilde{\boldsymbol{G}}^{\frac{1}{2}} = \tilde{\boldsymbol{G}}$. The equivalent representation of Eq. (5) is*

$$\begin{aligned}\hat{\boldsymbol{g}} &= \boldsymbol{X}\hat{\boldsymbol{\beta}} + \underline{\boldsymbol{Z}}\,\underline{\tilde{\boldsymbol{u}}}, \\ \underline{\tilde{\boldsymbol{u}}} &\sim N_{n-2}(\boldsymbol{0}, \gamma\boldsymbol{I_{n-2}}),\end{aligned} \tag{6}$$

Proof. Firstly, the expressions for the estimator $\hat{\boldsymbol{g}}$ are equivalent in (5) and (6), since

$$\hat{\boldsymbol{g}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \underline{\boldsymbol{Z}}\,\underline{\tilde{\boldsymbol{u}}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{Z}\tilde{\boldsymbol{G}}^{\frac{1}{2}}\tilde{\boldsymbol{G}}^{-\frac{1}{2}}\tilde{\boldsymbol{u}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{Z}\tilde{\boldsymbol{u}}.$$

Secondly, the distributions of the random effects in (5) and (6) are the same, since

$\tilde{\boldsymbol{u}} \sim N_{n-2}(\boldsymbol{0}, \gamma\tilde{\boldsymbol{G}}) \Leftrightarrow \underline{\tilde{\boldsymbol{u}}} \sim N_{n-2}(\boldsymbol{0}, \tilde{\boldsymbol{G}}^{-\frac{1}{2}}\gamma\tilde{\boldsymbol{G}}\tilde{\boldsymbol{G}}^{-\frac{1}{2}}) \Leftrightarrow \underline{\tilde{\boldsymbol{u}}} \sim N_{n-2}(\boldsymbol{0}, \gamma\boldsymbol{I_{n-2}})$. □

2.3. The D-optimality criterion

There exist a range of alphabetic optimality criteria to achieve different design objectives, but in this paper, we focus on the D-optimality criterion. Extensions to other optimality criteria are briefly discussed in Section 6. The D-optimality criterion minimizes the negative log-determinant of the information matrix, and has the geometrical interpretation of minimizing the volume of the ellipsoidal confidence region for the parameters (Atkinson et al., 2007, Chapter 10). The D-optimality criterion is an appropriate choice, because our interest lies in obtaining an accurate estimation of the entire curve, and D-optimality minimizes the overall variance of the parameters that define the spline model. From Eq. (6), the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ and the best linear unbiased predictor (BLUP) of $\boldsymbol{u}$ are

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{y}, \\ \underline{\tilde{\boldsymbol{u}}} &= (\underline{\boldsymbol{Z}}^T\underline{\boldsymbol{Z}} + \gamma\boldsymbol{I_{n-2}})^{-1}\underline{\boldsymbol{Z}}^T(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}),\end{aligned} \tag{7}$$

where $\boldsymbol{H} = \sigma^2(\boldsymbol{I_{n-2}} + (1/\gamma)\underline{\boldsymbol{Z}}^{\boldsymbol{T}}\underline{\boldsymbol{Z}})$ and

$$var\begin{bmatrix}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \tilde{\underline{\boldsymbol{u}}} - \underline{\boldsymbol{u}}\end{bmatrix} = \sigma^2\begin{bmatrix}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} & \boldsymbol{X}^{\boldsymbol{T}}\underline{\boldsymbol{Z}} \\ \underline{\boldsymbol{Z}}^{\boldsymbol{T}}\boldsymbol{X} & \underline{\boldsymbol{Z}}^{\boldsymbol{T}}\underline{\boldsymbol{Z}} + \eta\boldsymbol{I_{n-2}}\end{bmatrix}^{-1} \tag{8}$$

Hooks et al. (2009) derived the D-optimality criterion for the linear mixed model which incorporating both fixed and random effects. Denoting the information matrix in (8) by

$$\boldsymbol{M} = \begin{bmatrix}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} & \boldsymbol{X}^{\boldsymbol{T}}\underline{\boldsymbol{Z}} \\ \underline{\boldsymbol{Z}}^{\boldsymbol{T}}\boldsymbol{X} & \underline{\boldsymbol{Z}}^{\boldsymbol{T}}\underline{\boldsymbol{Z}} + \eta\boldsymbol{I_{n-2}}\end{bmatrix},$$

then the D-optimal design points $\boldsymbol{T}^*$ are found by solving

$$\begin{aligned}&\min_{\boldsymbol{T}} -\log(\det(\mathbf{M})) \\ \Leftrightarrow &\min_{\boldsymbol{T}} -\log(\det(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})\det(\underline{\boldsymbol{Z}}^{\boldsymbol{T}}\underline{\boldsymbol{Z}} + \eta\boldsymbol{I_{n-2}} - \underline{\boldsymbol{Z}}^{\boldsymbol{T}}\boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\underline{\boldsymbol{Z}}))\end{aligned} \tag{9}$$

The variables $\boldsymbol{T}$ are highly nonlinear in the optimization problem in (9), which complicates the evaluation of the gradient as required in the optimization routine. Therefore we use the Hooke–Jeeves derivative-free algorithm as implemented in the R package 'dfoptim' (Varadhan et al., 2016) to solve the optimization problem, without requiring evaluation of the gradient. Moreover, we introduce a set of slack variables to avoid singularity and impose constraints on the design points. If we let $s_1 = t_1, s_2 = t_2 - t_1, \ldots, s_n = t_n - t_{n-1}$, then we can impose the box constraints $s_i > \delta_i,\ i = 2, \ldots, n$, to guarantee that any two successive design points are separated by a meaningful distance, for example, two successive measurements t_i and t_{i+1} cannot be repeated within time δ_i.

We also add a penalization term $e^{k_2(\sum_{i=1}^n s_i - k_1)}$ (with k_2 taken to be a large positive number) to the objective function in (9) to prevent the last design point $t_n = \sum_{i=1}^n s_i$ exceeding the upper bound k_1 of the design region. For all the simulations and data applications in this paper, we chose $\delta_i = 0.001$ as the minimal distance between two successive measurements, $k_1 = 1$ as the upper bound of the design region and $k_2 = 500$ as an adequately large number. Now the optimization problem becomes

$$\begin{aligned}&\min_{\boldsymbol{S}} -\log\big(\det(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})\det(\underline{\boldsymbol{Z}}^{\boldsymbol{T}}\underline{\boldsymbol{Z}} + \eta\boldsymbol{I_{n-2}} - \underline{\boldsymbol{Z}}^{\boldsymbol{T}}\boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\underline{\boldsymbol{Z}})\big) + e^{k_2(\sum_{i=1}^n s_i - k_1)} \\ &\text{s.t. } s_i > \delta_i,\ i = 2, \ldots, n,\end{aligned} \tag{10}$$

with the optimization variables $\boldsymbol{S} = (s_1, \ldots, s_n)^T$. Then, the optimal design points $\boldsymbol{T}^*$ can be obtained by back-transformation the slack variables $\boldsymbol{S}^*$ found as the solution to (10).

3. Prior information

From our formulation (9), the optimal design is influenced by the inputs $\lambda(t)$ (embedded in $\underline{\boldsymbol{Z}}$), η and n. In this section we discuss some practical issues in estimating $\lambda(t)$ from historical data and when choosing the values of the smoothing parameter η and the number of design points n.

3.1. Estimation of $\lambda(t)$

Recall from Section 2, $\lambda(t) = \frac{1}{[f''(t)]^2}$, where $f''(t)$ is the prior curvature function chosen to reflect the expected behaviour of the curve. Previous studies have suggested some nonlinear parametric forms as surrogates for the plant growth modelling (Paine et al., 2012), in which case such a model can be assumed as prior information and the second derivative can be computed analytically. Alternatively, historical data from a similar experiment could be used to estimate the prior curvature information for the current design problem. In such cases, second order differences could be used to estimate the second derivatives at discrete time points t_i

$$f''(t_i) \approx \left(\frac{y_{i+1} - y_i}{t_{i+1} - t_i} - \frac{y_i - y_{i-1}}{t_i - t_{i-1}}\right)/\left(\frac{t_{i+1} - t_{i-1}}{2}\right), \tag{11}$$

and $f''(t)$ is assumed piece-wise constant between the discrete time points. With slightly more effort, a smooth curve fit could be estimated using a basis function expansion and hence used to estimate $f''(t)$ function. The curve $f(t)$ can be represented by K basis functions $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)^T$ as follows

$$f(t) = \sum_{k=1}^{K} c_k\phi_k(t) = \boldsymbol{c}^{\boldsymbol{T}}\boldsymbol{\phi}, \tag{12}$$

where the basis functions are often chosen to be the Fourier basis for periodic data or a B-spline basis for non-periodic data. Let Φ be the $n \times K$ matrix that contains the basis functions for all the observations, then the least squares estimates of the coefficients $\boldsymbol{c}$ are given by

$$\hat{\boldsymbol{c}} = (\boldsymbol{\Phi}^{\boldsymbol{T}}\boldsymbol{\Phi})^{-\mathbf{1}}\boldsymbol{\Phi}^{\boldsymbol{T}}\boldsymbol{y}. \tag{13}$$

Alternatively, a smoothed curve can be fitted using a penalization approach. The curve is expanded by basis functions as in (12) with a roughness penalty

$$\text{PEN} = \int_0^1 D^{(m)} f(s) ds = \boldsymbol{c}^{\boldsymbol{T}} [\int_0^1 D^{(m)} \boldsymbol{\phi}(s) D^{(m)} \boldsymbol{\phi}^{\boldsymbol{T}}(s) ds] \boldsymbol{c} = \boldsymbol{c}^{\boldsymbol{T}} \boldsymbol{P} \boldsymbol{c}, \tag{14}$$

where $D^{(m)}$ is the mth order derivative operator and the coefficients $\boldsymbol{c}$ can be estimated as

$$\hat{\boldsymbol{c}} = (\boldsymbol{\Phi}^{\boldsymbol{T}} \boldsymbol{\Phi} + \lambda_f \boldsymbol{P})^{-1} \boldsymbol{\Phi}^{\boldsymbol{T}} \boldsymbol{y}, \tag{15}$$

where λ_f is the smoothing parameter in $\boldsymbol{f}$.

Evaluating the second derivative of the curve is straightforward given a basis function expansion, since $f''(t)$ can be computed by taking the second derivative of the basis functions multiplied by the coefficients $\hat{\boldsymbol{c}}$ from (13) or (15)

$$f''(t) = D^2 \hat{\boldsymbol{c}}(\boldsymbol{t}) = \hat{\boldsymbol{c}}^{\boldsymbol{T}} D^2 \phi(t). \tag{16}$$

After obtaining the curvature estimate $f''(t)$, a 'representative' λ_i between two design points $[t_i, t_{i+1}]$ in Eq. (3) can be obtained by

$$\lambda_i = 1 / \left(\frac{\int_{t_i}^{t_{i+1}} [f''(t)]^2 dt}{t_{i+1} - t_i} \right). \tag{17}$$

For reasons of numerical stability and ease of choosing the smoothing parameter η, we divide $\lambda(t)$ by its maximum value in the design region as $\frac{\lambda(t)}{\max_{t \in \mathcal{T}} \lambda(t)}$, so that $\lambda(t)$ is between 0 and 1.

3.2. Choosing η

The tuning parameter η ($=\frac{4}{3}\rho$) in (9) controls the trade-off between goodness-of-fit and weighted smoothness of the curve. Decreasing η allows the prior curvature information to more strongly influence the construction of the optimal design. Therefore, a smaller value of η should be chosen when there is a strong belief that the curvature of the data to be collected will closely follow the prior curvature. We show in Simulation 1, that the design points are relatively insensitive to the choice of η for small to moderate values. However for large η, the design points are pushed towards the boundaries of the design region. In other simulations and real data examples, we fix $\eta = 1$. It is recommended to perform a sensitivity analysis over a range of η.

3.3. Choosing n

The optimal design problem also depends on the number of design points n. In this paper we treat n as fixed and vary the locations of the design points in the optimization problem. Treating n as another parameter to be determined when optimizing the location of the design points turns out to be a difficult problem for the following reasons. Firstly, for a parametric model with p parameters, the general equivalence theorem (Kiefer, 1974) shows that a D-optimal design has at most $p(p-1)/2$ support points, and the number of minimally supported points is p. However there are infinite number of parameters in a nonparametric model, so that the general equivalence theorem cannot be applied here. Secondly, the matrix $\boldsymbol{M}$ in (9) which defines D-optimality changes its dimension and the composition of fixed effects and random effects as the number of design points changes (Verbyla, 2019). Therefore, the log-determinant of $\boldsymbol{M}$ is not comparable between different n. While we do not propose a principled approach to choosing n, it is suggested to plot the prior curvature function to assess its shape, since intuitively, more design points will be needed for curves exhibiting complicated behaviour than for simple curves.

It is expected that the goodness-of-fit increases with more data points, however, improvements can be negligible as n increases beyond a certain value. Furthermore, as n increases, the gain in goodness-of-fit from the optimal design is often marginal when compared to the uniform design having the same n, therefore, considering the optimal design is the most beneficial when n is relatively small. Empirical support for these assertions is provided in Section 4.3 using the Berkeley growth data for females. In addition, the choice of n will be restricted by practical limitations of the data collection process, particularly cost considerations. Similar to the choice of η, it is recommended that designs for a range of n values be explored, in order to ensure that the selected n will provide sufficient coverage across the design region, particularly in areas of high curvature.

4. Simulations and real data applications

Here we conduct two simulation studies and apply our method to the Berkeley growth data for females. In the first simulation, we assume the growth curve follows a three-parameter logistic model, and derive optimal designs for a range of values of the smoothing parameter η and the number of design points n. The special case assuming constant curvature is presented in Appendix B, which reduces to the formulation of Dette et al. (2011) when $m = 2$. In the second simulation, we generate a more complicated growth curve that cannot be modelled by a simple nonlinear parametric model. In the simulations, we compare the optimal design with the uniform design in terms of the distribution of the design points and the goodness-of-fit. The application to the Berkeley growth data for females is presented in Section 4.3, and for males in Appendix C.

Table 1
Simulation 1, optimal design points $\boldsymbol{T_o}$ and uniform design points $\boldsymbol{T_u}$ for $n \in \{4, 5, 6\}$ and $\eta \in \{0.1, 0.5, 1, 10, 100, 1000\}$.

η	n	$\boldsymbol{T_u}$	$\boldsymbol{T_o}$
0.1	4	0.00 0.33 0.67 1.00	0 0.36 0.64 1
0.5			0 0.36 0.64 1
1			0 0.36 0.64 1
10			0 0.37 0.63 1
100			0 0.36 0.64 1
1000			0 0 1 1
0.1	5	0 0.25 0.50 0.75 1	0 0.31 0.50 0.69 1
0.5			0 0.31 0.50 0.69 1
1			0 0.31 0.50 0.69 1
10			0 0.31 0.50 0.69 1
100			0 0 0.5 1 1
1000			0 0 1 1 1
0.1	6	0 0.2 0.4 0.6 0.8 1	0 0.31 0.50 0.65 0.77 1
0.5			0 0.23 0.32 0.49 0.69 1
1			0 0.24 0.36 0.64 0.76 1
10			0 0.26 0.33 0.67 0.74 1
100			0 0 0.37 0.63 1 1
1000			0 0 0 1 1 1

4.1. Simulation 1: logistic growth curve — effect of η and n on the distribution of design points

We consider the 3-parameter logistic curve

$$f(t) = \frac{\beta_1}{1 + e^{\beta_2 t + \beta_3}}, \tag{18}$$

where β_1 is the asymptotic plateau, and $-\beta_3/\beta_2$ is the location of the point of inflection of the curve on the time axis. Taking the second derivative of $f(t)$, $\lambda(t)$ has the closed form solution

$$\lambda(t) = \frac{1}{[f''(t)]^2} = \frac{[1 + e^{\beta_2 t + \beta_3}]^6}{\beta_1^2 \beta_2^4 [e^{\beta_2 t + \beta_3} - e^{2(\beta_2 t + \beta_3)}]^2}. \tag{19}$$

In this simulation, the parameters were set as $\beta_1 = 1, \beta_2 = -10, \beta_3 = 5$, and optimal designs were obtained for all combinations of $\eta \in \{0.1, 0.5, 1, 10, 100, 1000\}$ and $n \in \{4, 5, 6\}$. We compare each of the optimal designs to the corresponding uniform design, in which the n design points are equally spanned over [0,1].

Table 1 lists the distributions of the uniform design $\boldsymbol{T_u}$ and the optimal design $\boldsymbol{T_o}$ under different simulation settings. For fixed n, increasing η moves the optimal design points towards the two boundaries, whereas decreasing η anchors the optimal design points at the locations with largest curvature values. This is also illustrated in Fig. 1 where we plot the logistic curve and its corresponding curvature function, and superpose the uniform and optimal design points for $n = 6$ and $\eta \in \{1, 100, 1000\}$. The optimal design points are symmetrically distributed about the inflection point, because the curvature function is antisymmetric about this point. All of the designs include the two boundaries as design points to capture the starting point and the asymptote of the curve. For $\eta = 1000$ (very large), all the design points are located at the boundaries (three on each side); when $\eta = 100$ (moderate), two of the design points are placed at the locations associated with the largest curvature values and four at the two ends; when $\eta = 1$ (small), two interior points are still placed at the large curvature locations with another two points not far away from them and two points at the boundaries.

4.2. Simulation 2: logistic growth curve with perturbation — comparing goodness-of-fit between optimal and uniform designs

We consider the mixture parameter model

$$f(t) = \frac{\beta_1}{1 + e^{\beta_2 t + \beta_3}} - 0.02 \frac{1}{\sigma_f} e^{\frac{-(t-\mu)^2}{2\sigma_f^2}}. \tag{20}$$

The curvature function is more complicated, but still has a closed form expression, and $\lambda(t)$ can be computed analytically as $\lambda(t) = \frac{1}{[f''(t)]^2} = 1/\left[\frac{\beta_1 \beta_2^2 \left(e^{\beta_2 t + \beta_3} - e^{2(\beta_2 t + \beta_3)}\right)}{(1 + e^{\beta_2 t + \beta_3})^3} - \frac{0.02^2}{\sigma_f^4} e^{\frac{-(t-\mu)^2}{2\sigma_f^2}} \left(\frac{(t-\mu)^2}{\sigma_f^2} - 1\right)\right]^2$.

The parameters were chosen as $\beta_1 = 1$, $\beta_2 = -10$, $\beta_3 = 5$, $\sigma_f = 0.002$, $\mu = 0.7$. Fig. 2 plots this growth curve, which mimics a scenario in which the growth of a plant is impacted by a stress treatment halfway through the experiment, before normal growth conditions are resumed. In this simulation we fixed $\eta = 1$ and obtained optimal and uniform design points for $n \in \{8, 9, 10, 11, 12\}$. For each set of design points, we simulated data from $f(t) + N(0, 0.01^2)$ at each design

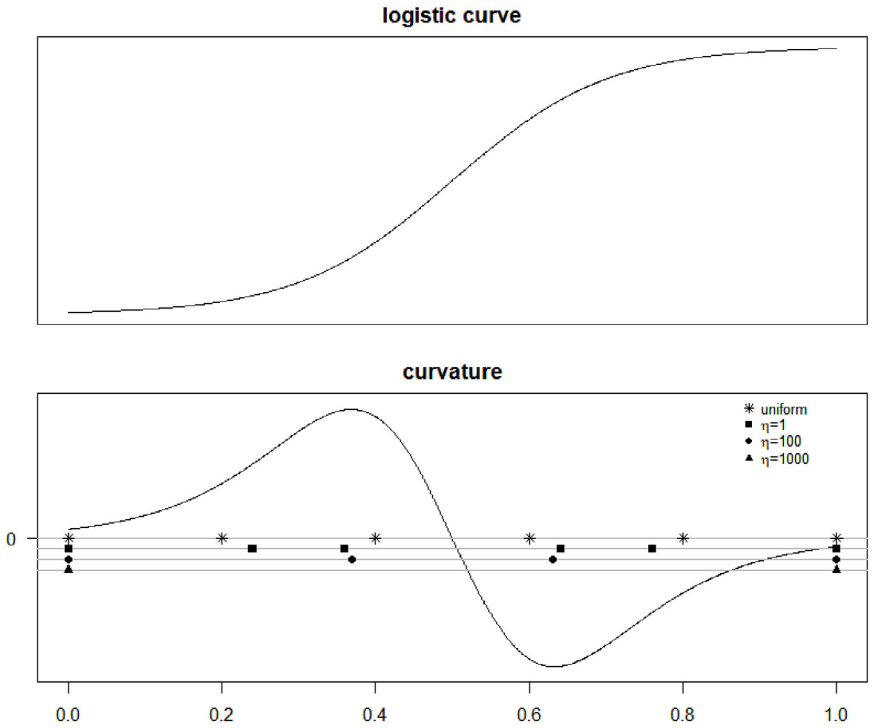

Fig. 1. Upper panel: the logistic curve. Lower panel: the corresponding curvature with uniform and optimal design points for $n = 6$ and $\eta \in \{1, 100, 1000\}$. Note that two points and three points coincide at each of the two boundaries when $\eta = 100$ and $\eta = 1000$ respectively.

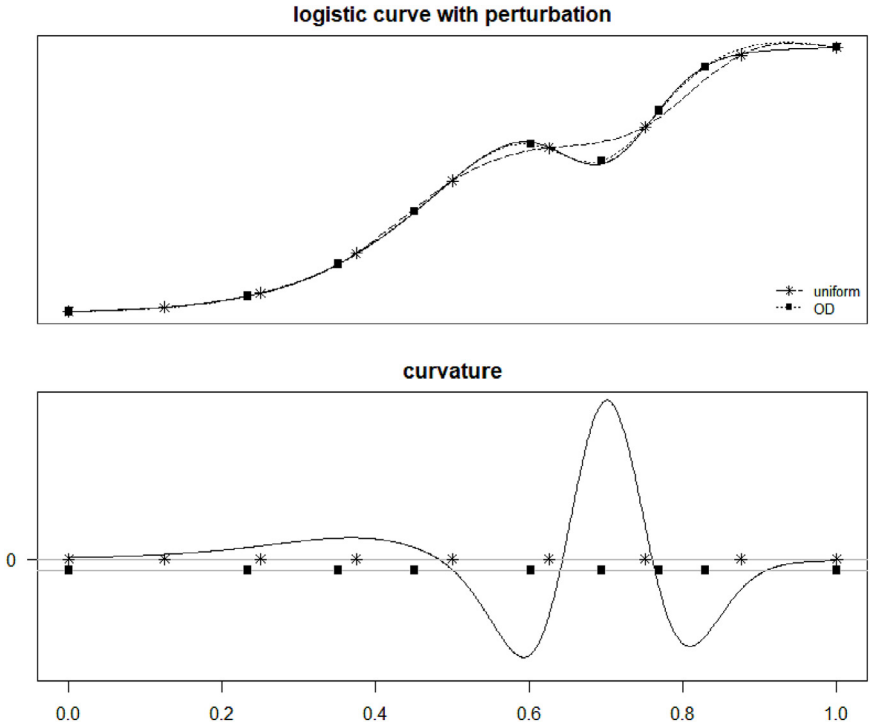

Fig. 2. Upper panel: the true logistic curve with perturbation and estimated smoothing splines with knots located at the uniform and optimal design points respectively, for $\eta = 1$, $\texttt{spar} = 0.1$ and $n = 9$. Lower panel: the corresponding curvature function with superposed uniform and optimal design points.

Table 2
Simulation 2, optimal and uniform design points $\boldsymbol{T}$ and mean square error (MSE) between the estimated curve and the true curve for $\eta = 1$, spar $\in \{0.1, 0.2\}$ and $n \in \{8, 9, 10, 11, 12\}$. A lower MSE implies a better goodness-of-fit.

n	Method	$\boldsymbol{T}$	Spar	MSE (*E+04)
8	OD	0.00 0.23 0.35 0.45 0.61 0.70 0.81 1.00	0.1	1.076
			0.2	4.435
	Uniform	0.00 0.14 0.29 0.43 0.57 0.71 0.86 1.00	0.1	3.569
			0.2	4.817
9	OD	0 0.23 0.35 0.45 0.60 0.69 0.77 0.83 1	0.1	0.787
			0.2	2.029
	Uniform	0 0.12 0.25 0.38 0.50 0.62 0.75 0.88 1	0.1	7.791
			0.2	9.680
10	OD	0 0.19 0.29 0.37 0.45 0.58 0.64 0.71 0.81 1	0.1	1.635
			0.2	1.168
	Uniform	0 0.11 0.22 0.33 0.44 0.56 0.67 0.78 0.89 1	0.1	3.055
			0.2	4.706
11	OD	0 0.19 0.29 0.37 0.45 0.58 0.64 0.70 0.77 0.83 1	0.1	0.700
			0.2	1.038
	Uniform	0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1	0.1	0.960
			0.2	1.282
12	OD	0 0.16 0.25 0.32 0.39 0.46 0.58 0.64 0.70 0.77 0.83 1	0.1	0.676
			0.2	0.967
	Uniform	0 0.09 0.18 0.27 0.36 0.45 0.55 0.64 0.73 0.82 0.91 1	0.1	0.783
			0.2	1.697

point, then used these data to fit cubic smoothing splines over the range [0,1] using the function 'smooth.spline' in R (R Core Team, 2018) for two choices of the smoothing parameter spar $\in \{0.1, 0.2\}$. The knots were chosen to be the same as the sampling (design) points. We repeated generating data 100 times. The mean square error (MSE) of the curve was defined as

$$\text{MSE} = \frac{1}{101}\sum_{i=1}^{101}(\hat{f}(t_i) - f(t_i))^2, \tag{21}$$

which averages the squared differences between the estimated cubic spline fit $\hat{\boldsymbol{f}}$ and the true curve $\boldsymbol{f}$ over 101 equally spaced grid points at $t_i = 0, 0.01, \ldots, 0.99, 1$, which provides a measure of the goodness-of-fit over the entire curve.

Fig. 2 compares the distributions of the optimal design points with the uniform design points, and the estimated curves with knots located at these two sets of design points respectively. The upper panel shows that the true curve (solid line) has a dip around $t = 0.7$ which is not captured by the uniform design points, but is captured by the optimal design points when fitting smoothing splines with spar $= 0.1$ and $n = 9$. As a consequence, the estimated smooth curve for the uniform design points fails to accurately capture the behaviour of the data in the vicinity of the stress treatment around $t = 0.7$, whereas the estimated smooth curve resulting from the optimal design points adheres well to the true curve. The lower panel of the plot displays the optimal and uniform design points, and the optimization of the design has selected points that lie close to the peaks in the curvature function.

Table 2 confirms that the MSEs for the optimal designs are smaller than those from the uniform designs with the same n and spar, indicating a better goodness-of-fit by sampling at the optimal design points.

4.3. Berkeley growth data for females

In real data applications, the prior curvature is unknown and hence needs to be estimated from historical data. In this section we derive an optimal design for the growth curve of females using the Berkeley growth data (Tuddenham, 1954).

In the Berkeley growth data set, the heights of 54 girls were measured at 31 unequally spaced ages. Four measurements were taken between ages one and two, followed by six measurements annually up to age eight, then two measurements per year until the individual reached eighteen years of age (Fig. 3). A prior curvature function for use in developing an optimal design was estimated by the penalized smoothing spline method as described in Eqs. (14)–(16). We penalized the derivatives of order three ($m = 3$ in Eq. (14)) and used the B-spline basis functions of order six to fit the smoothing spline. The tuning parameter λ_f was chosen by cross-validation. Because the timing and intensity of the growing trajectories differed between individuals, we performed a continuous registration to improve the estimation of curvature by minimizing the second eigenvalue for the matrix defined by the original curve and the registered curve (Ramsay and Silverman, 2005, Chapter 8). As noted in the lower panel of Fig. 3, the curvature values are amplified for some ages after registration, and the registered curvature is used as the prior knowledge to derive the optimal design. The large negative

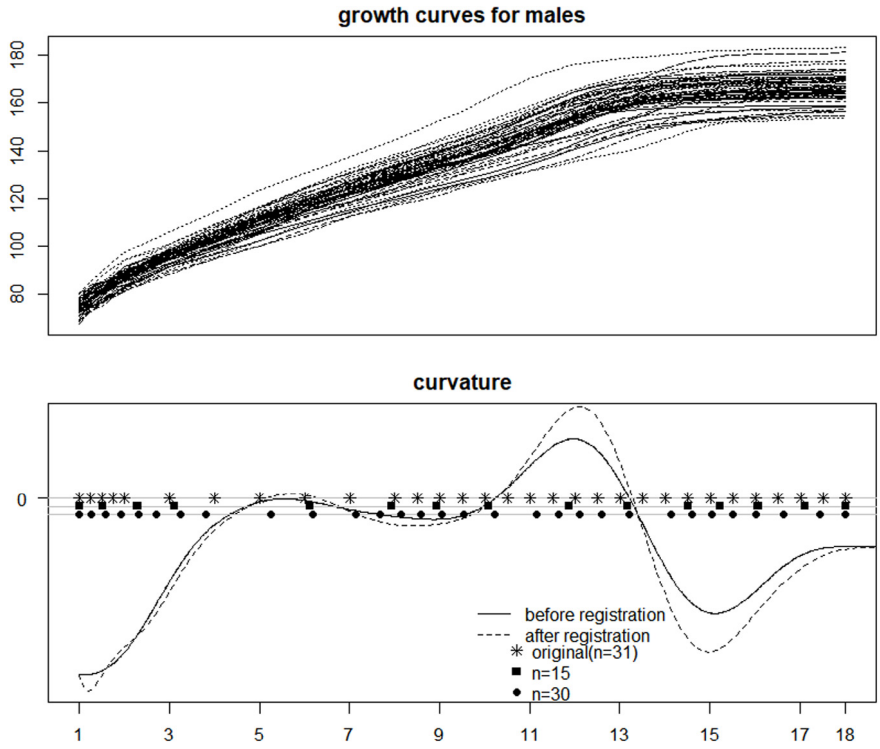

Fig. 3. Upper panel: height measurements for 54 girls in the Berkeley growth study data set. Lower panel: curvature estimates before and after registration, with superposed original sampling points ($n = 31$) and optimal design points for $\eta = 1$ and $n \in \{15, 30\}$.

curvature at age one reflects a sharp deceleration in growth at a very early age. There is a dip around age six followed by a positive acceleration peak around age ten. The maximum pubertal growth rate occurs at age eleven where the acceleration drops to zero and another negative acceleration peak happens at age thirteen. This data set shows that on average girls stop growing around age seventeen when the curvature approaches zero.

The original sampling points ($n = 31$) and the optimal design points for $n = \{15, 30\}$ are marked in the lower panel of Fig. 3, along with the estimated curvature functions derived from the Berkeley data. For $n = 15$, the optimal design suggests that measurements be taken four times before age three and five times between ages five and eleven, followed by five denser measurements between ages twelve and fifteen and the last measurement at age eighteen. When more measurements are allowed ($n = 30$), the optimal design suggests more data be collected before age three and between ages twelve and fifteen where the curvature values are largest, and slightly more data around ages seven and ten where the second and the third peak of the curvature are located. Although it is not recorded how the researchers designed the data collection protocol for the Berkeley growth study, it appears reasonable that they collected data four times between ages one to two, then only once per year between ages two to eight. However, it seems that collecting data biannually was not needed after age 16, because most of the girls stopped growing or grew at a very slow pace.

We also used this data set to examine the effect of the number of design points on goodness-of-fit. We fitted a smoothing spline to the data from all of the ages in the original data set and treated this as the true curve. Then we derived the optimal and uniform design points for n varying from seven to twenty. The corresponding heights at those design points were obtained from the true curve. The respective growth curve was estimated from those age–height data pairs at the optimal and uniform design points. We computed the MSE between the estimated curve and the true curve over 101 equally spaced grids, similar to Section 4.2. Fig. 4 shows that the MSE for an optimal design is always smaller than that for the corresponding uniform design across all n values. In this application, taking height measurements at fifteen ages seems to be adequate to accurately capture the growth curve, since the further reduction of MSE is marginal for $n > 14$. Fig. 4 also suggests that applying an optimal design can achieve efficiency since, for example, to reduce MSE below 0.05, nine data points from the optimal design are required compared to eleven points from the uniform design.

A similar graph to Fig. 3 obtained from the Berkeley growth data for males is presented in Appendix C (Figure S1). The registered curvature profile for males is similar to that for females but with a phase lag and an amplitude enlargement (Fig. 6). This is due to the phenomenon that boys tend to delay their pubertal growth compared to their female counterparts, but do grow taller eventually. As with the designs derived from the females, Figure S1 shows that more points are placed over the regions that exhibit higher variations, however, more points are allocated after age sixteen compared to the optimal design for girls. This motivated us to consider the optimal design in the presence of subpopulations where their respective curvatures show different patterns discussed in the next section.

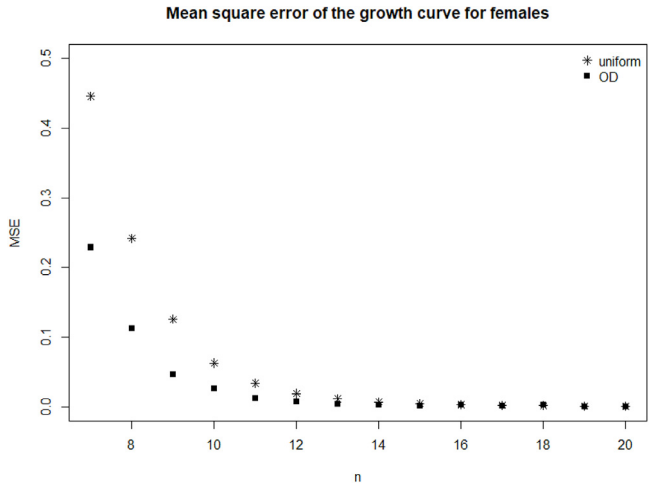

Fig. 4. Mean square error (MSE) between the assumed true Berkeley growth curve for females, and growth curves estimated from optimal and uniform design points obtained for $n \in \{7, \ldots, 20\}$. A lower MSE implies a better goodness-of-fit.

5. Optimal design with subpopulations

Here we derive the optimal design with the existence of subpopulations. The term 'subpopulation' can refer to some grouping of data that display different growth patterns, for example, heights of male and female children as they grow, or changes in phenotypic observations for different cultivars over time. This is non-trivial since it is sometimes not feasible/practical to derive the optimal design for each subpopulation and collect data separately, so instead we assume the measurements are to be taken at the same time points for every subpopulation and derive a design that is optimal for the entire population.

5.1. The D-optimality criterion

Following a similar derivation to that in Section 2, we represent the adaptive smoothing spline in the form of a linear mixed model. To avoid cluttered notation we will only present the development for two subpopulations, but extending to more than two subpopulations is straightforward. For the jth subpopulation, $j \in \{1, 2\}$, the model is

$$y_{ji} = g_j(t_i) + \epsilon_{ji}, \tag{22}$$

and the respective prior curvature functions are $f_j''(t)$, so that the weighted smoothing parameters are $\lambda_j(t) = 1/[f_j''(t)]^2$. Using similar notation to Eq. (5), but adding the index j to distinguish between the two subpopulations, the adaptive smoothing spline for subpopulation j in the form of linear mixed model is

$$\begin{aligned} &\boldsymbol{y_j} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{Z}\tilde{\boldsymbol{u}}_{\boldsymbol{j}} + \boldsymbol{\epsilon_j}, \\ &\tilde{\boldsymbol{u}}_{\boldsymbol{j}} \sim N(\boldsymbol{0}, \gamma\tilde{\boldsymbol{G}}_{\boldsymbol{j}}), \;\; \boldsymbol{\epsilon_j} \sim N(\boldsymbol{0}, \sigma_j^2\boldsymbol{I_n}), \end{aligned} \tag{23}$$

where $\tilde{\boldsymbol{G}}_{\boldsymbol{j}} = \boldsymbol{G}(\boldsymbol{G}_{\boldsymbol{j}}^*)^{-1}\boldsymbol{G}$.

Similar to Verbyla et al. (1999) where they considered a qualitative treatment factor, we stack the two models for the two subpopulations and then make the transformation as in Corollary 2. The combined model is

$$\begin{aligned} \boldsymbol{y} &= \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{X} \end{bmatrix} \hat{\boldsymbol{\beta}} + \begin{bmatrix} \boldsymbol{Z} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{u}}_{\boldsymbol{1}} \\ \tilde{\boldsymbol{u}}_{\boldsymbol{2}} \end{bmatrix} + \boldsymbol{\epsilon} \\ &= \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{X} \end{bmatrix} \hat{\boldsymbol{\beta}} + \begin{bmatrix} \underline{\boldsymbol{Z}}_{\boldsymbol{1}} & \boldsymbol{0} \\ \boldsymbol{0} & \underline{\boldsymbol{Z}}_{\boldsymbol{2}} \end{bmatrix} \begin{bmatrix} \underline{\tilde{\boldsymbol{u}}}_{\boldsymbol{1}} \\ \underline{\tilde{\boldsymbol{u}}}_{\boldsymbol{2}} \end{bmatrix} + \boldsymbol{\epsilon}, \\ &= \boldsymbol{\mathcal{X}}\hat{\boldsymbol{\beta}} + \boldsymbol{\mathcal{Z}}\,\underline{\tilde{\boldsymbol{u}}} + \boldsymbol{\epsilon} \\ \underline{\tilde{\boldsymbol{u}}} &= \begin{bmatrix} \underline{\tilde{\boldsymbol{u}}}_{\boldsymbol{1}} \\ \underline{\tilde{\boldsymbol{u}}}_{\boldsymbol{2}} \end{bmatrix} \sim N(\boldsymbol{0}, \gamma\boldsymbol{I}_{\boldsymbol{2(n-2)}}) \\ \boldsymbol{\epsilon} &= \begin{bmatrix} \boldsymbol{\epsilon}_{\boldsymbol{1}} \\ \boldsymbol{\epsilon}_{\boldsymbol{2}} \end{bmatrix} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{\boldsymbol{2n}}) \end{aligned} \tag{24}$$

where $\underline{\boldsymbol{Z}}_{\boldsymbol{j}} = \boldsymbol{Z}\tilde{\boldsymbol{G}}_{\boldsymbol{j}}^{\boldsymbol{1/2}}$, $\underline{\tilde{\boldsymbol{u}}}_{\boldsymbol{j}} = \tilde{\boldsymbol{G}}_{\boldsymbol{j}}^{\boldsymbol{-1/2}}\tilde{\boldsymbol{u}}_{\boldsymbol{j}}$, $\boldsymbol{\mathcal{X}} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{X} \end{bmatrix}$, $\boldsymbol{\mathcal{Z}} = \begin{bmatrix} \underline{\boldsymbol{Z}}_{\boldsymbol{1}} & \boldsymbol{0} \\ \boldsymbol{0} & \underline{\boldsymbol{Z}}_{\boldsymbol{2}} \end{bmatrix}$.

Table 3
Simulation 3, optimal and uniform design points $\boldsymbol{T}$, and squared bias values of inflection points for the two logistic curves, for $\eta = 1$ and $n \in \{6, 7, 8, 9\}$. A lower bias value implies a more accurate estimate of the inflection point.

n	Design	$\boldsymbol{T}$	Bias inflection 1 (*E+05)	Bias inflection 2 (*E+05)
6	Spline	0 0.28 0.42 0.58 0.81 1	1.03	4.00
	Uniform	0. 0.2 0.4 0.6 0.8 1	1.30	4.10
7	Spline	0 0.27 0.42 0.54 0.65 0.82 1	0.93	2.95
	Uniform	0 0.17 0.33 0.50 0.67 0.83 1	1.03	4.35
8	Spline	0 0.21 0.29 0.42 0.54 0.65 0.82 1	1.01	3.23
	Uniform	0 0.14 0.29 0.43 0.57 0.71 0.86 1	1.07	3.56
9	Spline	0 0.21 0.29 0.42 0.54 0.64 0.78 0.87 1	0.97	2.69
	Uniform	0 0.12 0.25 0.38 0.50 0.62 0.75 0.88 1	0.94	3.21

In Eq. (24), we assume the linear trend $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ is the same for the two groups but the random effects differ depending on $\lambda_j(t)$ which is embedded in the matrix $\tilde{\boldsymbol{G}}_{\boldsymbol{j}}$. We also assume that the variance parameters σ_j^2 are the same for the two groups, so that $\gamma = \sigma^2/\eta$. The variance–covariance of the parameters $(\hat{\boldsymbol{\beta}}, \tilde{\underline{\boldsymbol{u}}})$ is

$$var\begin{bmatrix}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \tilde{\underline{\boldsymbol{u}}} - \underline{\boldsymbol{u}}\end{bmatrix} = \sigma^2 \begin{bmatrix}\boldsymbol{X^T X} & \boldsymbol{X^T Z} \\ \boldsymbol{Z^T X} & \boldsymbol{Z^T Z} + \eta \boldsymbol{I_{2(n-2)}}\end{bmatrix}^{-1} = \sigma^2 \boldsymbol{\mathcal{M}^{-1}}. \tag{25}$$

The optimal design under the D-optimality criterion is obtained by solving

$$\min_{\boldsymbol{T}} -\log(\det(\boldsymbol{\mathcal{M}})). \tag{26}$$

5.2. Simulation 3: two logistic growth curves — compare bias of inflection points between optimal design and uniform design

We now consider two subpopulations having growth curves generated using the logistic curve (18) with different parameters. When considering sigmoidal growth curves, a parameter of interest to biologists is the location of the inflection point where the curvature changes its sign, and the maximum growth rate is achieved.

The simulation settings are summarized as follows.

- Curve 1: $\beta_1 = 1, \beta_2 = -12, \beta_3 = 5$; inflection 1 $= -\frac{\beta_3}{\beta_2} = 0.417$
- Curve 2: $\theta_1 = 1, \theta_2 = -9, \theta_3 = 6$; inflection 2 $= -\frac{\theta_3}{\theta_2} = 0.667$

We fixed $\eta = 1$ and varied the number of design points $n \in \{6, 7, 8, 9\}$. In a similar fashion to Simulation 2, we simulated data from the true curves at the design points (optimal and uniform). We fitted a logistic curve using the 'nls' function in R to each subpopulation and calculated the respective inflection point. We define the bias as the square of the difference between the estimated inflection point and the true value, hence the bias in the inflection point for subpopulation 1 is calculated as $(-\frac{\hat{\beta}_3}{\hat{\beta}_2} - 0.417)^2$, and for subpopulation 2 as $(-\frac{\hat{\theta}_3}{\hat{\theta}_2} - 0.667)^2$. Table 3 reports the mean bias values obtained from 100 sets of simulated data. Fig. 5 plots the two logistic curves and the distribution of the design points (uniform and optimal) along with the corresponding curvature functions for $n = 7$.

It can be noted from Fig. 5 that there is a delay in the growth in subpopulation 2 compared with subpopulation 1, and hence the inflection point occurs at a later time in subpopulation 2. The interior design points have been positioned near the large curvature values from both populations. From Table 3, the biases of both inflection points are smaller when estimated from the optimal design than from the uniform design, with the exception of the inflection point in subpopulation 1 for $n = 9$.

5.3. Berkeley growth data for males and females

Lastly we derived an optimal design from the Berkeley growth data incorporating both males and females subpopulations. The prior curvature function for each population was estimated separately following the same procedure as described in Section 4.3. Fig. 6 plots the curvature functions for both populations after registrations. The different patterns in the curvature functions strongly imply a need to take the subpopulation structure into account when optimizing a design to collect data, to ensure we accurately capture the relevant features of both female and male growth curves. The female-only design (solid circles) distributes design points more densely between ages twelve and fifteen, whereas points in the male-only design (solid squares) are more evenly distributed across the region. The optimal design derived for both subpopulations (asterisks) appears to achieve a satisfactory compromise between the female-only and male-only designs.

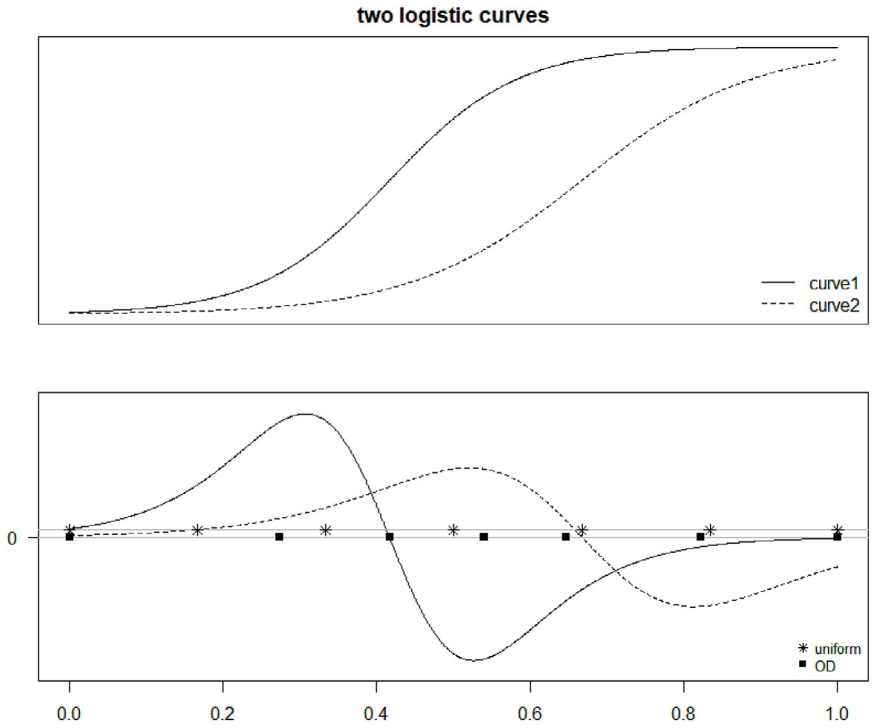

Fig. 5. Upper panel: the two logistic curves with different set of parameters. Lower panel: the corresponding curvature functions with superposed uniform and optimal design points for $\eta = 1$ and $n = 7$.

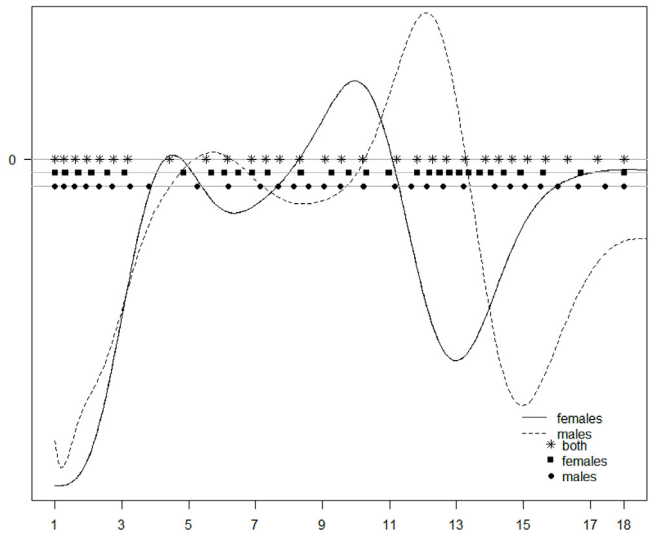

Fig. 6. Estimated curvature functions obtained after registration for males and females, with superposed optimal design points for female-only, male-only and both subpopulations for $\eta = 1$ and $n = 30$.

6. Conclusion

In this paper, we derived a method for obtaining D-optimal designs for temporal data. The prior knowledge is given by the curvature which can be interpreted as the weighted smoothness in the adaptive smoothing spline. The mathematical convenience of expressing the estimator of the curve in a linear mixed model form enables us to derive an analogue of information matrix from a linear model. An extension is made for when there exist subpopulations whose curvatures are known to have different patterns a priori. Three simulation studies are carried out to demonstrate the influence of

the choice of the smoothing parameter and the number of design points, and we use independent criteria to compare the optimal design with the commonly adopted uniform design. We also use the Berkeley growth data as a real data example to show the practicality of performing the optimal design when the curvature information is unknown but can be estimated from historical data.

There are a few extensions that could be made from the general framework described in this paper. Firstly, other criteria could be considered in addition to the D-optimality criterion to achieve different goals in a design. Denote by $\boldsymbol{M}$ the information matrix which is a function of a design $\boldsymbol{T}$, a more general form of the design optimization criterion is: $\min_{\boldsymbol{T}} \Psi(\boldsymbol{M})$, where Ψ is an operator that maps the information matrix to a real value. The D-optimality criterion takes the form $\Psi(\boldsymbol{M}) = -\log|\boldsymbol{M}|$, which is the most commonly used among a list of alphabetic criteria, see for example Chapter 10 in Atkinson et al. (2007). Dette et al. (2008) derived the optimal designs for D- and G-optimality criteria, where the former aims for a more precise estimation of the coefficients in the spline model while the latter aims to provide an accurate prediction of the curve, and it was shown through simulations that the G-optimal design was more sensitive to the smoothing parameter.

Secondly, a more general variance–covariance structure can be considered. In model (1), we assume the observations are independent and identically distributed with a common variance parameter σ^2. We could specify a more complicated $\boldsymbol{R}$ matrix to take into account heterogeneity and dependence of the residuals. For example in the Berkeley growth data, the variation in children's height is more pronounced at young age, and there are positive correlations in heights between neighbouring ages. However, adding more flexibility in the variance–covariance matrix comes at a cost of the necessity to provide prior values for more parameters, increasing the difficulty in the design problem.

Thirdly, experimental designs can be developed in a sequential manner. Suppose there is initially little knowledge about the data, a good starting point is to conduct a study with a uniform design to collect some data, and then use these data to fit a curve. The estimated curvature from this study can then be used as the prior knowledge to perform optimal design for a subsequent run of the experiment. The procedure of designing an experiment, collecting data and fitting a curve can be repeated until a satisfactory result is reached. Another more challenging task is to produce the sequential design in a single experimental run, in other words, to estimate future sampling points $(t_{m+1}, \ldots, t_n)$ after collecting data at $(t_1, \ldots, t_m)$. This is different from the more general sequential construction of an optimal design (Tsay, 1976) where the design points could be anywhere within the design region, but for temporal data the subsequent design points must be chosen from a constraint region after time t_m.

Acknowledgement

The authors would like to thank Prof. Eric Stone for his helpful discussions and comments.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jspi.2019.10.002.

References

Atkinson, A., Donev, A., Tobias, R., 2007. Optimum Experimental Designs, with SAS, Vol. 34. Oxford University Press.
Chaloner, K., Verdinelli, I., 1995. Bayesian experimental design: A review. Statist. Sci. 273–304.
Dette, H., Melas, V.B., Pepelyshev, A., 2008. Optimal designs for free knot least squares splines. Statist. Sinica 1047–1062.
Dette, H., Melas, V.B., Pepelyshev, A., 2011. Optimal design for smoothing splines. Ann. Inst. Statist. Math. 63 (5), 981–1003.
DiMatteo, I., Genovese, C.R., Kass, R.E., 2001. Bayesian curve-fitting with free-knot splines. Biometrika 88 (4), 1055–1071.
Donev, A.N., Tobias, R., Monadjemi, F., 2008. Cost-cautious designs for confirmatory bioassay. J. Statist. Plann. Inference 138 (12), 3805–3812.
Green, P.J., Silverman, B.W., 1993. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. CRC Press.
Gu, Y., Jin, Z., 2013. Neighborhood preserving d-optimal design for active learning and its application to terrain classification. Neural Comput. Appl. 23 (7–8), 2085–2092.
He, X., 2009. Laplacian regularized d-optimal design for active learning and its application to image retrieval. IEEE Trans. Image Process. 19 (1), 254–263.
Heiligers, B., 1998. E-optimal designs for polynomial spline regression. J. Stat. Plan. Inference 75 (1), 159–172.
Hooks, T., Marx, D., Kachman, S., Pedersen, J., 2009. Optimality criteria for models with random effects. Rev. Colombiana Estadíst. 32 (1), 17–31.
Kaishev, V., 1989. Optimal experimental designs for the b-spline regression. Comput. Statist. Data Anal. 8 (1), 39–47.
Kiefer, J., 1974. General equivalence theory for optimum designs (approximate theory). Ann. Statist. 849–879.
Li, G., 2012. Optimal and efficient designs for gompertz regression models. Ann. Inst. Statist. Math. 64 (5), 945–957.
Li, G., Majumdar, D., 2008. D-optimal designs for logistic models with three and four parameters. J. Statist. Plann. Inference 138 (7), 1950–1959.
Miyata, S., Shen, X., 2003. Adaptive free-knot splines. J. Comput. Graph. Stat. 12 (1), 197–213.
Montgomery, D.C., 2017. Design and Analysis of Experiments. John wiley & sons.
Paine, C.T., Marthews, T.R., Vogt, D.R., Purves, D., Rees, M., Hector, A., Turnbull, L.A., 2012. How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. Methods Ecol. Evol. 3 (2), 245–256.
Park, S.H., 1978. Experimental designs for fitting segmented polynomial regression models. Technometrics 20 (2), 151–154.
Pintore, A., Speckman, P., Holmes, C.C., 2006. Spatially adaptive smoothing splines. Biometrika 93 (1), 113–125.
R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis. Springer.
Tsay, J.-Y., 1976. On the sequential construction of d-optimal designs. J. Amer. Statist. Assoc. 71 (355), 671–674.

Tuddenham, R.D., 1954. Physical growth of california boys and girls from birth to eighteen years. Univ. Calif. Publ. Child Dev. 1, 183–364.
Varadhan, R., Borchers, H.W., Varadhan, M.R., 2016. Package 'dfoptim'.
Verbyla, A.P., 2019. A note on model selection using information criteria for general linear models estimated using reml. Aust. N. Z. J. Stat. 61 (1), 39–50.
Verbyla, A.P., Cullis, B.R., Kenward, M.G., Welham, S.J., 1999. The analysis of designed experiments and longitudinal data by using smoothing splines. J. R. Stat. Soc. Ser. C. Appl. Stat. 48 (3), 269–311.
Wahba, G., 1990. Spline Models for Observational Data, Vol. 59. Siam.
Wang, Y., 1998. Mixed effects smoothing spline analysis of variance. J. R. Stat. Soc. Ser. B Stat. Methodol. 60 (1), 159–174.