Metrika (2002) 55: 161–176

Metrika


# Designs for smoothing spline ANOVA models*

**Rong-Xian Yue, Fred J. Hickernell**

College of Mathematical Science, Shanghai Normal University, 100 Guilin Road, Shanghai 200234, China (E-mail: rxyue@online.sh.cn)
Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China (E-mail: fred@hkbu.edu.hk)

Received May 2000

**Abstract.** Smoothing spline estimation of a function of several variables based on an analysis of variance decomposition (SS-ANOVA) is one modern nonparametric technique. This paper considers the design problem for specific types of SS-ANOVA models. As criteria for choosing the design points, the integrated mean squared error (IMSE) for the SS-ANOVA estimate and its asymptotic approximation are derived based on the correspondence between the SS-ANOVA model and the random effects model with a partially improper prior. Three examples for additive and interaction spline models are provided for illustration. A comparison of the asymptotic designs, the $2^d$ factorial designs, and the glp designs is given by numerical computation.

**AMS 1991 subject classification:** 62G07, 62J02, 62K05

**Key words:** Factorial design; Model robust design; Random effects model; Smoothing spline ANOVA; Uniform design

## 1. Introduction

We are interested in the design problem of estimating a multivariate response function $f(\boldsymbol{t})$, $\boldsymbol{t} \in \mathscr{T}$, from observations of $f$ at a discrete set, $\xi_n$, of $n$ points in $\mathscr{T}$ (called the design).

In most practical problems, the true response functions $f$ are usually unknown and may be often very complicated. Then the standard purely parametric methods, such as linear models, quadratic models, or even cubic models are often inadequate for studying the responses. Smoothing spline analysis of

* This work was partially supported by a Hong Kong Research Grants Council Grant RGC/97-98/47, a Shanghai NSF Grant 00JC14057 and a Shanghai Higher Education STF Grant.

variance (SS-ANOVA) is one modern nonparametric technique for modelling or estimating multivariate functions.

In SS-ANOVA methods, suppose that there are observations $y_i$ generated according to the model

$$y_i = f(t_{i1}, \ldots, t_{id}) + \varepsilon_i = f(\boldsymbol{t}_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)' \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ is a Gaussian white noise vector, and $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{id})$ in $\mathscr{T} = \mathscr{T}^{(1)} \otimes \cdots \otimes \mathscr{T}^{(d)}$, the $\mathscr{T}^{(\alpha)}$ are some measurable spaces. The function $f$ is assumed to be in some reproducing kernel Hilbert space, $\mathscr{H}$, and $f$ has an ANOVA decomposition of the form

$$f(\boldsymbol{t}) = C + \sum_{\alpha} f_\alpha(t_\alpha) + \sum_{\alpha<\beta} f_{\alpha\beta}(t_\alpha, t_\beta) + \cdots,$$

where $C$, $f_\alpha$, $f_{\alpha\beta}$, etc, are the mean, main effects, two-factor interaction effects, etc. These effects can be determined in the following way: Let $d\mu_\alpha$ be a probability measure on $\mathscr{T}^{(\alpha)}$ and $d\mu(\boldsymbol{t}) = \prod_{\alpha=1}^{d} d\mu_\alpha(t_\alpha)$ be a product measure on $\mathscr{T}$. Let $u, v \subseteq D = \{1, \ldots, d\}$ denote subsets of the axes. We use $d\mu_u$ for integration with respect to the axes in $u$, leaving a function defined over the axes in $D - u$. That is, $d\mu_u = \prod_{\alpha \in u} d\mu_\alpha(t_\alpha)$. Similarly, $d\mu_{-u}$ indicates integration with respect to the complement axes $D - u$. The general form of an effect is

$$f_u = \int \left\{ f - \sum_{v \subset u} f_v \right\} d\mu_{-u}.$$

Thus the mean is $C = \int f(\boldsymbol{t})\, d\mu(\boldsymbol{t})$. The main effects are

$$f_\alpha(t_\alpha) = \int (f - C)\, d\mu_{-\alpha}, \quad 1 \le \alpha \le d.$$

The two-factor interaction effects are

$$f_{\alpha\beta}(t_\alpha, t_\beta) = \int (f - C - f_\alpha - f_\beta)\, d\mu_{-\alpha\beta}, \quad \alpha \ne \beta,$$

and so forth.

The estimate, $f_{n,\lambda}$, of $f$ is obtained by finding $f_{n,\lambda}$ in an appropriate subspace $\mathscr{M}$ of $\mathscr{H}$ to minimize an expression similar to

$$\frac{1}{n}\sum_{i=1}^{n}[y_i - f(\boldsymbol{t}_i)]^2 + \lambda \left[ \sum_{\alpha \in I_{\mathscr{M}}} \theta_\alpha^{-1} J_\alpha(f_\alpha) + \sum_{\alpha,\beta \in I_{\mathscr{M}}} \theta_{\alpha\beta}^{-1} J_{\alpha\beta}(f_{\alpha\beta}) + \cdots \right],$$

where $I_{\mathscr{M}}$ is the collection of indices for components to be included in the model, the $J_\alpha$, $J_{\alpha\beta}$ and so forth are roughness penalties, and the series may be truncated at some point. The scalar $\lambda$ is the main smoothing parameter, and the $\theta$'s are subsidiary smoothing parameters satisfying an appropriate constraint for identifiability. The details for fitting these models from given data sets can be found in Wahba (1990), Gu and Wahba (1993), Wahba et al. (1995) and references cited therein.

The SS-ANOVA methods provide the ability to visualize some of the relationships between the variables not easily observed with use of the standard parametric methods. They also overcome the "curse of dimensionality" since estimating a more general function $f(t_1, \ldots, t_d)$ will require truly large data sets for even moderate $d$. In the last decade, the SS-ANOVA methods have become popular in the analysis of real data. Recent applications are provided by Wahba and her collaborators, such as modelling environmental data (Gu and Wahba (1993)), meteorological data (Wahba and Luo (1997), and Luo and Wahba (1998)), and epidemiological data (Wahba et al. (1995), and Wang et al. (1997)).

It is known that the SS-ANOVA estimate of $f$ is determined by observations of $f$ at $n$ points in $\mathcal{T}$. As for the traditional parametric models, one may then ask how to collect the $n$ observations so that the SS-ANOVA estimate of $f$ is closest to $f$ in some appropriate sense among all designs $\xi_n$. This problem arises in many applications. The region $\mathcal{T}$ may be a surface, a sphere, or a rectangle, and $f$ the temperature, the concentration of some air pollutant, or the outcome of medical experiments. The design problem concerns optimal or nearly optimal choices of $\xi_n$.

In this paper, according to the correspondence between smoothing splines and random effects models with a partially improper prior (Gu and Wahba (1993)), we define the integrated mean squared error (IMSE) of the SS-ANOVA estimate. The definition of IMSE is based on the idea of model-robust design due to Box and Draper (1959). We choose the design so that IMSE is as small as possible. At this point, our consideration is somewhat similar to that of Steinberg (1985), Sacks et al. (1989) and Mitchell et al. (1994). In the asymptotic case where $n\lambda$ is assumed to be large enough, we derive an asymptotic approximation of IMSE which leads to a criterion for choosing nearly optimal designs.

In Section 2 we first briefly review the SS-ANOVA setting and its relationship to a random effects model with a partially improper prior. We then define the IMSE and derive its asymptotic approximation. In Section 3 we provide three examples including an additive spline model with only constant fixed effect, an additive spline model with constant and linear fixed effects, and an interaction spline model with constant and linear fixed effects. Several designs are compared numerically using IMSE and its approximation. Finally, a summary is given in Section 4.

## 2. The underlying model and design criteria

### *2.1 The SS-ANOVA models*

Let $\mathcal{H}^{(\alpha)}$ be a reproducing kernel Hilbert space of real valued functions on $\mathcal{T}^{(\alpha)}$ with $\int_{\mathcal{T}^{(\alpha)}} f_\alpha(t_\alpha)\, d\mu_\alpha = 0$ for $f_\alpha \in \mathcal{H}^{(\alpha)}$, and let $[1^{(\alpha)}]$ be the one dimensional space of constant functions on $\mathcal{T}^{(\alpha)}$. Any $f$ in the space $[1^{(\alpha)}] \oplus \mathcal{H}^{(\alpha)}$ has a unique decomposition $f = P_c f + (f - P_c f)$ with $P_c f = \int f\, d\mu_\alpha \in [1^{(\alpha)}]$ and $(f - P_c f) \in \mathcal{H}^{(\alpha)}$. We endow the space $[1^{(\alpha)}] \oplus \mathcal{H}^{(\alpha)}$ with the square norm $\|f\|^2 = (P_c f)^2 + \|f - P_c f\|^2_{\mathcal{H}^{(\alpha)}}$. Consider the tensor product of the $d$ Hilbert spaces $\mathcal{H} = \bigotimes_{\alpha=1}^{d} [[1^{(\alpha)}] \oplus \mathcal{H}^{(\alpha)}]$ which can be expanded as

$$\mathcal{H} = [1] \oplus \sum_{\alpha} \mathcal{H}^{(\alpha)} \oplus \sum_{\alpha<\beta} (\mathcal{H}^{(\alpha)} \oplus \mathcal{H}^{(\beta)}) \oplus \cdots.$$

Further, $\mathscr{H}^{(\alpha)}$ is decomposed into a parametric part and a smooth part by letting $\mathscr{H}^{(\alpha)} = \mathscr{H}_\pi^{(\alpha)} \oplus \mathscr{H}_s^{(\alpha)}$, where $\mathscr{H}_\pi^{(\alpha)}$ (the parametric part) is finite dimensional, and $\mathscr{H}_s^{(\alpha)}$ (the smooth part) is the orthocomplement of $\mathscr{H}_\pi^{(\alpha)}$ in $\mathscr{H}^{(\alpha)}$. Then $\mathscr{H}$ has been decomposed into sums of products of unpenalized finite dimensional subspaces, plus main effects smooth subspaces, plus two factor interaction spaces of the form $[\mathscr{H}_\pi^{(\alpha)} \otimes \mathscr{H}_s^{(\beta)}]$, $[\mathscr{H}_s^{(\alpha)} \otimes \mathscr{H}_\pi^{(\beta)}]$ and $[\mathscr{H}_s^{(\alpha)} \otimes \mathscr{H}_s^{(\beta)}]$, and so on for the three and higher factor subspaces.

Now suppose that we have decided which subspaces will be included into the model $\mathscr{M}$. Let $\mathscr{H}_0$ of dimension $q$ be the collection of all included unpenalized subspaces, and relabel the other included subspaces as $\mathscr{H}^\beta$, $\beta = 1, \ldots, p$. Let $\mathscr{H}_1 = \sum_\beta \mathscr{H}^\beta$. Then the SS-ANOVA estimate $f_{n,\lambda}$ of $f$ is in $\mathscr{M} = \mathscr{H}_0 \oplus \mathscr{H}_1$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}[y_i - f(\boldsymbol{t}_i)]^2 + \lambda \sum_{\beta=1}^{p} \theta_\beta^{-1} \|P^\beta f\|^2, \tag{1}$$

where $P^\beta$ is the orthogonal projector in $\mathscr{M}$ onto $\mathscr{H}^\beta$. $\lambda$ is the main smoothing parameter, and the $\theta_\beta$ are subsidiary smoothing parameters with $\theta_\beta > 0$. In a practical application, $\lambda$ and the $\theta_\beta$ may be chosen by generalized cross-validation ($\sigma^2$ unknown) or unbiased risk ($\sigma^2$ known); see Wahba (1990) and the references cited therein. Here, for the purpose of choosing design we assume that the subsidiary parameters, $\theta_\beta$, are given.

Let $g_1, \ldots, g_q$ be a basis for $\mathscr{H}_0$, and let $K_\beta(\boldsymbol{s}, \boldsymbol{t})$ be the reproducing kernel for $\mathscr{H}^\beta$, $\beta = 1, \ldots, p$. It is known from Gu and Wahba (1993) that the SS-ANOVA estimate $f_{n,\lambda}$ under the selected model $\mathscr{M}$ can be derived from the following random effects model with a partially improper prior:

$$\begin{aligned} F_{\text{ass}}(\boldsymbol{t}) &= \sum_{v=1}^{q} \tau_v g_v(\boldsymbol{t}) + b^{1/2} \sum_{\beta=1}^{p} \sqrt{\theta_\beta} Z_{1,\beta}(\boldsymbol{t}) \\ Y_i &= F_{\text{ass}}(\boldsymbol{t}_i) + \varepsilon_i, \quad i = 1, \ldots, n, \end{aligned} \tag{2}$$

where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)' \sim N(\boldsymbol{0}, a\boldsymbol{I})$ with $a \to \infty$, and the $Z_{1,\beta}$ are independent zero mean Gaussian stochastic processes independent of the $\tau_v$, with

$$E[Z_{1,\beta}(\boldsymbol{s}) Z_{1,\beta}(\boldsymbol{t})] = K_\beta(\boldsymbol{s}, \boldsymbol{t}).$$

It follows that $Z_1(\boldsymbol{t}) = \sum_\beta \sqrt{\theta_\beta} Z_{1,\beta}(\boldsymbol{t})$ satisfies $E[Z_1(\boldsymbol{s}) Z_1(\boldsymbol{t})] = K(\boldsymbol{s}, \boldsymbol{t})$, where

$$K(\boldsymbol{s}, \boldsymbol{t}) = \sum_{\beta=1}^{p} \theta_\beta K_\beta(\boldsymbol{s}, \boldsymbol{t}).$$

We call the model (2) the *assumed model* which we want to fit. Letting

$$\hat{F}_a(\boldsymbol{t}) = E[F_{\text{ass}}(\boldsymbol{t}) \mid Y_i = y_i, i = 1, \ldots, n],$$

then for each fixed $\boldsymbol{t} \in \mathscr{T}$

$$f_{n,\lambda}(\boldsymbol{t}) = \lim_{a \to \infty} \hat{F}_a(\boldsymbol{t}).$$

Now, let $\boldsymbol{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{g}(\boldsymbol{t}) = (g_1(\boldsymbol{t}), \ldots, g_q(\boldsymbol{t}))'$, $\boldsymbol{k}(\boldsymbol{t}) = (K(\boldsymbol{t}, \boldsymbol{t}_1), \ldots, K(\boldsymbol{t}, \boldsymbol{t}_n))'$. Let $\boldsymbol{X}$ be the $n \times q$ matrix whose $(i, v)$th entry is $g_v(\boldsymbol{t}_i)$, and let $\boldsymbol{K}$ be the $n \times n$ matrix whose $(i, j)$th entry is $K(\boldsymbol{t}_i, \boldsymbol{t}_j)$. It is always being assumed that $\boldsymbol{X}$ is of full column rank. Let $\boldsymbol{\Sigma} = \boldsymbol{K} + n\lambda \boldsymbol{I}$, $\eta = a/b$, and

$$\boldsymbol{h}(\boldsymbol{t}) = \eta \boldsymbol{X}\boldsymbol{g}(\boldsymbol{t}) + \boldsymbol{k}(\boldsymbol{t}), \quad \boldsymbol{A} = \eta \boldsymbol{X}\boldsymbol{X}' + \boldsymbol{\Sigma}. \tag{3}$$

Then $\hat{F}_a$ is given by

$$\hat{F}_a(\boldsymbol{t}) = \boldsymbol{h}(\boldsymbol{t})' \boldsymbol{A}^{-1} \boldsymbol{y}. \tag{4}$$

### *2.2 The integrated mean squared error*

Note that the assumed model (2) corresponding to the SS-ANOVA setting $\mathscr{M} = \mathscr{H}_0 \oplus \mathscr{H}_1$ is likely to be, at best, only a reasonable approximation to the true model for the response. In th spirit of *model-robust design* due to Box and Draper (1959), we require a design to provide protection against departures from the fitted model.

Let $\mathscr{H}_2$ be the orthocomplement of $\mathscr{M}$ in $\mathscr{H}$, and let $R(\boldsymbol{s}, \boldsymbol{t})$ be the reproducing kernel for $\mathscr{H}_2$. $R(\boldsymbol{s}, \boldsymbol{t})$ can be taken as follows:

$$R(\boldsymbol{s}, \boldsymbol{t}) = \prod_{\alpha=1}^{d} [1 + K^{(\alpha)}(s_\alpha, t_\alpha)] - \sum_{v=1}^{q} g_v(\boldsymbol{s}) g_v(\boldsymbol{t}) - \sum_{\beta=1}^{p} K_\beta(\boldsymbol{s}, \boldsymbol{t}). \tag{5}$$

Let $Z_2$ be the difference between the true unknown model and the assumed model. Then the *true model* can be expressed by

$$\begin{aligned} F(\boldsymbol{t}) &= \sum_{v=1}^{q} \tau_v g_v(\boldsymbol{t}) + b^{1/2} \sum_{\beta=1}^{p} \sqrt{\theta_\beta} Z_{1,\beta}(\boldsymbol{t}) + Z_2(\boldsymbol{t}) \\ Y_i &= F(\boldsymbol{t}_i) + \varepsilon_i, \quad i = 1, \ldots, n, \end{aligned} \tag{6}$$

where the $\tau_v$ and the $Z_{1,\beta}$ are defined as in (2), and $Z_2$ is zero mean Gaussian process independent of the $\tau_v$ and $Z_{1,\beta}$, with

$$E[Z_2(\boldsymbol{s}) Z_2(\boldsymbol{t})] = bR(\boldsymbol{s}, \boldsymbol{t}).$$

We refer to $Z_2$ in (6) as a *bias* or *contamination* term consistent with the terminology used in the design literature and introduced by Box and Draper (1959).

We now define the integrated mean squared error in the estimation as follows:

$$\Delta(\xi_n) \equiv \mathrm{IMSE}(\xi_n) = \lim_{a \to \infty} \frac{1}{b} E \int_{\mathscr{T}} [F(\boldsymbol{t}) - \hat{F}_a(\boldsymbol{t})]^2 \, d\mu(\boldsymbol{t}), \tag{7}$$

where $F(\boldsymbol{t})$ is given in (6), and $\hat{F}_a(\boldsymbol{t})$ is given by (4). That is, $\Delta(\xi_n)$ is the squared expected loss incurred if the design $\xi_n = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ and the estimator $\hat{F}_a(\boldsymbol{t})$ are used and $F(\boldsymbol{t})$ is the true response.

**Lemma 1.** *Let*

$$\boldsymbol{y}_{\text{ass}} = (F_{\text{ass}}(\boldsymbol{t}_1) + \varepsilon_1, \ldots, F_{\text{ass}}(\boldsymbol{t}_n) + \varepsilon_n)', \quad \hat{Z}_2(\boldsymbol{t}) = \boldsymbol{h}(\boldsymbol{t})'\boldsymbol{A}^{-1}\boldsymbol{z}_2,$$

*where* $\boldsymbol{z}_2 = (Z_2(\boldsymbol{t}_1), \ldots, Z_2(\boldsymbol{t}_n))'$. *Then* $\Delta(\xi_n)$ *can be expressed as*

$$\Delta(\xi_n) = \lim_{a\to\infty} \frac{1}{b} E \int_{\mathscr{T}} \{\text{var}[F_{\text{ass}}(\boldsymbol{t})|\boldsymbol{y}_{\text{ass}}] + E[Z_2(\boldsymbol{t}) - \hat{Z}_2(\boldsymbol{t})]^2\}\, d\mu(\boldsymbol{t}), \tag{8}$$

*where* $\text{var}[F_{\text{ass}}(\boldsymbol{t})|\boldsymbol{y}_{\text{ass}}]$ *is the posterior variance under the assumed model (2).*

*Proof.* Note that $\boldsymbol{y} = \boldsymbol{y}_{\text{ass}} + \boldsymbol{z}_2$ under the true model (6). Then $\hat{F}_a(\boldsymbol{t})$ given in (4) can be expressed as

$$\hat{F}_a(\boldsymbol{t}) = E[F_{\text{ass}}(\boldsymbol{t})|\boldsymbol{y}_{\text{ass}}] + \hat{Z}_2(\boldsymbol{t}),$$

where $E[F_{\text{ass}}(\boldsymbol{t})|\boldsymbol{y}_{\text{ass}}]$ is the posterior expectation under the assumed model (2). From the definition of $f$ in the true model (6) and the independence of $\boldsymbol{y}_{\text{ass}}$ and $\hat{Z}_2$, we have

$$E[F(\boldsymbol{t}) - \hat{F}_a(\boldsymbol{t})]^2 = \text{var}[F_{\text{ass}}(\boldsymbol{t})|\boldsymbol{y}_{\text{ass}}] + E[Z_2(\boldsymbol{t}) - \hat{Z}_2(\boldsymbol{t})]^2.$$

This immediately completes the proof of the lemma. □

Observe that the integrated squared mean error can be split up into two parts: one is due to the posterior variance without bias, and another due to bias. An average posterior variance was used in Steinberg (1985) as a criterion for finding model-robust designs. Our goal will be to minimize this integrated squared mean error.

This mean error can be expressed in a closed form. Let $\boldsymbol{r}(\boldsymbol{t}) = (R(\boldsymbol{t}, \boldsymbol{t}_1), \ldots, R(\boldsymbol{t}, \boldsymbol{t}_n))'$, and let $\boldsymbol{R}$ be the $n \times n$ matrix whose $(i, j)$th entry is $R(\boldsymbol{t}_i, \boldsymbol{t}_j)$. Define matrices $\boldsymbol{\Gamma}_{\boldsymbol{gg}}$, $\boldsymbol{\Gamma}_{\boldsymbol{kg}}$, $\boldsymbol{\Gamma}_{\boldsymbol{rg}}$, $\boldsymbol{\Gamma}_{\boldsymbol{kk}}$ and $\boldsymbol{\Gamma}_{\boldsymbol{rk}}$ by

$$\boldsymbol{\Gamma}_{\boldsymbol{uv}} = \int_{\mathscr{T}} \boldsymbol{u}(\boldsymbol{t})\boldsymbol{v}(\boldsymbol{t})'\, d\mu(\boldsymbol{t}).$$

**Theorem 1.** *Under the model (6) with a partially improper prior assigned to* $\boldsymbol{\tau}$, *the integrated squared mean error* $\Delta(\xi_n)$ *can be expressed as*

$$\Delta(\xi_n) = \int_{\mathscr{T}} [K(\boldsymbol{t}, \boldsymbol{t}) + R(\boldsymbol{t}, \boldsymbol{t})]\, d\mu(\boldsymbol{t}) + \text{tr}\{\boldsymbol{Q}(\boldsymbol{RQ} - \boldsymbol{I})\boldsymbol{\Gamma}_{\boldsymbol{kk}} - 2\boldsymbol{Q}\boldsymbol{\Gamma}_{\boldsymbol{rk}}\}$$
$$+ \text{tr}\{[(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1} + \boldsymbol{LRL}']\boldsymbol{\Gamma}_{\boldsymbol{gg}} + 2\boldsymbol{L}(\boldsymbol{RQ} - \boldsymbol{I})\boldsymbol{\Gamma}_{\boldsymbol{kg}} - 2\boldsymbol{L}\boldsymbol{\Gamma}_{\boldsymbol{rg}}\}, \tag{9}$$

*where*

$$\boldsymbol{L} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}, \quad \boldsymbol{Q} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}.$$

*Proof.* From the assumptions on the assumed model (2) and on the true model (6), we have

$$\frac{1}{b}\operatorname{var}[F_{\text{ass}}(\boldsymbol{t})|\boldsymbol{y}_{\text{ass}}] = K(\boldsymbol{t},\boldsymbol{t}) + \eta\boldsymbol{g}(\boldsymbol{t})'\boldsymbol{g}(\boldsymbol{t}) - \boldsymbol{h}(\boldsymbol{t})'\boldsymbol{A}^{-1}\boldsymbol{h}(\boldsymbol{t})$$
$$\frac{1}{b}E[Z_2(\boldsymbol{t}) - \hat{Z}_2(\boldsymbol{t})]^2 = R(\boldsymbol{t},\boldsymbol{t}) - 2\boldsymbol{h}(\boldsymbol{t})'\boldsymbol{A}^{-1}\boldsymbol{r}(\boldsymbol{t}) + \boldsymbol{h}(\boldsymbol{t})'\boldsymbol{A}^{-1}\boldsymbol{R}\boldsymbol{A}^{-1}\boldsymbol{h}(\boldsymbol{t}), \tag{10}$$

where $\boldsymbol{h}$ and $\boldsymbol{A}$ are defined as in (3). Upon collecting terms in the right sides of (10) and using the following formulas (Gu and Wahba (1993))

$$\lim_{\eta\to\infty}[\eta\boldsymbol{I} - \eta\boldsymbol{X}'\boldsymbol{A}^{-1}\boldsymbol{X}\eta] = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}$$

$$\lim_{\eta\to\infty}\eta\boldsymbol{X}'\boldsymbol{A}^{-1} = \boldsymbol{L}$$

$$\lim_{\eta\to\infty}\boldsymbol{A}^{-1} = \boldsymbol{Q}$$

gives the result (9) from Lemma 1. □

A good design should make the quantity $\Delta(\xi_n)$ as small as possible. We call a design optimal for the SS-ANOVA model if it minimizes the functional $\Delta(\xi_n)$. Of course, the minimizer of $\Delta(\xi_n)$ depends on the values of the smoothing parameters $\lambda$ and $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_p)$. These parameters reflect the prior belief of the experimenter as to the nature of the response function. In order to assess the influence of these parameters on the designs, we define the efficiency of a design, $\xi_n^*$, say, by the ratio

$$\text{Eff}(\xi_n^*) = \frac{\min_{\xi_n}\Delta(\xi_n)}{\Delta(\xi_n^*)}. \tag{11}$$

$\xi_n^*$ may be the design that minimizes $\Delta(\xi_n)$ for a set of preassigned parameters, or some other kind of designs.

### *2.3 Asymptotics*

In this section, we allow $n\lambda$ to be large, and give asymptotic criteria for the design problem.

**Theorem 2.** *Letting $\delta = 1/(n\lambda)$ and $\boldsymbol{M} = \boldsymbol{X}'\boldsymbol{X}$, it follows that*

$$\delta\Delta(\xi_n) = \delta\int_{\mathscr{T}}[K(\boldsymbol{t},\boldsymbol{t}) + R(\boldsymbol{t},\boldsymbol{t})]\,d\mu(\boldsymbol{t}) + \operatorname{tr}\{\boldsymbol{M}^{-1}\boldsymbol{\Gamma}_{gg}\}$$
$$+\,\delta\operatorname{tr}\{\boldsymbol{M}^{-1}\boldsymbol{X}'(\boldsymbol{K}+\boldsymbol{R})\boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{\Gamma}_{gg} - 2\boldsymbol{M}^{-1}\boldsymbol{X}'(\boldsymbol{\Gamma}_{kg}+\boldsymbol{\Gamma}_{rg})\} + O(\delta^2).$$

*Proof.* Formally,

$$\boldsymbol{\Sigma}^{-1} = (\boldsymbol{K}+n\lambda\boldsymbol{I})^{-1} = \delta(\boldsymbol{I}+\delta\boldsymbol{K})^{-1} = \delta[\boldsymbol{I} - \delta\boldsymbol{K} + O(\delta^2)].$$

We then have

$$(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1} = \delta^{-1}[\boldsymbol{M}^{-1} + \delta\boldsymbol{M}^{-1}\boldsymbol{X}'\boldsymbol{K}\boldsymbol{X}\boldsymbol{M}^{-1} + O(\delta^2)],$$

$$\boldsymbol{L} = \boldsymbol{M}^{-1}\boldsymbol{X}' + \delta(\boldsymbol{M}^{-1}\boldsymbol{X}'\boldsymbol{K}\boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{X}' - \boldsymbol{M}^{-1}\boldsymbol{X}'\boldsymbol{K}) + O(\delta^2),$$

$$\boldsymbol{Q} = \delta(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{X}') + O(\delta^2),$$

Substituting these expressions into (9) yields the result in the theorem. □

Thus, ignoring terms of order $O(\delta^2)$, minimizing $\Delta(\xi_n)$ is then equivalent to minimizing the following functional

$$\Delta_{\text{asy}}(\xi_n) = \operatorname{tr}(\boldsymbol{M}^{-1}\boldsymbol{\Gamma}_{gg}) + \delta \operatorname{tr}\{\boldsymbol{M}^{-1}\boldsymbol{X}'(\boldsymbol{K}+\boldsymbol{R})\boldsymbol{X}\boldsymbol{M}^{-1}\boldsymbol{\Gamma}_{gg} - 2\boldsymbol{M}^{-1}\boldsymbol{X}'(\boldsymbol{\Gamma}_{kg} + \boldsymbol{\Gamma}_{rg})\} \tag{12}$$

We call the design minimizing $\Delta_{\text{asy}}(\xi_n)$ a nearly optimal design.

In the particular case where the parametric part $\mathscr{H}_0$ in the SS-ANOVA model is only of constant functions, we may take $\boldsymbol{g}(\boldsymbol{t}) \equiv 1$. Then $\boldsymbol{X} = \boldsymbol{1}$, $\boldsymbol{\Gamma}_{kg} = \boldsymbol{\Gamma}_{rg} = \boldsymbol{0}$, and $\Delta_{\text{asy}}(\xi_n)$ becomes that

$$\Delta_{\text{asy}}(\xi_n) = \frac{1}{n} + \frac{\delta}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}[K(\boldsymbol{t}_i,\boldsymbol{t}_j) + R(\boldsymbol{t}_i,\boldsymbol{t}_j)]. \tag{13}$$

Recalling the approximate design theory due to Kiefer, we replace a design $\xi_n$ by a probability measure $\xi$ on $\mathscr{T}$. Then the problem corresponding to (13) becomes minimizing

$$D(\xi) = \int_{\mathscr{T}\times\mathscr{T}} [K(\boldsymbol{s},\boldsymbol{t}) + R(\boldsymbol{s},\boldsymbol{t})]\, d\xi(\boldsymbol{s})\, d\xi(\boldsymbol{t})$$

Note that the positive definiteness of $K(\boldsymbol{s},\boldsymbol{t}) + R(\boldsymbol{s},\boldsymbol{t})$ makes $D(\xi) \geq 0$. It follows that the minimizer of $D(\xi)$ is $\xi = \mu$, since $D(\mu) = 0$. Especially, if $\mu$ is Lebesgue measure on $\mathscr{T}$ then the continuous uniform design on $\mathscr{T}$ is a nearly optimal design for the SS-ANOVA model. In this special case, a design of $n$ points $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$ should be uniformly scattered on the experimental region. Such a design is called a *uniform design* by Fang and Wang (1994). For details of the construction of uniform design see the monograph by Fang and Wang (1994).

## 3. Examples for illustration

We consider the polynomial spline for illustration. Let $\mathscr{T}^{(\alpha)} = [0,1]$ with Lebesgue measure, and let $\mathscr{W}_m$ be the Sobolev-Hilbert space

$$\mathscr{W}_m = \{f : f, f', \ldots, f^{(m-1)} \text{ absolutely continuous, } f^{(m)} \in \mathscr{L}^2[0,1]\}$$

equipped with a square norm

$$\|f\|^2 = \sum_{v=0}^{m-1}\left(\int_0^1 f^{(v)}(x)\,dx\right)^2 + \int_0^1 (f^{(m)}(x))^2\,dx,$$

which was used in Wahba (1990, Chapter 10). Let $B_\ell$ be the $\ell$th Bernoulli polynomial and let $[B_\ell]$ denote the one-dimensional space spanned by $B_\ell$. The first few of Bernoulli polynomials which will be used here are as follows:

$$B_0(t) = 1, \quad B_1(t) = t - 1/2, \quad B_2(t) = t^2 - t + 1/6,$$

$$B_3(t) = t^3 - 3t^2/2 + t/2.$$

Then $\mathscr{W}_m$ can be decomposed as the direct sum of $m$ orthogonal subspaces $[B_\ell]$, $\ell = 0, \ldots, m-1$, and $\mathscr{H}_*$ which is the subspace (orthogonal to $\bigoplus_\ell [B_\ell]$) satisfying $\int f^{(v)}\,dt = 0$, $v = 0, \ldots, m-1$. That is

$$\mathscr{W}_m = [B_0] \oplus [B_1] \oplus \cdots \oplus [B_{m-1}] \oplus \mathscr{H}_*.$$

Let $\mathscr{H}$ be the tensor product of $\mathscr{W}_m$ with itself $d$ times. Then $\mathscr{H}$ may be decomposed into the direct sum of $(m+1)^d$ fundamental subspaces, each of the form

$$[\cdot] \otimes \cdots \otimes [\cdot] \quad (d \text{ boxes}), \tag{14}$$

where each box [ ] is filled with either $[B_\ell]$ for some $\ell$ or $\mathscr{H}_*$. In the following examples we use the case with $m = 1$ and let $\mathscr{H}^{(\alpha)} = \mathscr{H}_*$ for $\alpha = 1, \ldots, d$. Then the reproducing kernel for $\mathscr{H}^{(\alpha)}$ is given by

$$K_*(s_\alpha, t_\alpha) = B_1(s_\alpha)B_1(t_\alpha) + \frac{1}{2}B_2(\{s_\alpha - t_\alpha\}), \tag{15}$$

where $\{t\}$ is the fractional part of a real number $t$, i.e., $\{t\} = t - \lfloor t \rfloor$, where $\lfloor t \rfloor$ is the greatest integer less than or equal to $t$.

**Example 1.** *Additive spline model with only constant fixed effect*: Let the parametric part $\mathscr{H}_0$ be the one-dimensional space of constant functions on $\mathscr{T} = [0,1]^d$, i.e., $\boldsymbol{g}(\boldsymbol{t}) = 1$. Let the penalty part $\mathscr{H}_1$ be given by $\mathscr{H}_1 = \bigoplus_{\alpha=1}^d \mathscr{H}^\alpha$, where $\mathscr{H}^\alpha$ is of the form (14) with $\mathscr{H}_*$ in the $\alpha$th box and $[B_0]$ in the other boxes. Therefore, we have

$$K(\boldsymbol{s}, \boldsymbol{t}) = \sum_{\alpha=1}^d \theta_\alpha K_*(s_\alpha, t_\alpha),$$

where $K_*$ is given in (15). The kernel $R$ defined in (5) becomes

$$R(\boldsymbol{s}, \boldsymbol{t}) = \prod_{\alpha=1}^d [1 + K_*(s_\alpha, t_\alpha)] - 1 - \sum_{\alpha=1}^d K_*(s_\alpha, t_\alpha). \tag{16}$$

For any given $n$ points $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n \in [0,1]^d$ we have $\boldsymbol{X} = \mathbf{1}_n$, $\boldsymbol{\Gamma}_{kg} = \boldsymbol{\Gamma}_{rg} = \mathbf{0}$, $\boldsymbol{\Gamma}_{rk} = \mathbf{0}$. The $(i,j)$th entry of $\boldsymbol{\Gamma}_{kk}$ is given by

$$\gamma_{ij} = \sum_{\alpha=1}^{d} \theta_\alpha^2 \psi_{ij\alpha}^{(1)},$$

where

$$\psi_{ij\alpha}^{(1)} = \frac{1}{45} - \frac{u_{ij\alpha}^2 + v_{ij\alpha}^2}{6} - \frac{u_{ij\alpha}^4 - v_{ij\alpha}^4}{24} - \frac{u_{ij\alpha}^2 v_{ij\alpha}^2}{4} + \frac{u_{ij\alpha}^2 v_{ij\alpha}}{2} + \frac{v_{ij\alpha}^3}{6}, \tag{17}$$

where $u_{ij\alpha} = \min\{t_{i\alpha}, t_{j\alpha}\}$ and $v_{ij\alpha} = \max\{t_{i\alpha}, t_{j\alpha}\}$. For given $n$ and $\lambda$, $\boldsymbol{\theta}$, we then can find numerically the design that minimizes $\Delta(\xi_n)$ in (9) for the model here. Shown in Figure 1 are the designs of $n = 12$ and 24 points in 2-dimensional unit cube $[0,1]^2$ for different values of $\lambda$, where we take $\boldsymbol{\theta} = (1,1)$. We also find 2-dimensional designs for some other values of $\boldsymbol{\theta}$. It turns out that the influence of the prior parameters, $\lambda$ and $\boldsymbol{\theta}$, on the location of design points is slight. All design points for the additive model with only a constant parametric part should be uniformly scattered on the design region. For the $d > 2$ dimensional case, the designs that minimize $\Delta(\xi_n)$ in (9) can be also found numerically.

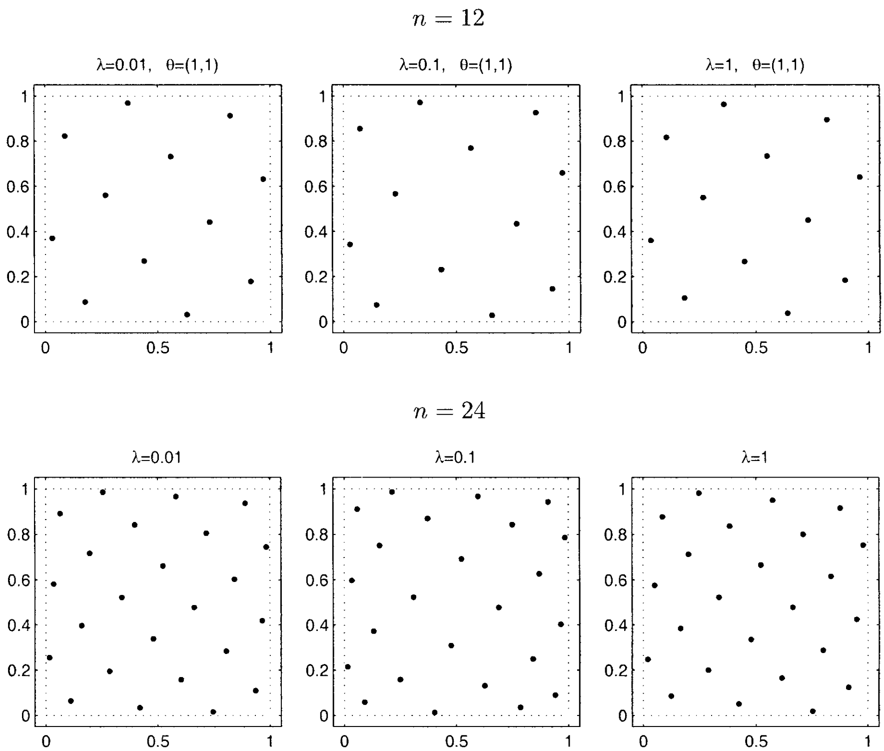


**Fig. 1.** The designs that minimize $\Delta(\xi_n)$ for the additive spline model with only constant fixed effect in Example 1, with $\theta_1 = \theta_2 = 1$.

Here, we compare the following two kinds of designs for the additive model:

$\xi_n^{\mathrm{asy}}$ – To minimize $\Delta_{\mathrm{asy}}(\xi_n)$ in (13) with $\lambda = 1$ and $\theta_1 = \cdots = \theta_d = 1$;

$\xi_n^{\mathrm{glp}}$ – A good lattice point (glp) set of $n$ points in $[0, 1]^d$.

The definition of a glp set is as follows (Fang and Wang (1994)): Let $(n; h_1, \ldots, h_d)$ be a vector with integral components satisfying $1 \leq h_\alpha < n$, $h_\alpha \neq h_\beta$ $(\alpha \neq \beta)$, $d < n$ and the greatest common divisor $(n, h_\alpha) = 1$, $\alpha = 1, \ldots, d$. Let

$$t_{i\alpha} = \left\{ \frac{2ih_\alpha - 1}{2n} \right\}, \quad i = 1, \ldots, n, \ \alpha = 1, \ldots, d.$$

Then the set $\{\boldsymbol{t}_i = (t_{i1}, \ldots, t_{id}), i = 1, \ldots, n\}$ is called the lattice point set of the generating vector $(n; h_1, \ldots, h_d)$. If this set has the smallest discrepancy among all possible generating vectors, then it is called a glp set. It is known that the points of a glp set are uniformly scattered in $[0, 1]^d$ and they are widely used in numerical problems. A lot of glp sets are provided in Hua and Wang (1981), and Fang and Wang (1994)). The generating vectors of the glp sets used in this section are found by a greedy algorithm described in Hickernell (1996). Specifically, for each $n$ we set $h_1 = 1$ and iteratively search for the $h_2, h_3, \ldots,$ that minimize the quantity below

$$[D_n]^2 = -1 + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \prod_{r=1}^{d} \left[ \frac{4}{3} + \frac{t_{ir}^2 + t_{jr}^2}{2} - \max(t_{ir}, t_{jr}) \right].$$

This $D_n$ is a worst-case quadrature error for a reproducing kernel Hilbert space of integrands (Hickernell (1996)).

Table 1 provides the efficiencies of the glp designs, $\xi_n^{\mathrm{glp}}$, and nearly optimal designs, $\xi_n^{\mathrm{asy}}$, described above, for the additive spline model with constant fixed effect in Example 1. These efficiencies are computed by the definition in (11), where the minimum, $\min_{\xi_n} \Delta(\xi_n)$, is taken over all sets of $n$ points in $[0, 1]^d$ for $\theta_1 = \cdots = \theta_d = 1$ and each $\lambda = 0.01, 0.1, 1$, respectively. We also compute the efficiencies for the values of unequal $\theta$'s, not listed here. All numerical results show that the glp designs and nearly optimal designs have high efficiencies for this additive spline model.

**Table 1.** The efficiencies, defined in (11), of designs $\xi_n^{\mathrm{glp}}$ and $\xi_n^{\mathrm{asy}}$ for the additive spline model with constant fixed effect in Example 1 with $\theta_1 = \cdots = \theta_d = 1$.

| | $\lambda$ | $\xi_n^{\mathrm{glp}}$ | $\xi_n^{\mathrm{asy}}$ |
|---|---|---|---|
| | 0.01 | 0.9812 | 0.9805 |
| $d = 2$, $n = 12$ | 0.1 | 0.9823 | 0.9815 |
| | 1 | 0.9995 | 0.9994 |
| | 0.01 | 0.9817 | 0.9789 |
| $d = 3$, $n = 16$ | 0.1 | 0.9846 | 0.9844 |
| | 1 | 0.9994 | 0.9993 |
| | 0.01 | 0.9811 | 0.9823 |
| $d = 4$, $n = 32$ | 0.1 | 0.9859 | 0.9860 |
| | 1 | 0.9992 | 0.9992 |

**Example 2.** *Additive spline model with constant and linear fixed effects*: Let $\mathscr{H}^{(\alpha)} = \mathscr{H}_*$ be described as in Example 1. Let $\mathscr{H}_{*\pi}$, the parametric part of in $\mathscr{H}_*$, be spanned by $B_1$, and $\mathscr{H}_{*s}$, the smooth part, be the orthocomplement of $\mathscr{H}_{*\pi}$ in $\mathscr{H}_*$. Then the reproducing kernel for $\mathscr{H}_{*s}$ is

$$K_{*s}(s_\alpha, t_\alpha) = \frac{1}{2} B_2(\{s_\alpha - t_\alpha\}). \tag{18}$$

Then the parametric part for the problem is $\mathscr{H}_0 = \operatorname{span}\{1, B_1(t_1), \ldots, B_1(t_d)\}$. We take

$$\boldsymbol{g}(\boldsymbol{t}) = (1, B_1(t_1), \ldots, B_1(t_d))'.$$

The smooth part for the problem is $\mathscr{H}_1 = \bigoplus_{\alpha=1}^{d} \mathscr{H}^\alpha$ where $\mathscr{H}^\alpha$ is of the form (14) with $\mathscr{H}_{*s}$ in the $\alpha$th box and $[B_0]$ in the other boxes. We then have

$$K(\boldsymbol{s}, \boldsymbol{t}) = \sum_{\alpha=1}^{d} \theta_\alpha K_{*s}(s_\alpha, t_\alpha).$$

The kernel $R(\boldsymbol{s}, \boldsymbol{t})$ is the same as in (16). We find that $\boldsymbol{\Gamma}_{\boldsymbol{rg}} = \mathbf{0}$, $\boldsymbol{\Gamma}_{\boldsymbol{rk}} = \mathbf{0}$, and

$$\boldsymbol{\Gamma}_{\boldsymbol{gg}} = \operatorname{diag}(1, 1/12, \ldots, 1/12),$$

$$\boldsymbol{\Gamma}_{\boldsymbol{kg}} = -\frac{1}{6\sqrt{3}} \begin{pmatrix} 0 & \theta_1 B_3(t_{11}) & \cdots & \theta_d B_3(t_{1d}) \\ \vdots & \vdots & & \vdots \\ 0 & \theta_1 B_3(t_{n1}) & \cdots & \theta_d B_3(t_{nd}) \end{pmatrix}.$$

The $(i, j)$th entry of $\boldsymbol{\Gamma}_{\boldsymbol{kk}}$ is given by

$$\gamma_{ij} = \sum_{\alpha=1}^{d} \theta_\alpha^2 \psi_{ij\alpha}^{(2)},$$

where

$$\psi_{ij\alpha}^{(2)} = \psi_{ij\alpha}^{(1)} - \frac{1}{12} B_1(t_{i\alpha}) B_1(t_{j\alpha}) + \frac{1}{6} B_1(t_{i\alpha}) B_3(t_{j\alpha}) + \frac{1}{6} B_3(t_{i\alpha}) B_3(t_{j\alpha}), \tag{19}$$

where $\psi_{ij\alpha}^{(1)}$ is defined as in (17).

We numerically find the optimal designs that minimize $\Delta(\xi_n)$ in (9) and the asymptotic optimal designs that minimizes $\Delta_{\text{asy}}(\xi_n)$ in (12) for a given values of $\lambda$ and $\boldsymbol{\theta}$. Specifically, we take the parameters $\theta_\alpha$ all to be 1. The number $n$ of points of a design in the $d$-dimensional unit cube $[0, 1]^d$ is taken to be a multiple of $2^d$. Shown in Figure 2 are some of the 2-dimensional optimal designs with multiples of $2^2 = 4$ points in $[0, 1]^2$.

Figure 3 shows the efficiencies of both the designs $\xi_n^{\text{glp}}$ and $\xi_n^{\text{asy}}$ for the additive spline model with constant and linear fixed effects described in this example. Here, $\xi_n^{\text{glp}}$ is the glp design defined as in Example 1, and $\xi_n^{\text{asy}}$ is the nearly optimal design that minimizes $\Delta_{\text{asy}}(\xi_n)$ in (12) under this model, where the parameters $\theta_\alpha$ are still taken to be 1.

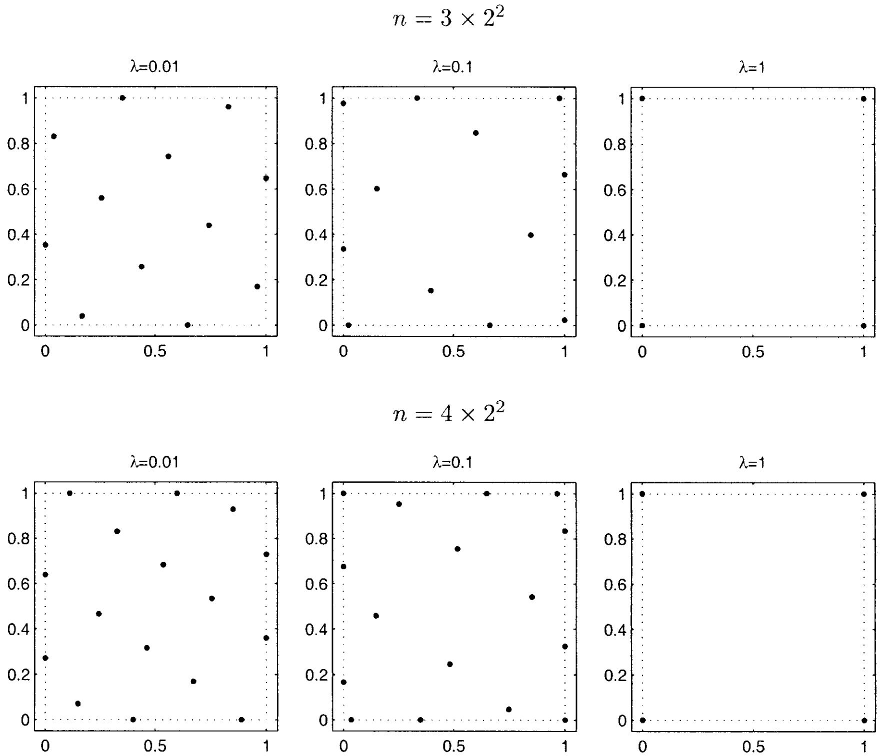


**Fig. 2.** The designs that minimize $\Delta(\xi_n)$ for the additive spline model with constant and linear fixed effects in Example 2, with $\theta_1 = \theta_2 = 1$.

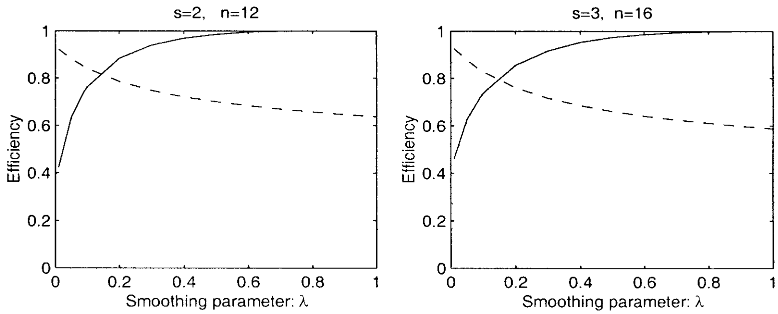


**Fig. 3.** The efficiencies of the glp design $\xi_n^{\mathrm{glp}}$ (dashed line) and the nearly optimal design $\xi_n^{\mathrm{asy}}$ (solid line) for the additive spline model with constant and linear fixed effects in Example 2, with $\theta_1 = \cdots = \theta_d = 1$.

We observe from Figures 2–3 that the locations of the design points in the cube are far more dependent on the parameter $\lambda$. When $\lambda$ is small, the designs have a reasonably even distribution of design points. However, as $\lambda$ increases, the designs tend towards the optimal designs under a classical first-order model, which are $2^d$ factorial designs with $n/2^d$ replicates at each corners of the cube.

Also, the glp design are better than the nearly optimal design for small $\lambda$ close to 0, while as $\lambda$ increases, the nearly optimal design are getting better than the glp design and is of high efficiency. This also indicates that our asymptotic criterion for selecting the design is efficient for the parameter $\lambda$ not too small. On the other hand, the efficiencies of the two designs change a little as the dimension $d$ increases.

**Example 3.** *Interaction spline model with constant and linear fixed effects*: Let $\mathscr{H}^{(\alpha)} = \mathscr{H}_{*\pi} \oplus \mathscr{H}_{*s}$, where $\mathscr{H}_{*\pi}$ and $\mathscr{H}_{*s}$ are described as in Example 2. Let the smooth part for the problem, $\mathscr{H}_1$, be the direct sum of the $d + d(d-1)/2$ fundamental subspaces of the form (14) with $\mathscr{H}_{*s}$ in one box and $[B_0]$ in the other boxes for $d$ main effects subspaces, and with $\mathscr{H}_{*s}$ in two boxes and $[B_0]$ in the other boxes for two factor interaction subspaces. Then the basis for the parametric part $\mathscr{H}_0$ is

$$\boldsymbol{g}(\boldsymbol{t}) = (1, B_1(t_1), \ldots, B_1(t_d))',$$

and the kernel for the smooth part is

$$K(\boldsymbol{s}, \boldsymbol{t}) = \sum_{\alpha=1}^{d} \theta_\alpha K_{*s}(s_\alpha, t_\alpha) + \sum_{1 \le \alpha < \beta \le d} \theta_{\alpha\beta} K_{*s}(s_\alpha, t_\beta) K_{*s}(s_\beta, t_\beta),$$

where $K_{*s}$ is defined by (18). The kernel for the subspace not included in the model is

$$R(\boldsymbol{s}, \boldsymbol{t}) = \prod_{\alpha=1}^{d} [1 + K_*(s_\alpha, t_\alpha)] - 1 - \sum_{\alpha=1}^{d} K_*(s_\alpha, t_\alpha) - \sum_{1 \le \alpha < \beta \le d} K_{*s}(s_\alpha, t_\alpha) K_{*s}(s_\beta, t_\beta),$$

where $K_*$ is defined in (15). We find that $\boldsymbol{\Gamma}_{rg} = \boldsymbol{0}$, $\boldsymbol{\Gamma}_{rk} = \boldsymbol{0}$, and $\boldsymbol{\Gamma}_{gg}$ and $\boldsymbol{\Gamma}_{kg}$ are the same as in Example 2. The $(i, j)$th entry of $\boldsymbol{\Gamma}_{kk}$ here is given by

$$\gamma_{ij} = \sum_{\alpha=1}^{d} \theta_\alpha^2 \psi_{ij\alpha}^{(2)} + \sum_{1 \le \alpha < \beta \le d} \theta_{\alpha\beta}^2 \psi_{ij\alpha}^{(2)} \psi_{ij\beta}^{(2)},$$

where $\psi_{ij\alpha}^{(2)}$ is defined as in (19).

As done in Example 2, we find numerically the optimal designs that minimize $\Delta(\xi_n)$ defined in (9), and the nearly optimal designs that minimize $\Delta_{\text{asy}}(\xi_n)$ defined in (12). The $\theta$'s in the criteria are all taken to be 1 in finding these designs. Figure 4 shows the plots of 2-dimensional optimal designs where $n$ are multiples of $2^2$. The efficiencies of the glp designs and the asymptotic designs are shown in Figure 5. The behaviours of these designs for the interaction spline model with constant and linear fixed effects are similar to those designs for the additive spline model with the same fixed effects in Example 2.

## 4. Summary

The design problem for specific smoothing spline ANOVA models has been considered in this paper based on the relationship between SS-ANOVA models

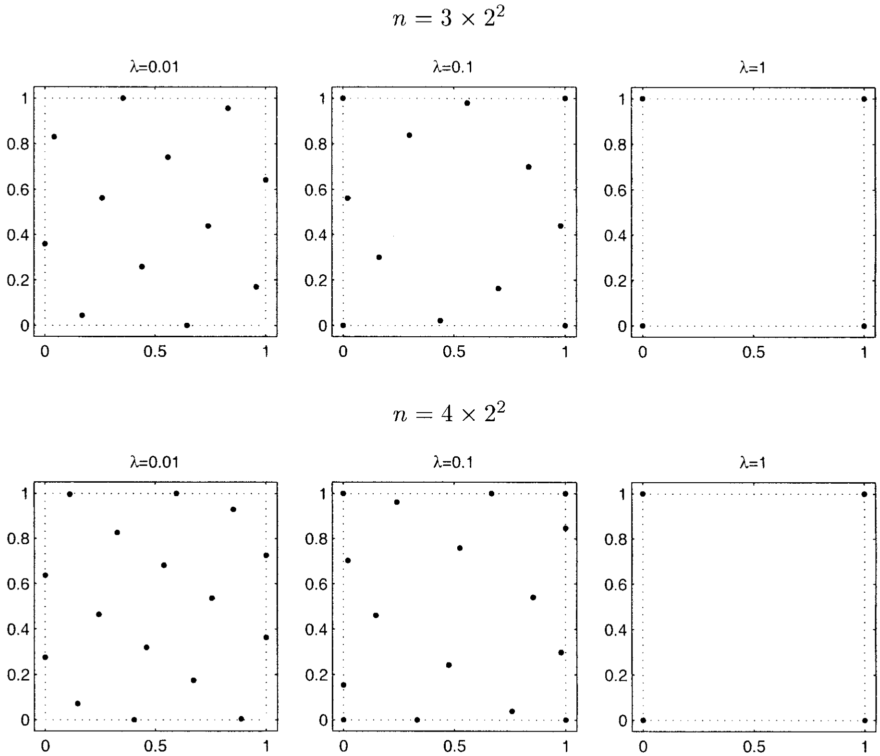


**Fig. 4.** The designs that minimize $\Delta(\xi_n)$ for the interaction spline model with constant and linear fixed effects in Example 3, with $\theta_1 = \theta_2 = \theta_{12} = 1$.

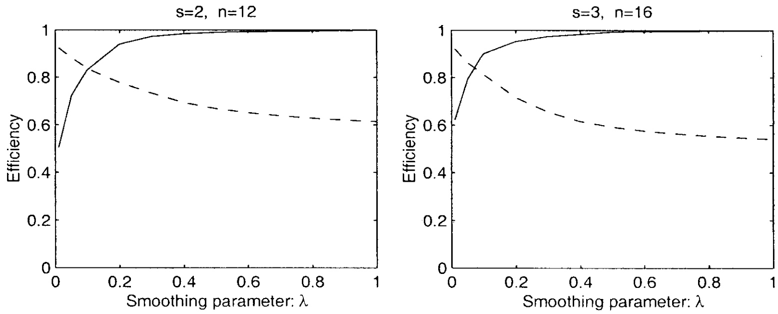


**Fig. 5.** The efficiencies of the glp design $\xi_n^{\text{glp}}$ (dashed line) and the nearly optimal design $\xi_n^{\text{asy}}$ (solid line) for the interaction spline model with constant and linear fixed effects in Example 3, with the $\theta_\alpha = \theta_{\alpha\beta} = 1$.

and random effects models with a partially improper prior. Designs are compared numerically using the integrated mean squared error and an asymptotic approximation of IMSE assuming that $n\lambda$ is large. Three examples are provided for illustration for the designs in the $d$-dimensional unit cube, namely:

i. An additive spline model with only constant fixed effect;
ii. An additive spline model with constant and linear fixed effects; and
iii. An interaction spline model with constant and linear fixed effects.

Our conclusions on the design problem can be summarized as follows: If the model only includes constant fixed effect, the design points should be evenly distributed in the design region, whatever the size of the parameter $\lambda$ is. We suggest to use a glp set of $n$ points for the design, since the glp set is relatively simple and has high efficiency for the design problem. If the smoothing spline model includes constant and linear fixed effects, the design for small $\lambda$ has a reasonably even distribution of design points, and so the glp design is preferred. However, as $\lambda$ increases, the design tends towards to the optimal design under a first-order regression model. In this case, when $n$ is a multiple of $2^d$, one can use the $2^d$ factorial design with $n/2^d$ replicates at each corners of the cube.

**Acknowledgement.** We are grateful to the editor and two anonymous referees for helpful comments and extensive suggestions. We thank Professor Kai-Tai Fang who provided valuable comments.

## References

Box GEP, Draper NR (1959) A basis for the selection of a response surface design. J. Amer. Statist. Assoc. 54:622–654

Fang KT, Wang Y (1994) Number-theoretic methods in statistics. London: Chapman and Hall

Gu C, Wahba G (1993) Smoothing spline ANOVA with component-wise Bayesian "confidence intervals". Journal of Computational and Graphical Statistics 2:97–117

Hickernell FJ (1996) Quadrature error bounds with applications to lattice rules. SIAM, J. Numer. Anal. 33:1195–2016

Hua LK, Wang Y (1981) Applications of number theory to numerical analysis. Spring Verlag, Berlin; Science Press, Beijing

Luo Z, Wahba G (1998) Spatial-temporal analysis of temperature using smoothing spline ANOVA. J. Climate 11:18–28

Mitchell T, Sacks J, Ylvisaker D (1994) Asymptotic Bayes criteria for nonparametric response surface design. Ann. Statist. 22:634–651

Sacks J, Welch W, Mitchell T, Wynn H (1989) Design and analysis of computer experiments (with discussion). Statist. Sci. 4:409–423

Steinberg DM (1985) Model robust response surface designs: Scaling two-level factorials. Biometrika 72:513–526

Wahba G (1990) Spline models for observational data. SIAM, Philadelphia

Wahba G, Luo Z (1997) Smoothing spline ANOVA fits for very large, nearly regular data sets, with application to historical global climate data. Ann. Numer. Math. 4:579–596

Wahba G, Wang Y, Gu C, Klein R, Klein B (1995) Smoothing spline ANOVA for exponential families, with application to the Wisconsin Epidemiological Study of Diabetic Retinopathy. Ann. Statist. 23:1865–1895

Wang Y, Wahba G, Gu C, Klein R, Klein B (1997) Using smoothing spline ANOVA to examine the relation of risk factors to the incidence and progression of diabetic retinopathy. Statistics in Medicine 16:1357–1376