Chapter

Computationally Efficient Kalman Filter Approaches for Fitting Smoothing Splines

Joel Parker, Yifan Zhang, Bonnie J. Lafleur and Xiaoxiao Sun

Abstract

Smoothing spline models have shown to be effective in various fields (e.g., engineering and biomedical sciences) for understanding complex signals from noisy data. As nonparametric models, smoothing spline ANOVA (Analysis Of variance) models do not fix the structure of the regression function, leading to more flexible model estimates (e.g., linear or nonlinear estimates). The functional ANOVA decomposition of the regression function estimates offers interpretable results that describe the relationship between the outcome variable, and the main and interaction effects of different covariates/predictors. However, smoothing spline ANOVA (SS-ANOVA) models suffer from high computational costs, with a computational complexity of $O(N^3)$ for N observations. Various numerical approaches can address this problem. In this chapter, we focus on the introduction to a state space representation of SS-ANOVA models. The estimation algorithms based on the Kalman filter are implemented within the SS-ANOVA framework using the state space representation, reducing the computational costs significantly.

Keywords: Kalman filter, smoothing spline, functional ANOVA, state space representation, Markov structure

1. Introduction

Smoothing spline ANOVA (SS-ANOVA) has been widely used in various applications [1–3]. The representer theorem enables an exact solution of regression function in SS-ANOVA models by minimizing a regularized function in a finite-dimensional space, even though the problem resides in a infinite-dimensional space. While SS-ANOVA models have strong theoretical properties, the estimation algorithms used to fit these models are computational intensive, with a computational complexity of $O(N^3)$ for datasets with N observations. Numerous approaches have been developed to reduce the heavy computational costs of SS-ANOVA [4–7]. For example, Kim and Gu (2004) proposed to select a $q \ll N$ basis functions from N ones and reduced the computational complexity to $O(Nq^2)$. Sun et al. (2021) synergistically combined asymptotic results with the smoothing parameter estimates based on randomly

selected samples with sizes of $\tilde{N} \ll N$ to reduce the computational complexity of selecting smoothing parameters to $O\left(\tilde{N}^3\right)$.

In this book chapter, we focus on the estimation approaches based on the Kalman filter. The Kalman filter was originally created to solve linear filtering and prediction problems used to generate simulations for the Apollo 11 project [8]. More recently, it has been implemented in a variety of engineering and biomedical fields [9, 10]. The Kalman filter is naturally used to fit state space models, methods that use recursive calculations on each observation entered one at a time and resulting in calculations on more accurate unknown variables after each iteration. The Kalman filter updates the state of the dynamic system given a new observation based on the state after the previous observation and the information gained from the new observation. This memory-less property and its simple recursive formulas make Kalman filter approaches computationally efficient, making them a useful tool for big data analytics. SS-ANOVA models can be reformulated to a state space representation, allowing computationally efficient Kalman filter-based model fitting and reducing computational costs to $O(N)$ for estimating univariate smoothing spline models [11]. An extension to the bivariate setting also significantly reduces the computational costs of SS-ANOVA models [12, 13].

Section 2 of this chapter will provide the theoretical background of SS-ANOVA models. Section 3 provides a brief background on state space models. The state space representation of SS-ANOVA models can be found in Section 4, along with a simulation study under the univariate setting in Section 5. Section 6 concludes this chapter.

2. Smoothing spline ANOVA models

We assume the data $(y_i, \mathbf{x}_i)$ and $i = 1,2, \ldots N$ are independent and identically distributed where $y_i \in Y \in \mathbb{R}$ is the outcome/response variable and $\mathbf{x}_i \in \mathcal{X} \in \mathbb{R}^d$ represents the covariates/predictors. A nonparametric model can then be written by

$$y_i = f(\mathbf{x}_i) + e_i, \tag{1}$$

where f is a function of covariates and $e_i \sim \mathcal{N}(0, \sigma^2)$ represents the random errors. For this nonparametric model, the structure of f is not fixed and can be estimated by minimizing a penalized least squares score,

$$\frac{1}{N}\sum_{i=1}^{N}\left(y_i - f(\mathbf{x}_i)\right)^2 + \lambda J(f), \tag{2}$$

where the first term measures the goodness of fit of f, and the smoothing parameter λ controls the trade-off between the goodness of fit and the roughness of f measured by $J(f)$ [2, 3, 14]. SS-ANOVA models can also handle responses from exponential families and/or correlated responses. For readers who are interested in these topics, more examples of the model estimation and implementation exist (e.g., [2, 14]). The nonparametric estimation allows f to vary in a high-dimensional (possibly infinite) space leading to more flexible results that can balance the bias-variance trade-off [2, 15]. The functional analysis of variance (ANOVA) is applied to the regression function f to improve the interpretability of model estimates by

decomposing the function into main and interaction effects of covariates. These main and interaction effects can be estimated in the corresponding subspaces of the reproducing kernel Hilbert space (RKHS), which is introduced in the next section.

2.1 ANOVA

2.1.1 Classical ANOVA

Classical ANOVA can be used to help to understand the decomposition of regression function in (1). We use a one-way classical ANOVA model as an example. The outcome y_i can be modeled by $y_{ij} = \mu_i + e_{ij}$, where μ_i is the mean treatment levels with $i = 1,2,\cdots,K_1$ and $j = 1,2,\cdots,K_2$. The terms in this model can be rewritten as

$$y_{ij} = \mu + \delta_i + e_{ij}, \tag{3}$$

where μ is the overall mean effect and δ_i is the treatment effect. Side conditions are added to ensure the uniqueness of this decomposition. Now consider the univariate nonparametric function in (1). The regression function can be written as

$$f(x) = Af + (I - A)f = f_0 + f_1 \tag{4}$$

where A is an averaging operator that averages the effect of x and I is the identity operator. We also need to add some side conditions for this decomposition to ensure the uniqueness of the decomposition of the regression function.

2.1.2 Functional ANOVA

The multivariate function $f\left(x_{\langle 1\rangle}, x_{\langle 2\rangle}, \ldots, x_{\langle d\rangle}\right)$ on a d-dimensional product domain $\mathcal{X} = \prod_{j=1}^{d}\mathcal{X}_j \in \mathcal{R}^d$ can be decomposed similarly to the classical ANOVA in the RKHS. The construction of the RKHS on $\Pi_{j=1}^{d}\mathcal{X}_j$ is by taking the tensor product over the marginal domains $\mathcal{X}_j$. We need the following theorem to construct the tensor-product space.

Theorem 1.1 If $R_1\left(x_{\langle 1\rangle}, \tilde{x}_{\langle 1\rangle}\right)$ is nonnegative definite on $\mathcal{X}_1$ and $R_2\left(x_{\langle 2\rangle}, \tilde{x}_{\langle 2\rangle}\right)$ is nonnegative definite on $\mathcal{X}_2$, then $R\left(x_{\langle 1\rangle}, x_{\langle 2\rangle}\right) = R_1\left(x_{\langle 1\rangle}, \tilde{x}_{\langle 1\rangle}\right)R_2\left(x_{\langle 2\rangle}, \tilde{x}_{\langle 2\rangle}\right)$ is nonnegative definite on $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$.

Theorem 1.1 implies that the RKHS $\mathcal{H}$ on $\Pi_{j=1}^{d}\mathcal{X}_j$ has the reproducing kernel $R = \Pi_{j=1}^{d}R_{\langle j\rangle}$, where $R_{\langle j\rangle}$ is the reproducing kernel for $\mathcal{H}_{\langle\gamma\rangle}$ on $\mathcal{X}_{\langle\gamma\rangle}$. Additionally, the Hilbert space $\mathcal{H}_{\langle j\rangle}$ can be decomposed into $\mathcal{H}_{\langle j\rangle} = \mathcal{H}_{\langle j\rangle(0)} \oplus \mathcal{H}_{\langle j\rangle(1)}$, where $\mathcal{H}_{\langle j\rangle(0)}$ is the null space and $\mathcal{H}_{\langle j\rangle(1)}$ is the orthogonal complement to $\mathcal{H}_{\langle j\rangle(0)}$. Then $\mathcal{H} = \otimes_{j=1}^{d}\mathcal{H}_{\langle j\rangle}$ can be decomposed as

$$\begin{aligned}\mathcal{H} &= \otimes_{j=1}^{d}\left(\mathcal{H}_{\langle j\rangle(0)} \oplus \mathcal{H}_{\langle j\rangle(1)}\right)\\ &= \oplus_{\mathcal{S}}\left\{\left(\otimes_{j\in\mathcal{S}}\mathcal{H}_{\langle j\rangle(1)}\right) \otimes \left(\otimes_{j\notin\mathcal{S}}\mathcal{H}_{\langle j\rangle(0)}\right)\right\}\\ &= \oplus_{\mathcal{S}}\mathcal{H}_{\mathcal{S}}\end{aligned} \tag{5}$$

where $\mathcal{S}$ denotes all of the subsets of $\{1, \ldots, d\}$. The term $\mathcal{H}_{\mathcal{S}}$ has the reproducing kernel $R_{\mathcal{S}} \propto \Pi_{j\in\mathcal{S}}R_{\langle j\rangle(1)}$.

In general, the inner product in $\mathcal{H}$ can be specified as

$$J(f,g)=\sum_{j=1}^{B}\theta_j^{-1}\left\langle f_j,g_j\right\rangle_j \tag{6}$$

where $\theta_j\geq 0$ are tuning parameters and B is the number of smoothing parameters. The roughness penalty in (2) can be written in the form of (6). Then the reproducing kernel associated with (6) can be written as

$$R=\sum_{j=1}^{B}\theta_j R_j, \tag{7}$$

where R_j is the reproducing kernel for the corresponding tensor product RKHS in (5).

2.2 An example of RKHS on $[0, 1]^2$

We will use one example of an RKHS on $[0, 1]^2$ to demonstrate the decomposition of RKHS. More examples of discrete and/or continuous domains can be found in Gu (2013) [2]. We consider the following tensor sum decomposition on $[0, 1]$,

$$\begin{aligned}\mathcal{H}_{\langle 1\rangle}&=\left\{f:\int_0^1\left(f^{(2)}(x)\right)^2dx<\infty\right\}\\&=\{f:f\propto 1\}\oplus\{f:f\propto k_1(x)\}\\&\oplus\left\{f:\int_0^1 f dx=\int_0^1 f^{(1)}dx=\int_0^1 f^{(2)}dx=0,f^{(2)}\in\mathscr{L}_2[0,\ 1]\right\}\\&=\mathcal{H}_{\langle 1\rangle(00)}\oplus\mathcal{H}_{\langle 1\rangle(01)}\oplus\mathcal{H}_{\langle 1\rangle(1)},\end{aligned} \tag{8}$$

where $\mathcal{H}_{\langle 1\rangle 01}\oplus\mathcal{H}_{\langle 1\rangle 1}$ forms the contrast in the one-way ANOVA decomposition, and the function k_r is a scaled Bernoulli polynomial function with $k_r(x)=B_r(x)/r!$ [2, 16]. The RKHS has three reproducing kernels $R_{\langle 1\rangle 00}(x,\tilde{x})=1$, $R_{\langle 1\rangle 01}(x,\tilde{x})=k_1(x)k_1(\tilde{x})$, and $R_{\langle 1\rangle 1}(x,\tilde{x})=k_2(x)k_2(\tilde{x})-k_4(|x-\tilde{x}|)$.

Now consider the RKHS $\mathcal{H}$ on $[0, 1]\times[0, 1]$. Here $\mathcal{H}$ can be the tensor product spaces of $\mathcal{H}_{\langle 1\rangle}$ on $[0, 1]$ and $\mathcal{H}_{\langle 2\rangle}$ on $[0, 1]$. Based on the tensor sum decomposition in (8), we have

$$\begin{aligned}\mathcal{H}&=\mathcal{H}_{\langle 1\rangle}\otimes\mathcal{H}_{\langle 2\rangle}\\&=\left(\mathcal{H}_{\langle 1\rangle(00)}\oplus\mathcal{H}_{\langle 1\rangle(01)}\oplus\mathcal{H}_{\langle 1\rangle(1)}\right)\otimes\left(\mathcal{H}_{\langle 2\rangle(00)}\oplus\mathcal{H}_{\langle 2\rangle(01)}\oplus\mathcal{H}_{\langle 2\rangle(1)}\right)\\&=\left(\mathcal{H}_{\langle 1\rangle(00)}\otimes\mathcal{H}_{\langle 2\rangle(00)}\right)\oplus\left(\mathcal{H}_{\langle 1\rangle(00)}\otimes\mathcal{H}_{\langle 2\rangle(01)}\right)\oplus\left(\mathcal{H}_{\langle 1\rangle(01)}\otimes\mathcal{H}_{\langle 2\rangle(00)}\right)\\&\oplus\left(\mathcal{H}_{\langle 1\rangle(01)}\otimes\mathcal{H}_{\langle 2\rangle(01)}\right)\oplus\left(\mathcal{H}_{\langle 1\rangle(00)}\otimes\mathcal{H}_{\langle 2\rangle(1)}\right)\oplus\left(\mathcal{H}_{\langle 1\rangle(1)}\otimes\mathcal{H}_{\langle 2\rangle(00)}\right)\\&\oplus\left(\mathcal{H}_{\langle 1\rangle(01)}\otimes\mathcal{H}_{\langle 2\rangle(1)}\right)\oplus\left(\mathcal{H}_{\langle 1\rangle(1)}\otimes\mathcal{H}_{\langle 2\rangle(01)}\right)\oplus\left(\mathcal{H}_{\langle 1\rangle(1)}\otimes\mathcal{H}_{\langle 2\rangle(1)}\right),\end{aligned} \tag{9}$$

where the first four terms in (9) are in the null space $\mathcal{H}_{\langle\mathbf{0}\rangle}$ and the remaining five terms are in the orthogonal complement (i.e., $\mathcal{H}_{\langle\mathbf{1}\rangle}$). The reproducing kernel for the cubic spline (m=2) for each subspace can be found in **Table 1**.

Subspace	Reproducing Kernel
$\mathscr{H}_{\langle 1\rangle(00)} \otimes \mathscr{H}_{\langle 2\rangle(00)}$	1
$\mathscr{H}_{\langle 1\rangle(01)} \otimes \mathscr{H}_{\langle 2\rangle(00)}$	$k_1(x_{\langle 1\rangle})k_1(\tilde{x}_{\langle 1\rangle})$
$\mathscr{H}_{\langle 1\rangle(01)} \otimes \mathscr{H}_{\langle 2\rangle(01)}$	$k_1(x_{\langle 1\rangle})k_1(\tilde{x}_{\langle 1\rangle})k_1(x_{\langle 2\rangle})k_1(\tilde{x}_{\langle 2\rangle})$
$\mathscr{H}_{\langle 1\rangle(1)} \otimes \mathscr{H}_{\langle 2\rangle(00)}$	$k_2(x_{\langle 1\rangle})k_2(\tilde{x}_{\langle 1\rangle}) - k_4(\lvert x_{\langle 1\rangle} - \tilde{x}_{\langle 1\rangle}\rvert)$
$\mathscr{H}_{\langle 1\rangle(1)} \otimes \mathscr{H}_{\langle 2\rangle(01)}$	$[k_2(x_{\langle 1\rangle})k_2(\tilde{x}_{\langle 1\rangle}) - k_4(\lvert x_{\langle 1\rangle} - \tilde{x}_{\langle 1\rangle}\rvert)]k_1(x_{\langle 2\rangle})k_1(\tilde{x}_{\langle 2\rangle})$
$\mathscr{H}_{\langle 1\rangle(1)} \otimes \mathscr{H}_{\langle 2\rangle(1)}$	$[k_2(x_{\langle 1\rangle})k_2(\tilde{x}_{\langle 1\rangle}) - k_4(\lvert x_{\langle 1\rangle} - \tilde{x}_{\langle 1\rangle}\rvert)][k_2(x_{\langle 2\rangle})k_2(\tilde{x}_{\langle 2\rangle}) - k_4(\lvert x_{\langle 2\rangle} - \tilde{x}_{\langle 2\rangle}\rvert)]$

Table 1.
Subspaces and their corresponding reproducing kernels for the RKHS, $\mathscr{H}$, on $[0, 1] \times [0, 1]$.

2.3 Estimation

The estimation algorithm of SS-ANOVA models relies on the following representer theorem.

Theorem 1.2 (Representer Theorem) There exist coefficient vectors $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_M)' \in \mathcal{R}^M$ and $\mathbf{c} = (c_1, \ldots, c_N)' \in \mathcal{R}^N$ such that the minimizer of (2) in $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ has the following representation:

$$f(\mathbf{x}) = \sum_{m=1}^{M} \xi_m \phi_m(\mathbf{x}) + \sum_{i=1}^{N} c_i R(\mathbf{x}_i, \mathbf{x}), \tag{10}$$

where $\{\phi_m, m = 1, \ldots, M\}$ are the basis functions of the null space $\mathcal{H}_0$ and $R(\cdot, \cdot)$ is the reproducing kernel of $\mathcal{H}_1$.

Taking into consideration model (1), the function f can be estimated by minimizing (2). Using the representer theorem, the function f can be written as

$$\mathbf{f} = \mathbf{S}\boldsymbol{\xi} + \mathbf{Q}\mathbf{c} \tag{11}$$

where $\mathbf{f} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_N))'$, $\mathbf{S}$ is a $N \times M$ matrix (e.g., $M = 2$ for cubic spline) where the $(i, j)^{th}$ entry is $\phi_j(\mathbf{x}_i)$ and $\mathbf{Q}$ is a $N \times N$ matrix where the $(i, j)^{th}$ entry is $R(\mathbf{x}_i, \mathbf{x}_j)$ of the form (7). Plugging (11) into (2), the penalized least squares can be written as

$$\frac{1}{N}(\mathbf{y} - \mathbf{S}\boldsymbol{\xi} - \mathbf{Q}\mathbf{c})'(\mathbf{y} - \mathbf{S}\boldsymbol{\xi} - \mathbf{Q}\mathbf{c}) + \lambda \mathbf{c}'\mathbf{Q}\mathbf{c} \ . \tag{12}$$

We differentiate (12) with respect to $\boldsymbol{\xi}$ and $\mathbf{c}$ to obtain the linear system

$$\begin{aligned} (\mathbf{Q} + N\lambda\mathbf{I})\mathbf{c} + \mathbf{S}\boldsymbol{\xi} &= \mathbf{y}, \\ \mathbf{S}'\mathbf{c} &= 0. \end{aligned} \tag{13}$$

For given smoothing parameters, solving for $\mathbf{c}$ and $\boldsymbol{\xi}$ provides the estimation for f. The selection of smoothing parameters for SS-ANOVA models is introduced below.

2.4 Selection of smoothing parameters

The smoothing parameters $\lambda/\boldsymbol{\theta}$ balance the trade-off between the goodness of fit for f and the roughness of f in (2). Choosing the optimal smoothing parameters is data specific and should be performed prior to nonparametric regression analysis. Several smoothing parameter selection methods have been developed for the SS-ANOVA models [17]. Generalized cross-validation (GCV) is one of the most popular methods for selecting the optimal smoothing parameters $\lambda/\boldsymbol{\theta}$ [18, 19].

To avoid overparameterization, let $\boldsymbol{\lambda} = (\lambda/\theta_1, \ldots \lambda/\theta_B)'$. The GCV score is defined as

$$V(\boldsymbol{\lambda}) = \frac{N^{-1}\mathbf{y}'(I - A(\boldsymbol{\lambda}))^2\mathbf{y}}{\left[N^{-1}tr(I - \alpha A(\boldsymbol{\lambda}))\right]^2} \tag{14}$$

where $A(\boldsymbol{\lambda})$ is symmetric matrix similar to the hat matrix in linear regression, and $tr(\cdot)$ represents trace. The parameter $\alpha \geq 1$ is a fudge factor [4]. When $\alpha = 1$ it is the original GCV score. Larger α's yield smoother estimates. By default, we set $\alpha = 1.4$. Then optimal smoothing parameters $\boldsymbol{\lambda}$ can be chosen by minimizing the GCV score (14) using Newton-Raphson methods.

2.5 Computational complexity

In this section, we will discuss the computation complexity for calculating $\mathbf{c}$ and $\boldsymbol{\xi}$ from (12). One requires $N^3/3 + O(N^2)$ operations to obtain estimates of $\mathbf{c}$ and $\boldsymbol{\xi}$ for the fixed smoothing parameters. In practice, the optimization of the smoothing parameter is also needed, which requires operations of $4BN^3/3 + O(N^2)$, where B is the number of smoothing parameters. Therefore, to minimize (12), the estimation algorithms have a computational complexity of $O(N^3)$. The following sections will discuss how the Kalman filter can be used to fit SS-ANOVA models, which reduces the computation complexity to $O(N)$.

3. State space models

State space methodology was traditionally used to study dynamic problems (e.g., space tracking settings) because the procedure allows for "real-time" updating as data are collected [20]. In this chapter, we use the linear Gaussian state space model as an example to introduce concepts of state space models using the Kalman filter approach. More applications of state space models can be found in a study by Douc, Moulines and Stoffer (2014) and Durbin and Koopman (2001) [21, 22]. A state space model consists of two equations: state equation and measurement equation. The state equation describes the dynamics of the state variables:

$$\mathbf{z}_{t+1} = \mathbf{G}_t\mathbf{z}_t + \Psi_t\boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \Delta_t), \tag{15}$$

where $\mathbf{z}_t$ is the $h \times 1$ state vector, $\boldsymbol{\eta}_t$ is the $g \times 1$ disturbance vector with zero mean and a covariance matrix Δ_t, and $\mathbf{G}_t$ and Ψ_t are fixed design matrices of dimensions

$h \times h$ and $h \times g$, respectively. The measurement equation shows the relationship between the observed variable and the unobserved state variable:

$$\mathbf{o}_t = \Phi_t \mathbf{z}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \boldsymbol{V}_t), \tag{16}$$

where $\mathbf{o}_t$ is a $n \times 1$ vector with n observations, $\boldsymbol{\varepsilon}_t$ is a $n \times 1$ disturbance vector with zero mean and a covariance matrix $\mathbf{V}_t$, and Φ_t is a fixed design matrix of dimension $n \times h$. The initial state vector $\mathbf{z}_0$ is assumed to be normally distributed with mean μ_0 and covariance matrix $\mathbf{P}$. The two vectors $\boldsymbol{\eta}_t$ and $\boldsymbol{\varepsilon}_t$ are assumed to be mutually uncorrelated, that is,

$$\begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\varepsilon}_t \end{pmatrix} \overset{\text{iid}}{\sim} \mathcal{N}\left[\mathbf{0}, \begin{pmatrix} \Delta_t & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_t \end{pmatrix}\right]. \tag{17}$$

These two vectors are also uncorrelated to the initial state vector $\mathbf{z}_0$.

The Kalman filter utilizes a set of recursive equations to estimate $\mathbf{z}_t$, given the observations $\mathbf{O}_t = \{\mathbf{o}_1, \ldots, \mathbf{o}_t\}$ at time t and its error variance matrix $\mathbf{P}_t$ [13]. Define

$$\begin{aligned} \hat{\mathbf{z}}_t &= \mathbb{E}[\mathbf{z}_t | \mathbf{O}_t], \\ \mathbf{P}_t &= \mathbb{E}\big[(\mathbf{z}_t - \hat{\mathbf{z}}_t)\big((\mathbf{z}_t - \hat{\mathbf{z}}_t)' | \mathbf{O}_t\big], \end{aligned} \tag{18}$$

where $\hat{\mathbf{z}}_t$ is the Kalman filter estimation of $\mathbf{z}_t$, with $\mathbf{z}_0 = \mathbb{E}[\mathbf{z}_0] = \mu_0$, and $\mathbf{P}_0 = \mathbf{P}$. From the state and measurement Eqs. (15) and (16), the estimated $\mathbf{z}_{t|t-1}$ and the covariance matrix given $\mathbf{z}_{t-1}$, and $\mathbf{P}_{t-1}$ become

$$\begin{aligned} \hat{\mathbf{z}}_{t|t-1} &= \mathbb{E}[\mathbf{z}_t | \mathbf{O}_{t-1}] = \mathbf{G}_t \mathbf{z}_{t-1}, \\ \mathbf{P}_{t|t-1} &= \mathbb{E}\Big[\big(\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1}\big)\Big(\big(\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1}\big)' | \mathbf{O}_{t-1}\Big] = \mathbf{G}_t \mathbf{P}_{t-1} \mathbf{G}_t' + \Psi_t \Delta_t \Psi_t', \end{aligned} \tag{19}$$

and

$$\mathbf{o}_{t|t-1} = \Phi_t \ \hat{\mathbf{z}}_{t|t-1}. \tag{20}$$

Then the prediction error vector $\mathbf{v}_t$ is

$$\mathbf{v}_t = \mathbf{o}_t - \mathbf{o}_{t|t-1} = \mathbf{o}_t - \Phi_t \ \hat{\mathbf{z}}_{t|t-1} = \Phi_t\big(\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1}\big) + \boldsymbol{\varepsilon}_t, \tag{21}$$

with the covariance matrix

$$\mathbb{E}\big[\mathbf{v}_t \mathbf{v}_t'\big] = \Lambda_t = \Phi_t \mathbf{P}_{t|t-1} \Phi_t' + \mathbf{V}_t. \tag{22}$$

Using the facts of the joint distribution of $\mathbf{z}_t$ and $\mathbf{v}_t$, the Kalman filter estimator $\hat{\mathbf{z}}_t = \mathbb{E}(\mathbf{z}_t | \mathbf{O}_t)$ at time t and its covariance matrix can be updated using

$$\begin{aligned} \hat{\mathbf{z}}_t &= \hat{\mathbf{z}}_{t|t-1} + \mathbf{P}_{t|t-1} \Phi_t' \Lambda_t^{-1}\big(\mathbf{o}_t - \Phi_t \hat{\mathbf{z}}_{t|t-1}\big) \\ &= \mathbf{G}_t \mathbf{z}_{t-1} + \mathbf{P}_{t|t-1} \Phi_t' \Lambda_t^{-1} \mathbf{v}_t, \\ \mathbf{P}_t &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \Phi_t' \Lambda_t^{-1} \Phi_t \mathbf{P}_{t|t-1}. \end{aligned} \tag{23}$$

We further define Kalman gain as

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\Phi_t'\Lambda_t^{-1}. \tag{24}$$

Then the filtered estimate of $\mathbf{z}_t$ and its covariance matrix $\mathbf{P}_t$ is

$$\begin{aligned} \hat{\mathbf{z}}_t &= \hat{\mathbf{z}}_{t|t-1} + \mathbf{K}_t\mathbf{v}_t, \\ \mathbf{P}_t &= (\mathbf{I} - \mathbf{K}_t\Phi_t)\mathbf{P}_{t|t-1}. \end{aligned} \tag{25}$$

4. State space representation of SS-ANOVA models

Due to the Markov structure of SS-ANOVA models after reparameterization, the SS-ANOVA models can be represented by the state space models, allowing for efficient estimation by algorithms based on the Kalman filter. Wecker and Ansley (1983) showed the state space representation for univariate smoothing spline models [11]. Such an approach reduces the computational complexity of smoothing splines from $O(N^3)$ to $O(N)$. Based on the fast algorithm for the multivariate Kalman filter, Qin and Guo (2006) extended the univariate case to multivariate SS-ANOVA models [12]. The two-dimensional procedure was implemented, with computational complexity of $O(n_1 n_2^3)$ for data of size $N = n_1 n_2$. The extension to higher dimensions was also discussed. In this chapter, we focus on the univariate setting to demonstrate the procedure to derive the state space representation of smoothing splines.

4.1 Univariate setting

We can use the state space formulation to represent model (1) ($d = 1$) and apply the Kalman filter algorithm to estimate the model parameters of smoothing spline models efficiently. Based on the pioneered work in Wahba (1978) [23], the univariate function $f(x)$ can be written in the following form

$$f(x) = \sum_{\nu=0}^{m-1} \alpha_\nu \frac{(x - x_l)^\nu}{\nu!} + \sqrt{\lambda}\sigma \int_{x_l}^{x} \frac{(x-h)^{m-1}}{(m-1)!} dW(h), \tag{26}$$

where the covariate $x \in [x_l, x_u]$ and $W(h)$ is a Wiener process with the unit dispersion parameter. When all α's have the diffuse prior distribution, the conditional expectation of $f(x)$ given all data is the function minimizing (2) with the smoothing parameter $1/\lambda$. The model (26) can be rewritten as

$$y_i = \mathbf{1}'\mathbf{U}(x_i) + e_i, \quad i = 1,\cdots,N, \tag{27}$$

where $\mathbf{1}' = [1, 0, \cdots, 0]$ and $\mathbf{U}(x) = \left[U^{(m)}(x), \cdots, U^{(1)}(x)\right]'$ is the m-dimensional stochastic process. In particular, we define

$$U^{(j)}(x) = \sum_{\nu=0}^{j-1} U^{(m-\nu)}(x_l) \frac{(x - x_l)^\nu}{\nu!} + \sqrt{\lambda}\sigma \int_{x_l}^{x} \frac{(x-h)^{m-1}}{(m-1)!} dW(h), \quad j = m,\cdots,1. \tag{28}$$

The vector $\mathbf{U}$ contains $U^{(m)}(x)$ and its first $(m-1)$ derivatives. Let $\alpha_\nu = U^{(m-\nu)}(x_l)$, and we can easily verify the model (27).

The state space formulation relies on the Markov structure of $\mathbf{U}(x)$, which is demonstrated below. We define a $m \times m$ matrix, $\Gamma_m(x_b, x_a)$, for any x_b and x_a within the interval $[x_l, x_u]$ as

$$\Gamma_m(x_b, x_a) = \begin{pmatrix} 1 & (x_b - x_a) & \cdots & \frac{(x_b - x_a)^{m-1}}{(m-1)!} \\ & 1 & \cdots & \frac{(x_b - x_a)^{m-2}}{(m-2)!} \\ & & & \vdots \\ & & & 1 \end{pmatrix}. \tag{29}$$

We can easily verify

$$\Gamma_m(x_c, x_a) = \Gamma_m(x_c, x_b)\Gamma_m(x_b, x_a). \tag{30}$$

To show the Markov structure of $\mathbf{U}(x)$, we also define a $m \times 1$ random vector $\omega(x_b, x_a) = \left[\omega^{(1)}(x_b,\ x_a), \cdots, \omega^{(m)}(x_b,\ x_a)\right]'$, where

$$\omega^{(\nu)}(x_b, x_a) = \sqrt{\lambda}\sigma \int_{x_a}^{x_b} \frac{(x_b - h)^{\nu-1}}{(\nu-1)!} dW(h), \quad \nu = 1,\cdots,m. \tag{31}$$

For any $\nu = 1,\cdots,m$, we have

$$\omega^{(\nu)}(x_c, x_a) = \omega^{(\nu)}(x_c, x_b) + \sum_{j=0}^{\nu-1} \frac{(x_c - x_b)^j}{j!} \omega^{(\nu-j)}(x_b, x_a). \tag{32}$$

Thus we have

$$\omega(x_b, x_a) = \Gamma_m(x_c, x_b)\boldsymbol{\omega}(x_b, x_a) + \boldsymbol{\omega}(x_c, x_b). \tag{33}$$

We now apply (30) and (33) to obtain the Markov structure of $\mathbf{U}$

$$\begin{aligned} \mathbf{U}(x_b) &= \Gamma_m(x_b, x_l)\mathbf{U}(x_l) + \omega(x_b, x_l) \\ &= \Gamma_m(x_b, x_a)\Gamma_m(x_a, x_l)\mathbf{U}(x_l) + \Gamma_m(x_b, x_a)\boldsymbol{\omega}(x_a, x_l) + \boldsymbol{\omega}(x_b, x_a) \\ &= \Gamma_m(x_b, x_a)\mathbf{U}(x_a) + \boldsymbol{\omega}(x_b, x_a). \end{aligned} \tag{34}$$

The Markov structure is the key to the state space representation of SS-ANOVA models. For $x_l = x_1 \leq \cdots \leq x_n = x_u$, we have

$$y_i = \sum_{\nu=0}^{m-1} \alpha_\nu \frac{(x_i - x_l)^\nu}{\nu!} + \mathbf{1}'\Omega(x_i) + e_i, \tag{35}$$

as the measurement equation, where $\Omega(x_i) = \left[\Omega^{(m)}(x_i), \cdots, \Omega^{(1)}(x_i)\right]'$ and $\Omega^{(\nu)}(x_i) = \omega^{(\nu)}(x_i, x_l)$ for $\nu = 1,\cdots,m$ and $i = 1,\cdots,N$. From (33), we have the state equation

$$\Omega(x_i) = \Gamma_m(x_i, x_{i-1})\Omega(x_{i-1}) + \boldsymbol{\omega}(x_i, x_{i-1}), \tag{36}$$

for $i = 1,\cdots,N$. Given the parameters λ and σ, the Kalman filtering and smoothing algorithms can be implemented to perform estimation for this state space representation.

5. Simulation studies

We used simulated data to compare the estimates of smoothing splines with those based on state space representation under the univariate setting. The following model was used to simulate $N = 1{,}000$ observations.

$$y_i = 7\sin(\pi^* x_i) + e_i, \tag{37}$$

where $x_i = t_i/100$, $t_i = 1,\cdots,1{,}000$, and $e_i \sim \mathcal{N}(0, 1)$. To apply the univariate SS-ANOVA model to the simulated data, we used the *ssanova* function in the *gss* package (version 2.2-3) [24]. The GCV algorithm was used to select the smoothing parameter of SS-ANOVA models. To implement the Kalman filtering algorithm to fit the smoothing splines, we used Eq. (35) as the measurement equation and Eq. (36) as the state equation. The parameters λ and σ were set to 0.01 and 1, respectively. We fitted the cubic spline (i.e., $m = 2$) in the simulation studies. In the state Eq. (36), we have

$$\Gamma_2(x_i, x_{i-1}) = \begin{pmatrix} 1 & (x_i - x_{i-1}) \\ 0 & 1 \end{pmatrix}, \tag{38}$$

for $i = 2,\cdots,1{,}000$. The $\nu\nu'$th element of the variance matrix of $\boldsymbol{\omega}(x_i, x_{i-1})$ is

$$\lambda\sigma^2 \frac{(x_i - x_{i-1})^{\nu+\nu'-1}}{(\nu+\nu'-1)(\nu-1)!(\nu'-1)!}, \tag{39}$$

where $\nu,\nu' = 1,\cdots,m$. Given the above information, we utilized the *fkf* function in the *FKF* package (version 0.2.3) to implement the Kalman filtering and smoothing

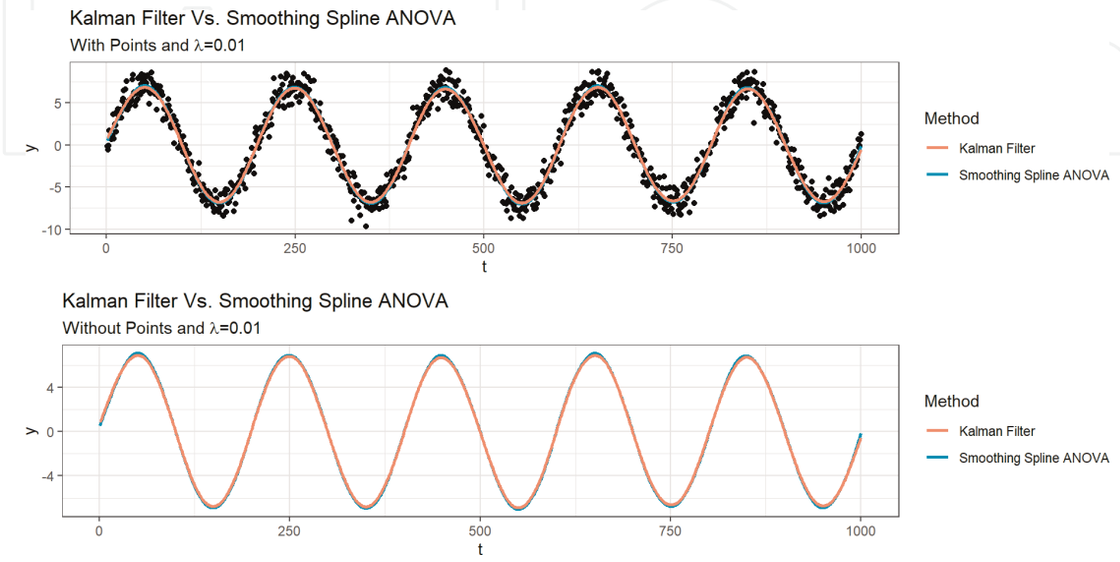

Figure 1.
Comparison between the SS-ANOVA model fit and the model fit with the Kalman filter given $\lambda = 0.01$ on simulated data.

algorithms for the SS-ANOVA model. The details of iteration procedures are available in the help document of *fkf*, which is similar to the procedures described in Section 3. **Figure 1** shows the similarity between the SS-ANOVA model fit with the generic algorithm from the *gss* package and the model fit based on the state space representations in (35) and (36).

6. Conclusions

In this chapter, we have introduced the theoretical foundation (e.g., representer theorem) and estimation algorithms of SS-ANOVA models. Given tensor product operations, the SS-ANOVA models can handle the multivariate data and study the main and interaction effects in the corresponding subspaces via functional ANOVA. The estimation algorithms of SS-ANOVA models need $O(N^3)$ operations, which might be prohibitive computationally for analyzing super large data. Utilizing the Markov structure of SS-ANOVA models, the Kalman filter can be used to fit SS-ANOVA models when reparameterized into a state space formulation [11]. This state space representation reduces the computational complexity from $O(N^3)$ to $O(N)$ for the univariate case, allowing SS-ANOVA models to be applicable to big data applications. Additional research has been done to extend this representation to the multidimensional setting [12]. For the two-dimensional data with the dimensions of n_1 and n_2, the SS-ANOVA models can be fitted with the computational complexity of $O(n_1 n_2^3)$, where $N = n_1 n_2$. Furthermore, we provided a simulated example to compare estimates from the state space representation and the estimates from the SS-ANOVA model for the univariate case.

Acknowledgements

Sun's research was supported by the 18th Mile TRIF Funding from the University of Arizona and the NIH grant U19AG065169. LaFleur's research was supported by the NIH grant U19AG065169.

Conflict of interest

The authors declare no conflict of interest.

Author details

Joel Parker[1,2], Yifan Zhang[1], Bonnie J. Lafleur[2,3] and Xiaoxiao Sun[1*]

1 Department of Epidemiology and Biostatistics, The University of Arizona, Tucson, Arizona, USA

2 BIO5 Institute, The University of Arizona, Tucson, Arizona, USA

3 R. Ken Coit College of Pharmacy, The University of Arizona, Tucson, Arizona, USA

*Address all correspondence to: xiaosun@arizona.edu

References

[1] Brinkman ND. Ethanol fuel-single-cylinder engine study of efficiency and exhaust emissions. SAE Transactions. 1981;**90**:1410-1424

[2] Gu C. Smoothing Spline ANOVA Models. New York, NY: Springer; 2013

[3] Wahba G. Spline Models for Observational Data. Philadelphia, PA: Society for Industrial and Applied Mathematics; 1990

[4] Kim YJ, Gu C. Smoothing spline Gaussian regression: More scalable computation via efficient approximation. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2004;**66**(2):337-356

[5] Ma P, Huang JZ, Zhang N. Efficient computation of smoothing splines via adaptive basis sampling. Biometrika. 2015;**102**(3):631-645

[6] Helwig NE, Ma P. Smoothing spline ANOVA for super-large samples: Scalable computation via rounding parameters. Statistics and Its Interface. 2016;**9**:433-444

[7] Sun X, Zhong W, Ma P. An asymptotic and empirical smoothing parameters selection method for smoothing spline ANOVA models in large samples. Biometrika. 2021;**108**(1): 149-166

[8] Kalman RE. A new approach to linear filtering and prediction problems. Journal of Basic Engineering. 1960;**82**(1): 35-45

[9] Moreno VM, Pigazo A. Kalman Filter: Recent Advances and Applications. London, UK: IntechOpen; 2009

[10] Shumway RH. Longitudinal data with serial correlation: A state-space approach. Journal of Statistical Planning and Inference. 1995;**47**(3):395-397

[11] Wecker WE, Ansley CF. The signal extraction approach to nonlinear regression and spline smoothing. Journal of the American Statistical Association. 1983;**78**(381):81-89

[12] Qin L, Guo W. State space representation for smoothing spline ANOVA models. Journal of Computational and Graphical Statistics. 2006;**15**(4):830-847

[13] Koopman SJ, Durbin J. Fast filtering and smoothing for multivariate state space models. Journal of Time Series Analysis. 2000;**21**(3):281-296

[14] Wang Y. Smoothing Splines : Methods and Applications. Boca Raton, FL: CRC Press; 2011

[15] Yang Y. Model selection for nonparametric regression. Statistica Sinica. 1999;**9**:475-499

[16] Abramowitz M, Stegun IA. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Washington, DC: U.S. Govt. Print. Off; 1964

[17] Wahba G. A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. The Annals of Statistics. 1985;**13**:1378-1402

[18] Xiang D, Wahba G. A generalized approximate cross validation for smoothing splines with non-Gaussian data. Statistica Sinica. 1996:675-692

[19] Graven P, Wahba G. Smoothing noisy data with spline function:

Estimating the correct degree of smoothing by the method of Generalized Cross-Validaton. Numerische Mathematik. 1979;**31**:377-403

[20] De Jong P. The likelihood for a state space model. Biometrika. 1988;**75**(1): 165-169

[21] Douc R, Moulines E, Stoffer D. Nonlinear Time Series: Theory, Methods and Applications with R Examples. Boca Raton, FL: CRC Press; 2014

[22] Durbin J, Koopman SJ. A Simple and Efficient Simulation Smoother for State Space Time Series Analysis. Biometrika, 2002;**89**(3);603-615

[23] Wahba G. Improper priors, spline smoothing and the problem of guarding against model errors in regression. Journal of the Royal Statistical Society: Series B (Methodological). 1978;**40**(3): 364-372

[24] Gu C. Smoothing Spline ANOVA Models: R Package gss. Journal of Statistical Software. 2014;**58**(5):1-25