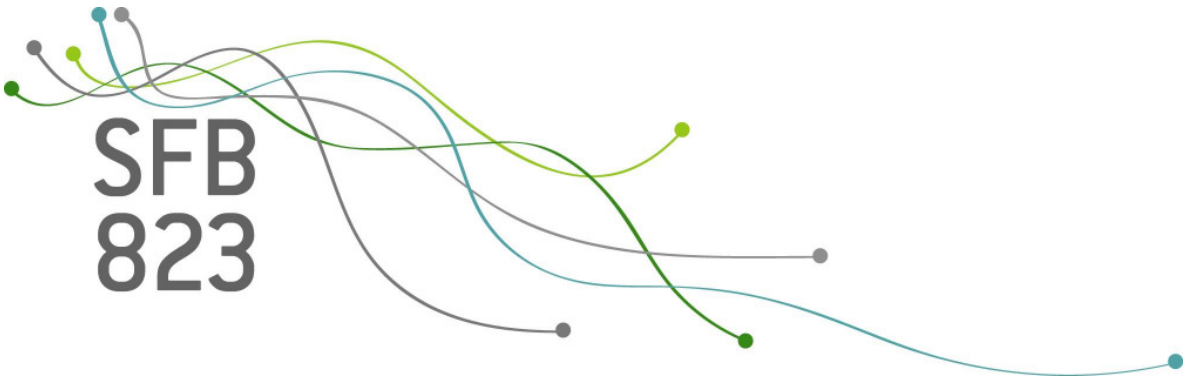
SFB 823

Optimal designs for multivariable spline models

Stefanie Biedermann, Holger Dette,
David C. Woods

Nr. 26/2009

Discussion Paper



Optimal designs for multivariable spline models

Stefanie Biedermann
University of Southampton
School of Mathematics
Highfield SO17 1BJ
UK

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany

David C. Woods
University of Southampton
School of Mathematics
Highfield SO17 1BJ
UK

September 24, 2009

Abstract

In this paper, we investigate optimal designs for multivariate additive spline regression models. We assume that the knot locations are unknown, so must be estimated from the data. In this situation, the Fisher information for the full parameter vector depends on the unknown knot locations, resulting in a non-linear design problem. We show that locally, Bayesian and maximin D-optimal designs can be found as the products of the optimal designs in one dimension. A similar result is proven for Q-optimality in the class of all product designs.

Keywords and Phrases: Additive spline model, Bayesian D-optimality, maximin D-optimality, Q-optimality, product designs,

1 Introduction

Polynomial spline modeling is a popular statistical technique for nonparametric function estimation, because of its similarity to polynomials and its conceptual simplicity [see e.g. De Boor

(1978), Dierckx (1995) or Eubank (1999) among many others]. These concepts have first been considered in the case of a univariate predictor and have then been extended to the situation of a multivariate predictor usually combining univariate basis functions from different components of the predictor [see for example Friedman (1991), Stone et al. (1997) among others]. Several authors propose to use least squares splines [see e.g. Hartley (1961) and Gallant and Fuller (1973) for some early references]. If the knots are assumed to be fixed, this approach is particularly attractive because of its computational simplicity. Grove et al. (2004) use B-splines to analyse data from an engine-mapping experiment from the automotive industry. For some further applications of spline models in the context of dynamic programming, computer models and chromatography we refer to Chen et al. (1999), Siddapa et al. (2007), Fang (2006) and Put et al. (2004). If the knots are also estimated from the data the estimation problem is a nonlinear least squares problem and the computation of the estimate and appropriate designs is substantially more difficult [see e.g. Jupp (1978) or Mao and Zhao (2003)].

In the case where the knots are assumed to be known several authors have studied the problem of constructing optimal designs for the corresponding segmented univariate polynomial regression models [see e.g. Studden and Van Arman (1969), Studden (1971), Murty (1971a,b), Park (1978), Kaishev (1989), Heiligers (1998, 1999), Woods and Lewis (2006) among others]. If the design of experiment should also address the precise estimation of the knots, the situation is substantially more difficult. Recently, Dette et al. (2008) considered the problem of constructing optimal designs for (univariate) free knot least squares splines, while Dette et al. (2009) discussed optimal designs for (univariate) smoothing splines.

Most of the literature on design of experiments for spline regression models discusses the univariate case. Recently Yue and Hickernell (2002) investigated smoothing splines for ANOVA models while Woods et al. (2003) considered optimal design problems for multivariate B-spline models, where the knots are assumed to be known and do not have to be estimated from the data.

This is the first paper to consider multivariable designs for spline models with unknown knots, which is the most prevalent situation in practical problems encountered in industry. The paper is organised as follows. After a brief introduction into the terminology in Section 2 we consider optimal design problems in an additive spline regression model with a K-dimensional predictor and a truncated power basis, where the knots for each coefficient are also estimated from the data. We derive theoretical results on locally, Bayesian and standardised maximin D-optimal designs for these models in Section 3, and apply these results to the engine mapping problem considered by Grove et al. (2004). We also show that the same assertions hold true for D_s-optimal designs for estimating the knots and briefly consider models containing interaction terms. In Section 4, we investigate local and robust Q-optimal designs for predictions at unobserved locations of the predictor. In Section 5, we show that the results derived in Sections 3 and 4 hold for any regression spline basis spanning the same space. Our results are illustrated by several examples throughout the paper, and some conclusions are given in Section 6. Finally, the proofs of our results can be found in an Appendix.

2 Optimal design for additive spline models

The general form of a spline regression function in one variable $x_k \in [a_k, b_k] \subset \mathbb{R}$ using a truncated power basis is given by

$$\mu_k(x_k) \;=\; \theta_{k,1} + \sum_{i=2}^{l_k} \theta_{k,i} x_k^{i-1} + \sum_{i=1}^{r_k} \sum_{j=0}^{l_{ik}-1} \theta_{k,i,j} (x_k - \lambda_{k,i})_+^{m_k - j} =: \theta_{k,1} + \gamma_k^T(x_k, \lambda_k)\theta_{(k)} \qquad (1)$$

where the last identity defines the vector γ_k in an obvious manner, $\theta_{(k)}$ is the vector of all linear parameters in model (1) except the intercept, and $\lambda_k = (\lambda_{k,1}, \ldots, \lambda_{k,r_k})^T \in \Lambda_k$ is the vector of all knots in model (1). Throughout the paper, we define $z_+ = \max(0, z)$ and $\Lambda_k = \{\lambda_k \in [a_k, b_k]^{r_k} \,|\, \lambda_{k,1} < \lambda_{k,2} < \ldots < \lambda_{k,r_k}\}$. Moreover, we assume that $l_k \leq m_k + 1$ and $l_{ik} \leq m_k - 1$ to ensure that the regression function is continuously differentiable.
In many real life problems, there will be more than one explanatory variable, and this situation can be accommodated by fitting additive spline models in K variables of the form

$$\begin{aligned} Y_i \;&=\; \mu(x_{1,i}, \ldots, x_{K,i}) + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2), \quad i = 1, \ldots, n, \\ \mu(x) \;&=\; \theta_1 + \sum_{k=1}^{K} \gamma_k^T(x_k, \lambda_k)\theta_{(k)}, \quad x = (x_1, \ldots, x_K)^T. \end{aligned} \qquad (2)$$

Example 1 *To familiarise with the notation, consider a simple example. Assume we have the quadratic two-factor model ($l_1 = l_2 = 3$) with one knot in each direction ($r_1 = r_2 = 1$), i.e.*

$$\mu(x_1, x_2) = \theta_1 + \theta_{1,2}x_1 + \theta_{1,3}x_1^2 + \theta_{1,1,0}(x_1 - \lambda_{1,1})_+^2 + \theta_{2,2}x_2 + \theta_{2,3}x_2^2 + \theta_{2,1,0}(x_2 - \lambda_{2,1})_+^2. \qquad (3)$$

All exponents of the terms including the knots ($m_1 - j$ and $m_2 - j$) are 2, so $l_{11} - 1 = l_{12} - 1 = 0$ ($l_{11} = l_{12} = 1$) and $m_1 = m_2 = 2$. Thus $l_k = m_k + 1$ and $l_{ik} = m_k - 1$ for $k = 1, 2$, $i = 1$, so the conditions on the degree of the polynomial part and the smoothnesses of the spline part are also satisfied.

The full parameter vector to be estimated is $(\theta^T, \lambda^T)^T$ where $\theta = (\theta_1, \theta_{(1)}^T, \ldots, \theta_{(K)}^T)^T$ and $\lambda = (\lambda_1^T, \ldots, \lambda_K^T)^T$. Obviously, model (2) is linear in the θ-component of the parameter vector, and non-linear in λ. Throughout this paper, we consider approximate designs. An approximate design

$$\xi \;=\; \begin{pmatrix} x_{(1)} & x_{(2)} & \ldots & x_{(m)} \\ w_1 & w_2 & \ldots & w_m \end{pmatrix}$$

is a probability measure with finite support on the design space $\chi = [a_1, b_1] \times \ldots \times [a_K, b_K]$, i.e. $x_{(i)} \in \chi$, $i = 1, \ldots, m$. Without loss of generality, we let $a_k = -1$, $b_k = 1$ for all $k = 1, \ldots, K$. The observations are taken at the support points of the design, and the number of observations in each point $x_{(i)}$ is proportional to the weight w_i.

The Fisher information of a design ξ is then given by the matrix

$$M(\xi,\theta,\lambda) = \int g(x,\theta,\lambda)g^T(x,\theta,\lambda)\,d\xi(x) = \sum_{i=1}^{m} w_i\, g(x_{(i)},\theta,\lambda)g^T(x_{(i)},\theta,\lambda)$$

where $g(x,\theta,\lambda)$ is the vector of derivatives of the regression function $\mu(x)$ with respect to the model parameters, i.e. $g(x,\theta,\lambda) = (1, g_1^T(x_1,\theta_{(1)},\lambda_1)),\ldots,g_K^T(x_K,\theta_{(K)},\lambda_K))^T$ with

$$g_k^T(x_k,\theta_{(k)},\lambda_k)) \;=\; \left(\left(\frac{\partial\mu_k(x_k)}{\partial\theta_{(k)}}\right)^T, \left(\frac{\partial\mu_k(x_k)}{\partial\lambda_k}\right)^T\right).$$

The Fisher information can be expressed as

$$M(\xi,\theta,\lambda) \;=\; C_\theta I(\xi,\lambda)C_\theta^T \tag{4}$$

where C_θ is a nonsingular block-diagonal square matrix depending only on θ, but neither on λ nor the design ξ, and

$$I(\xi,\lambda) = \int f(x,\lambda)f^T(x,\lambda)\,d\xi(x).$$

The vector $f(x,\lambda)$ is defined by $f(x,\lambda) = (1, \tilde{f}_1^T(x_1,\lambda_1),\ldots,\tilde{f}_K^T(x_K,\lambda_K))^T$ where the components of the vector $\tilde{f}_k(x_k,\lambda_k)$, $k=1,\ldots,K$, are given by

$$(\tilde{f}_k(x_k,\lambda_k))_q \;=\; \begin{cases} x_k^q & q = 1,\ldots,l_k-1 \\ (x_k-\lambda_{k,1})_+^{m_k-q+1+\alpha_0} & q = \alpha_0+1,\ldots,\alpha_1 \\ (x_k-\lambda_{k,2})_+^{m_k-q+1+\alpha_1} & q = \alpha_1+1,\ldots,\alpha_2 \\ \vdots & \vdots \\ (x_k-\lambda_{k,r_k})_+^{m_k-q+1+\alpha_{r_k-1}} & q = \alpha_{r_k-1}+1,\ldots,\alpha_{r_k} \end{cases} \tag{5}$$

where $\alpha_j = l_k - 1 + \sum_{s=1}^{j}(l_{sk}+1)$, $j=0,\ldots,r_k$, and $q=1,\ldots,p_k$ with p_k+1 the number of parameters in the k^{th} single factor model (1). Similarly, for each single factor model $\mu_k(x_k)$, the Fisher information can be expressed as

$$M_k(\xi_k,\theta_k,\lambda_k) = C_{\theta_k} I(\xi_k,\lambda_k)C_{\theta_k}^T$$

where $I_k(\xi_k,\lambda_k) = \int f_k(x_k,\lambda_k)f_k^T(x_k,\lambda_k)\,d\xi_k(x_k)$ and the vector $f_k(x_k,\lambda_k)$ equals $(1,\tilde{f}_k^T(x_k,\lambda_k))^T$. The non-singular matrices C_{θ_k}, $k=1,\ldots,K$, without their first rows and columns, respectively, form the blocks around the main diagonal of the matrix C_θ, together with the first block consisting of the value 1.

From (4) we obtain that both the single factor models (1) and the additive model (2) are partially nonlinear models in the sense of Hill (1980) and Khuri (1984). Using properties of the determinant, it follows that

$$|M(\xi,\theta,\lambda)| = |C_\theta I(\xi,\lambda)C_\theta^T| = |C_\theta|^2\,|I(\xi,\lambda)|,$$

so the same design $\xi^*_{D,\lambda}$ will maximise the determinants of the Fisher information $M(\xi, \theta, \lambda)$ and of the more manageable matrix $I(\xi, \lambda)$, which will be denoted as information matrix in what follows. The design $\xi^*_{D,\lambda}$ will only depend on the vector of the unknown knot locations λ, but not on the linear parameters θ. Following Chernoff (1953), we call a design $\xi^*_{D,\lambda}$ locally D-optimal if it maximises the determinant of the Fisher information matrix for given λ, i.e.

$$\xi^*_{D,\lambda} = \arg\max_{\xi} |I(\xi, \lambda)|.$$

3 D- and D_s-optimal designs

3.1 Locally and robust D-optimal designs

The concept of local D-optimality requires knowledge of the unknown parameter vector λ. If λ is misspecified at the design stage, the design may be inefficient, i.e. will not allow accurate estimation of the model parameters. Several approaches to overcome the parameter dependency of optimal designs in nonlinear models have been suggested. We will focus on two non-sequential concepts: Bayesian D-optimality (see, e.g. Chaloner and Verdinelli, 1995) and standardised maximin D-optimality (Imhof, 2001).

When some prior knowledge about the location of the knots is available, which can be summarised in a prior distribution $\pi(\lambda)$ on Λ where $\Lambda = \Lambda_1 \times \Lambda_2 \times \ldots \times \Lambda_K$, it is reasonable to use a Bayesian optimality criterion which averages the original criterion over the plausible values for λ. The Bayesian D-optimality criterion function with respect to the prior π on Λ is given by

$$\Phi_{D,\pi}(\xi) \;=\; \int_{\Lambda} \log |I(\xi, \lambda)| \, d\pi(\lambda), \tag{6}$$

and is maximised with respect to the design ξ.

If only an interval for each knot can be specified, the problem of specifying a prior on the knots can be avoided by using a maximin approach guarding the experiment against the worst case scenario. This is a more cautious approach than the Bayesian, and is recommended in the absence of sufficient prior knowledge. The standardised maximin D-optimality criterion is

$$\Psi_{D,\Lambda_M}(\xi) \;=\; \inf_{\lambda \in \Lambda_M} \Phi(\xi, \lambda) = \inf_{\lambda \in \Lambda_M} \frac{|I(\xi, \lambda)|}{|I(\xi^*_{D,\lambda}, \lambda)|} \tag{7}$$

where $\xi^*_{D,\lambda}$ is the locally D-optimal design with respect to λ and $\Lambda_M \subseteq \Lambda$ is the set in which λ is supposed to lie.

The following result states how Bayesian and standardised maximin D-optimal designs for the additive model (2) can be constructed from the corresponding Bayesian and standardised maximin D-optimal designs for the single factor models (1).

Theorem 1 *(a) Let $\pi(\lambda)$ be a prior for $\lambda \in \Lambda$ with marginals $\pi_k(\lambda_k)$, $k = 1, \ldots, K$. Then the product design $\xi^*_{D,\pi} = \xi^*_{D,\pi_1} \otimes \xi^*_{D,\pi_2} \otimes \ldots \otimes \xi^*_{D,\pi_K}$ is Bayesian D-optimal with respect*

to $\pi(\lambda)$ for the additive model, where $\xi^*_{D,\pi_1},\ldots,\xi^*_{D,\pi_K}$ are the Bayesian D-optimal designs with respect to $\pi_k(\lambda_k)$ in the single factor models (1).

(b) Let $\xi^*_{D,\Lambda_{M,1}},\ldots,\xi^*_{D,\Lambda_{M,K}}$ be the standardised maximin D-optimal designs with respect to $\Lambda_{M,k}$, $k=1,\ldots,K$, in the single factor models (1) for compact parameter spaces $\Lambda_{M,k}\subset\Lambda_k$. Then the product design $\xi^*_{D,\Lambda_M}=\xi^*_{D,\Lambda_{M,1}}\otimes\xi^*_{D,\Lambda_{M,2}}\otimes\ldots\otimes\xi^*_{D,\Lambda_{M,K}}$ is standardised maximin D-optimal with respect to $\Lambda_M=\Lambda_{M,1}\times\ldots\times\Lambda_{M,K}$ in the additive model (2).

See Appendix A.1 for the proof of Theorem 1. Local D-optimality can be viewed as a special case of Bayesian D-optimality with a point mass prior on λ.
Applying Theorem 1 and Theorem 3.1 in Dette et al. (2008) we can find an explicit solution of the locally D-optimal design problem for the additive model if the regression function has exactly one continuous derivative at each knot.

Corollary 1 *Let $m_k\geq l_k-1$ and $l_{ik}=m_k-1$ for all i. Then the locally D-optimal design $\xi^*_{D,\lambda}$ with respect to $\lambda=(\lambda_1^T,\ldots,\lambda_K^T)^T$ in the additive model (2) on the design space χ is the product design with marginals ξ^*_{D,λ_k}, $k=1,\ldots,K$, where the support points $x_{k,1},\ldots,x_{k,p_k+1}$ of ξ^*_{D,λ_k} have equal weights $w_{k,i}=1/(p_k+1)$, $i=1,\ldots,p_k+1$, where p_k+1 is the number of parameters in the marginal model μ_k. The support points are given by*

$$\begin{aligned} x_{k,i} &= a_k+(\nu_{k,i,l_k}+1)\Big(\frac{\lambda_{k,1}-\lambda_{k,0}}{2}\Big), \quad i=1,\ldots,l_k,\\ x_{k,i-1+l_k+(s-1)m_k} &= \lambda_s+(\nu_{k,i,m_k+1}+1)\Big(\frac{\lambda_{k,s+1}-\lambda_{k,s}}{2}\Big), \quad i=2,\ldots,m_k+1,\ s=1,\ldots,r_k, \end{aligned}$$

where $\lambda_{k,0}=a_k$, $\lambda_{k,r_k+1}=b_k$, $\nu_{k,1,t},\ldots,\nu_{k,t,t}$ are the ordered roots of the polynomial $(x_k^2-1)L'_{t-1}(x_k)$ and $L'_t(x_k)$ is the derivative of the t^{th} Legendre polynomial.

Example 2 *Consider Example 1 on the design space $[-1,1]^2$. Note that the condition $l_{ik}=m_k-1$ for all i from Corollary 1 is satisfied. For both marginal models $l_k=m_k+1=3$, so the required values are $\nu_{k,1,3}=-1,\nu_{k,2,3}=0$ and $\nu_{k,3,3}=1$ for $k=1,2$. Applying Corollary 1, we obtain the general form of the locally D-optimal marginal designs as*

$$\xi^*_{D,\lambda_k}=\begin{pmatrix} \lambda_{k,0} & (\lambda_{k,0}+\lambda_{k,1})/2 & \lambda_{k,1} & (\lambda_{k,1}+\lambda_{k,2})/2 & \lambda_{k,2}\\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \end{pmatrix}.$$

For $\lambda_{1,1}=-0.5$ and $\lambda_{2,1}=0.5$ ($\lambda_{k,0}=-1,\lambda_{k,2}=1$) we obtain

$$\xi^*_{D,\lambda_1}=\begin{pmatrix} -1 & -0.75 & -0.5 & 0.25 & 1\\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \end{pmatrix},$$

$$\xi^*_{D,\lambda_2}=\begin{pmatrix} -1 & -0.25 & 0.5 & 0.75 & 1\\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \end{pmatrix}$$

and the product design of these marginals is locally D-optimal in the two-factor model.

Although the D-optimal information matrix $I(\xi^*_{D,\lambda}, \lambda)$ for the additive model (2) is uniquely determined due to the strict log-concavity of the determinant criterion, the locally D-optimal designs are not necessarily unique even if the D-optimal designs for the single factor models are unique. Remark 1 characterises the support of any D-optimal design for the additive model (2).

Remark 1 *The support of any multi-factor D-optimal design must be contained in the support of the product of the corresponding one-dimensional D-optimal designs.*

The proof of Remark 1 can be found in Appendix A.2.
The number of support points of product designs quickly increases in higher dimensions. To investigate if all support points of the product design are required, or if a subset will be sufficient we consider the model from Examples 1 and 2. The product design consists of 25 support points to estimate 9 parameters. Numerical evidence shows that there is, however, no locally D-optimal design with fewer support points than 25. Figure 3.1 shows the efficiencies of the locally D-optimal m-point designs, found numerically, relative to the product design for $\lambda = (-0.5, 0.5)^T$, where $9 \leq m \leq 25$ is the number of support points. The efficiency is an increasing function of m up to $m = 25$. We obtained similar results numerically for a variety of models in two variables.

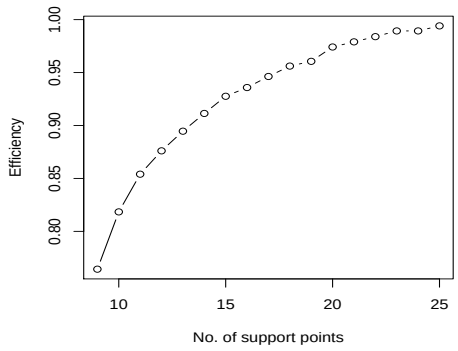


Figure 1: *Efficiencies of locally D-optimal designs with 9 - 25 support points, relative to the product optimal design, for the two factor spline model with terms* $(1, x_1, x_1^2, (x_1 + 0.5)_+^2, x_2, x_2^2, (x_2 - 0.5)_+^2)$.

3.2 D_s-optimal designs for estimating the knots

In some practical problems, the experimenter's main interest is in estimating the knot locations, since they may indicate at which experimental conditions the behaviour of the regression function changes and provide insight into the complexity of the response. In what follows, we therefore investigate D_s-optimal designs for the estimation of the knots. This means we minimise the

determinant of the asymptotic covariance matrix for the estimator of λ, or equivalently, maximise the function

$$\psi_s(M(\xi,\theta,\lambda)) \;=\; |(A^T M^-(\xi,\theta,\lambda)A)^{-1}|. \tag{8}$$

The matrix $A^T = (J_s \mid 0_{s\times(p-s)})$ consists of two blocks, where J_s is the identity matrix of size $s = \sum_{k=1}^K r_k$, $0_{s\times(p-s)}$ is a zero matrix of size $s\times(p-s)$, and $p = 1+\sum_{k=1}^K p_k$ is the total number of model parameters. Without loss of generality, throughout this section we have re-ordered the rows and columns of the information matrix $M(\xi,\theta,\lambda)$ such that the top left corner of size $s\times s$ of this matrix corresponds to the derivatives of the regression function with respect to the knots, and also re-ordered the rows and columns of $I(\xi,\lambda)$ accordingly. Here, $M^-(\xi,\theta,\lambda)$ and $I^-(\xi,\lambda)$ denote the respective generalised inverses of the matrices $M(\xi,\theta,\lambda)$ and $I(\xi,\lambda)$. The design ξ must ensure that the parameters in λ are estimable, i.e. the matrix $A^T M^-(\xi,\theta,\lambda)A$ must be non-singular.
Lemma 1 shows that we can restrict ourselves to considering the simpler problem of maximising $\psi_s(I(\xi,\lambda)) = |(A^T I^-(\xi,\lambda)A)^{-1}|$, and that consequently D_s-optimal designs for estimating the knots in model (2) do not depend on the linear model parameters θ.

Lemma 1 *There exists a positive constant c_θ, depending only on θ but neither on λ nor on the design ξ, such that*

$$|(A^T M^-(\xi,\theta,\lambda)A)^{-1}| = c_\theta |(A^T I^-(\xi,\lambda)A)^{-1}|,$$

where $I(\xi,\lambda)$ is the re-ordered version of the information matrix defined in (5).

The proof of Lemma 1 can be found in Appendix A.3.
We now consider Bayesian and standardised maximin D_s-optimality, where a Bayesian D_s-optimal design with respect to a prior π on Λ maximises

$$\Phi_{D_s,\pi}(\xi) \;=\; \int_\Lambda \log \psi_s(I(\xi,\lambda))\, d\pi(\lambda),$$

and a standardised maximin D_s-optimal design with respect to Λ_M maximises

$$\Psi_{D_s,\Lambda_M}(\xi) \;=\; \inf_{\lambda\in\Lambda_M} \frac{\psi_s(I(\xi,\lambda))}{\psi_s(I(\xi^*_{D_s,\lambda},\lambda))}.$$

Here $\xi^*_{D_s,\lambda}$ denotes the locally D_s-optimal design with respect to λ. Analogous to Section 3.1, we show that the product of designs which are Bayesian (standardised maximin) D_s-optimal for estimating the knots λ_k in the k^{th} marginal model (1) are Bayesian (standardised maximin) D_s-optimal for estimating all knots $\lambda = (\lambda_1^T,\ldots,\lambda_K^T)^T$ in the additive model (2). Local D_s-optimality is embedded in this result as the special case of π being a point mass prior concentrated in some $\lambda\in\Lambda$. The proof of Theorem 2 is in Appendix A.4.

Theorem 2 *(a) Let $\pi(\lambda)$ be a prior for $\lambda\in\Lambda$ with marginals π_k, $k=1,\ldots,K$. Let $\xi^*_{D_s,\pi_1},\ldots,\xi^*_{D_s,\pi_K}$ denote the Bayesian D_s-optimal designs with respect to $\pi_k(\lambda_k)$ in the single factor models. Then the product design $\xi^*_{D_s,\pi} = \xi^*_{D_s,\pi_1}\otimes\xi^*_{D_s,\pi_2}\otimes\ldots\otimes\xi^*_{D_s,\pi_K}$ is Bayesian D_s-optimal with respect to $\pi(\lambda)$ for the additive model.*

(b) Let $\xi^*_{D_s,\Lambda_{M,1}}, \ldots, \xi^*_{D_s,\Lambda_{M,K}}$ be standardised maximin D_s-optimal designs with respect to $\Lambda_{M,k}$, $k = 1, \ldots, K$, in the single factor models (1) for compact parameter spaces $\Lambda_{M,k} \subset \Lambda_k$. Then the product design $\xi^*_{D_s,\Lambda_M} = \xi^*_{D_s,\Lambda_{M,1}} \otimes \xi^*_{D_s,\Lambda_{M,2}} \otimes \ldots \otimes \xi^*_{D_s,\Lambda_{M,K}}$ is standardised maximin D_s-optimal with respect to the parameter space $\Lambda_M = \Lambda_{M,1} \times \ldots \times \Lambda_{M,K}$ in the additive model.

3.2.1 Examples

A numerical study was carried out to investigate locally D_s-optimal designs, and found they have the same support points as the locally D-optimal designs for the same model, but different weights. Some selected single factor locally D_s-optimal designs for quadratic models with one and two knots, respectively, i.e.

$$\begin{aligned} \mu(x_1) &= \theta_1 + \theta_{1,2}x_1 + \theta_{1,3}x_1^2 + \theta_{1,1,0}(x_1 - \lambda_{1,1})_+^2 \\ \mu(x_1) &= \theta_1 + \theta_{1,2}x_1 + \theta_{1,3}x_1^2 + \theta_{1,1,0}(x_1 - \lambda_{1,1})_+^2 + \theta_{1,2,0}(x_1 - \lambda_{1,2})_+^2, \end{aligned} \tag{9}$$

are given in Table 1.

Table 1: *Selected locally D_s-optimal designs for the quadratic single factor models with one or two knots, respectively. The last weight is omitted since the weights sum up to one.*

$\lambda_{1,1}$	$\lambda_{1,2}$	Support points							Weights					
-0.5	–	-1	-0.75	-0.5	0.25	1			0.094	0.375	0.375	0.125		
0	–	-1	-0.5	0	0.5	1			0.063	0.250	0.375	0.250		
0.2	–	-1	-0.4	0.2	0.6	1			0.050	0.200	0.375	0.300		
-0.5	0.5	-1	-0.75	-0.5	0	0.5	0.75	1	0.047	0.188	0.207	0.116	0.207	0.188
0.2	0.5	-1	-0.4	0.2	0.35	0.5	0.75	1	0.018	0.073	0.238	0.246	0.250	0.141

The locally D_s-optimal designs in Table 1 all generate information matrices of full rank, so the full parameter vector is estimable when using these designs. Even if the experimenter's main interest is in estimating the knots, the other parameters are still of some importance. We therefore investigate how efficient locally D_s-optimal designs are for estimating the full parameter vector, i.e. how D-efficient they are. Similarly, when using a locally D-optimal design, it will be interesting to see how well this design performs for estimating the knots only, i.e. to assess its D_s-efficiency. The D-efficiency and the D_s-efficiency of a design ξ are defined by

$$\text{eff}_D(\xi, \lambda) = \left(\frac{|I(\xi, \lambda)|}{|I(\xi^*_{D,\lambda}, \lambda)|} \right)^{1/p}, \quad \text{eff}_{D_S}(\xi, \lambda) = \left(\frac{\psi_s(\xi, \lambda)}{\psi_s(\xi^*_{D_s,\lambda}, \lambda)} \right)^{1/s}.$$

In Table 2, we present D-efficiencies of locally D_s-optimal designs (for the same parameter λ) as well as D_s-efficiencies of locally D-optimal designs in models (9) and some selected two-dimensional models.

Table 2: *Selected D- and D_s-efficiencies of locally D_s- respective D-optimal designs for quadratic models in one or two variables with knots $\lambda_{1,1}$, $\lambda_{1,2}$ and $\lambda_{2,1}$.*

$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\text{eff}_D(\xi^*_{D_s,\lambda}, \lambda)$	$\text{eff}_{D_s}(\xi^*_{D,\lambda}, \lambda)$
-0.5	–	–	0.694	0.652
0	–	–	0.779	0.731
0.2	–	–	0.766	0.718
-0.5	0.5	–	0.850	0.820
0.2	0.5	–	0.695	0.696
-0.5	–	0.5	0.665	0.653
0.2	–	0.5	0.703	0.685
0	–	0	0.757	0.732

We can see from Table 2 that the D- and D_s-efficiencies in these examples are between 65% and 85%. This was confirmed for a variety of different scenarios (not listed). Hence D- (D_s)-optimal designs maintain moderate efficiency for the estimation of only the knot locations (all parameters). However, to avoid a loss in accuracy it is recommended to run the design which has been constructed for the respective purpose.

3.3 Application - Robustness of Bayesian D-optimal designs

The purpose of engine mapping experiments as considered in Grove et al. (2004) is to model a measure of engine performance as a function of several adjustable engine variables. The data for such an experiment described in Woods et al. (2003) give rise to an additive spline model for the maximum brake torque timing of an engine in the three variables "speed", "load" and "air-fuel ratio". The corresponding single factor models are the cubic spline model

$$\mu_1(x_1) \;=\; \theta_{1,1} + \theta_{1,2}x_1 + \theta_{1,3}x_1^2 + \theta_{1,4}x_1^3 + \theta_{1,1,0}(x_1 - \lambda_{1,1})_+^3 \tag{10}$$

for the variable "speed" and quadratic polynomials for "load" and "air-fuel ratio", respectively. We use this model to assess the robustness of locally and Bayesian D-optimal designs, in order to investigate if it is necessary to calculate the numerically more demanding Bayesian D-optimal design, or if locally D-optimal designs will be sufficiently robust.
The data imply that the knot $\lambda_{1,1}$ should be in the interval $[0, 0.6]$. We compare the locally D-optimal design for the midpoint, i.e. $\lambda_{1,1} = 0.3$, with the Bayesian D-optimal design with respect to the uniform prior on $[0, 0.6]$ using two different approximations for this prior: π_1, the uniform distribution on the seven points $0, 0.1, \ldots, 0.6$ as a crude but simple approximation; π_2, the uniform distribution on 121 equidistant points from 0 to 0.6 as an approximation close to the continuous prior. To compare designs we define the relative D-efficiency of a design ξ_1 compared

with a design ξ_2 as

$$\text{eff}_{rel,D}(\xi_1, \xi_2, \lambda) = \left(\frac{|I(\xi_1, \lambda)|}{|I(\xi_2, \lambda)|} \right)^{1/p}.$$

The Bayesian D-optimal designs with respect to the priors π_1 and π_2 were calculated numerically, and the locally D-optimal design for $\lambda_{11} = 0.3$ and the Bayesian D-optimal design with respect to π_1 for the single factor model (10) are depicted in Figure 2. The Bayesian D-optimal design with respect to π_2 is very similar to the corresponding design for the cruder approximation and therefore not shown. Figure 2 also shows the relative D-efficiencies of the Bayesian D-optimal designs with respect to π_1 and π_2, respectively, compared with the locally D-optimal design for $\lambda_{11} = 0.3$. Here $p = 10$, as we consider the three-factor model described above.

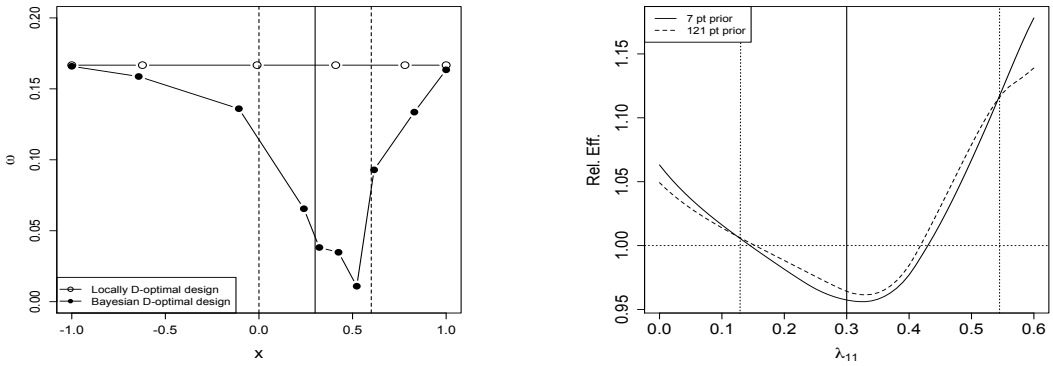


Figure 2: *Left: Support points and weights of the locally D-optimal design for $\lambda_{11} = 0.3$ and the Bayesian D-optimal design with respect to π_1 for the single factor model (10). Right: Relative D-efficiencies of the Bayesian D-optimal designs with respect to π_1 and π_2 compared with the locally D-optimal design for $\lambda_{11} = 0.3$, plotted against the possible knots in the interval $[0, 0.6]$.*

Figure 2 shows that in the interval from about 0.14 to 0.42 the Bayesian D-optimal designs are slightly less efficient, but outperform the locally D-optimal designs if the knot is closer to the boundary. Both Bayesian designs have similar relative D-efficiencies, with the Bayesian D-optimal design with respect to π_1 being slightly better around the boundary, and the Bayesian D-optimal design with respect to π_2 being somewhat more efficient in the interior. For a large area of uncertainty it is recommended to use a Bayesian D-optimal design, where the level of approximation to a continuous prior does not seem to have a large impact on design performance.

3.4 D-optimal designs for models with interactions

So far we have considered the additive model (2), but there might occur situations where the factors do not work independently on the response but interact. In this case, the models are no longer partially nonlinear. The decomposition (4) of the Fisher information can no longer be achieved with a matrix C which does not depend on the design. Therefore, optimal designs

(with respect to any of the criteria considered in this paper) do not only depend on the knots but also on some of the linear parameters.
For the D-optimal designs we have found numerically for the model

$$\begin{aligned}\mu(x) &= \theta_1 + \theta_2 x_1 + \theta_3 x_1^2 + \theta_4 (x_1 - \lambda_{1,1})_+^2 + \theta_5 x_2 + \theta_6 x_1 x_2 + \theta_7 x_1^2 x_2 + \theta_8 (x_1 - \lambda_{1,1})_+^2 x_2 \\ &+ \theta_9 x_2^2 + \theta_{10} x_1 x_2^2 + \theta_{11} x_1^2 x_2^2 + \theta_{12} (x_1 - \lambda_{1,1})_+^2 x_2^2 + \theta_{13} (x_2 - \lambda_{2,1})_+^2 + \theta_{14} x_1 (x_2 - \lambda_{2,1})_+^2 \\ &+ \theta_{15} x_1^2 (x_2 - \lambda_{2,1})_+^2 + \theta_{16} (x_1 - \lambda_{1,1})_+^2 (x_2 - \lambda_{2,1})_+^2 \end{aligned} \quad (11)$$

the dependence of design performance on θ is quite weak. This can also be seen in Table 3, which shows the D-efficiencies of several D-optimal designs for the additive model (2) (which do not depend on θ) relative to the D-optimal designs for the full model for different values of θ. We can also see from Table 3 that the D-optimal designs for the additive model are highly efficient in the full interaction model with efficiencies of over 90%.

Table 3: Selected D-efficiencies of D-optimal designs for the additive model in the full interaction model. $\theta_1 = (1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)^T$, $\theta_2 = (1,1,1,1,1,1,1,1,2,2,2,3,3,9,4,5,6)^T$, and $\theta_3 = (1,1,2,3,-3,-2,-1,4,2,-4,2,-1,1,-7,8,-11)^T$.

$\lambda_{1,1}$	$\lambda_{1,2}$	θ	$\text{eff}_D(\xi^*_{D,\lambda}, \lambda)$
-0.5	0.5	θ_1	0.901
-0.5	0.5	θ_2	0.901
-0.5	0.5	θ_3	0.930
0.2	0.5	θ_1	0.912
0.2	0.5	θ_2	0.913
0.2	0.5	θ_3	0.913

Although somewhat counter-intuitive, the D-optimal designs for the full interaction model found here have fewer support points than the D-optimal designs for the additive model, i.e. a model with much fewer parameters than the interaction model. For example, the D-optimal design for the full interaction model with $\lambda = (0.2, 0.5)$ and $\theta = \theta_1$ has 19 support points, many of which actually coincide with the support points of the D-optimal design for the additive model. For illustration, this design, together with the D-optimal design for the additive model is shown in Figure 3, together with the corresponding standardised variances.

4 Optimal designs for prediction of the response surface

Often, the experimenter is rather interested in the prediction of the response surface at different points than in the particular values of the unknown parameters. A first order approximation to

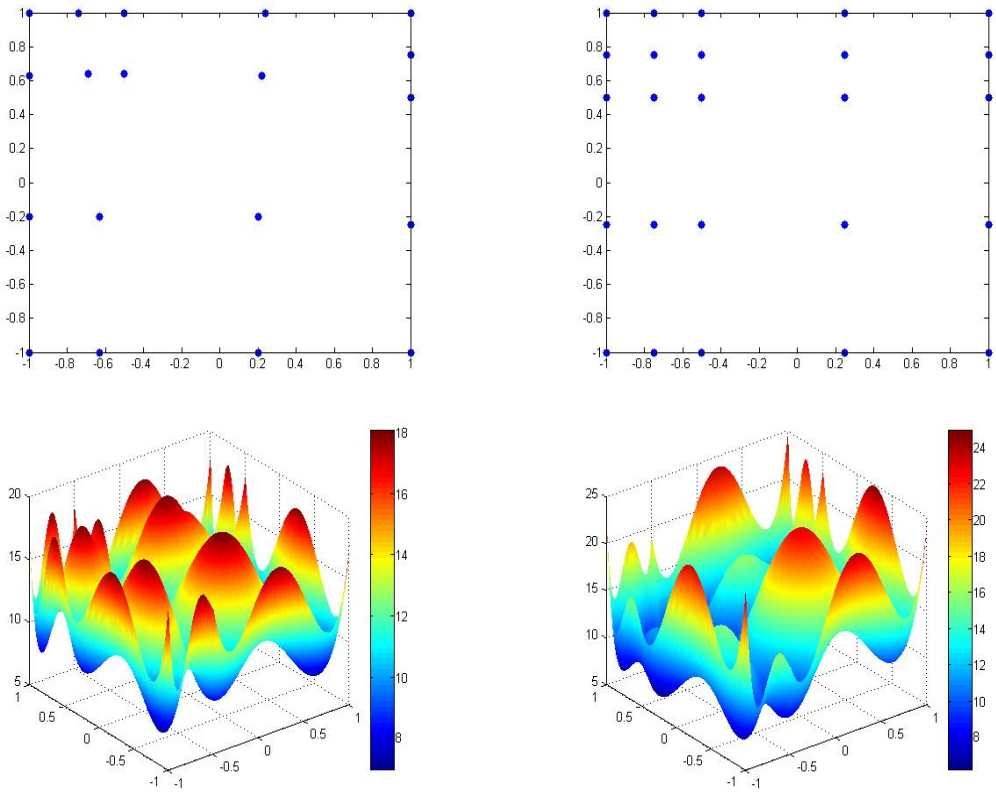


Figure 3: *Top left: Support points of the D-optimal design for the full interaction model with* $\lambda = (0.2, 0.5)$ *and* $\theta = \theta_1$*. Top right: Support points of the D-optimal design for the additive model with* $\lambda = (0.2, 0.5)$*. Bottom left: Standardised variance of the D-optimal design for the full interaction model. Bottom right: Standardised variance of the D-optimal design for the additive model evaluated under the full interaction model.*

the variance of $\hat{\mu}(x)$ at some point $x = (x_1, \ldots, x_K) \in \mathbb{R}^K$ is given by

$$Var\,(\hat{\mu}(x)) \;=\; g^T(x, \theta, \lambda) M^{-1}(\xi, \theta, \lambda) g(x, \theta, \lambda) = f^T(x, \lambda) I^{-1}(\xi, \lambda) f(x, \lambda).$$

Naturally, it is appealing to minimise this variance jointly for a user-selected choice of values for x, reflected in a distribution $H(x)$. So the goal is to minimise the objective function

$$Q(\xi, \lambda) \;=\; \int f^T(x, \lambda) I^{-1}(\xi, \lambda) f(x, \lambda)\, dH(x). \tag{12}$$

Since the matrix $Q := \int f(x, \lambda) f^T(x, \lambda)\, dH(x)$ is non-negative definite we can decompose it into $Q = GG^T$ where G is a lower triangular matrix (Cholesky Decomposition), and the objective function can be expressed as

$$Q(\xi, \lambda) = \mathrm{tr}(G^T I^{-1}(\xi, \lambda) G) = \mathrm{tr}(G G^T I^{-1}(\xi, \lambda)).$$

To achieve robustness against misspecification of the knots, we seek Bayesian Q-optimal designs with respect to a prior $\pi(\lambda)$, which minimise

$$\Phi_{Q,\pi}(\xi) \;=\; \int_{\Lambda} Q(\xi,\lambda)\, d\pi(\lambda). \tag{13}$$

Similarly, a minimax Q-optimal design minimises

$$\Psi_{Q,\Lambda_M}(\xi) \;=\; \max_{\lambda\in\Lambda_M} Q(\xi,\lambda). \tag{14}$$

Theorem 3, which is proven in Appendix A.5, establishes the main result of this section, i.e. that the product design of the Bayesian (minimax) Q-optimal designs in the marginal models (1) is Bayesian (minimax) Q-optimal for the additive model (2) in the class of all product designs.

Theorem 3 *Let π be a prior on $\lambda \in \Lambda$ with marginals π_k on Λ_k, $k = 1,\dots,K$, $\Lambda_M = \Lambda_{M,1} \times \dots \times \Lambda_{M,K}$ a compact subset of Λ, and the weighting measure $H(x)$ be a product measure with marginals $H_1(x_1),\dots,H_K(x_K)$.*

(a) The product design of the Bayesian Q-optimal designs for the single factor models with respect to H_k and π_k, $k = 1,\dots,K$, is Bayesian Q-optimal within the class of all product designs with respect to H and π.

(b) The product design of the minimax Q-optimal designs for the single factor models with respect to H_k and Λ_k, $k = 1,\dots,K$, is minimax Q-optimal within the class of all product designs with respect to H and Λ.

4.1 Performance of Q-optimal product designs

Since the product designs of the Q-optimal designs in the single factor models (1) are Q-optimal within the class of product designs, but not necessarily optimal among all designs, we investigate their performance relative to the Q-optimal designs found by numerical search.
Table 4 shows a selection of Q-optimal designs for the quadratic single factor models (9). Here, H_1 is the uniform distribution on the design interval $[-1, 1]$, i.e. the goal is to predict accurately over the whole range of the variable x_1. The designs were found numerically, using a multiplicative algorithm which updates the weights on a grid in each step, and H_1 was approximated by a discrete uniform distribution on 1000 equidistant points in this interval.
We can see from Table 4 that although the Q-optimal designs are all minimally supported they are far from being equally weighted. All Q-optimal designs are supported at the end points of the design interval, -1 and 1, at the knots, and at the points approximately in the middle between the endpoints and the knots. The support is thus almost identical to the support of the locally D-optimal designs with respect to the same knot locations.
Comparing the products of these designs with the corresponding Q-optimal designs in two variables, we found that the support was identical, but the weights differed to some extent. The

Table 4: Selected Q-optimal designs for the quadratic single factor models with one or 2 knots, respectively. The last weight is omitted since the weights sum up to one.

$\lambda_{1,1}$	$\lambda_{1,2}$	Support points							Weights					
-0.5	–	-1	-0.76	-0.5	0.25	1			0.101	0.195	0.196	0.338		
-0.2	–	-1	-0.61	-0.2	0.405	1			0.123	0.241	0.191	0.295		
0	–	-1	-0.51	0	0.505	1			0.136	0.269	0.191	0.269		
0.2	–	-1	-0.41	0.2	0.605	1			0.150	0.295	0.191	0.241		
0.5	–	-1	-0.255	0.5	0.755	1			0.171	0.338	0.196	0.195		
-0.5	0.5	-1	-0.755	-0.5	0	0.5	0.755	1	0.084	0.162	0.140	0.229	0.140	0.162
-0.5	0.2	-1	-0.755	-0.5	-0.155	0.2	0.605	1	0.083	0.160	0.124	0.189	0.139	0.202

Q-optimal product designs are therefore not Q-optimal among all designs, but they turn out to be very highly efficient. Table 5 shows some Q-efficiencies of Q-optimal product designs $\xi_{Q,\lambda}$ where the Q-efficiency of a design ξ is defined as the ratio

$$\text{eff}_Q(\xi, \lambda) = \frac{Q(\xi^*_{Q,\lambda}, \lambda)}{Q(\xi, \lambda)},$$

with $\xi^*_{Q,\lambda}$ denoting the locally Q-optimal design. The last column shows the Q-efficiency of the D-optimal design for the same knot locations for comparison.

Table 5: *Selected Q-efficiencies of locally Q-optimal product designs and locally D-optimal designs with respect to the knots $\lambda_{1,1}$, $\lambda_{1,2}$ and $\lambda_{2,1}$ in the quadratic two factor model.*

$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\text{eff}_Q(\xi_{Q,\lambda}, \lambda)$	$\text{eff}_Q(\xi^*_{D,\lambda}, \lambda)$
-0.5	–	0.5	0.9978	0.8505
-0.5	–	-0.2	0.9984	0.8746
0	–	0	0.9991	0.9071
0.2	–	0.5	0.9985	0.8746
-0.5	0.5	0.5	0.9989	0.8767
-0.5	0.2	-0.2	0.9993	0.9100

We find that all Q-optimal product designs have an efficiency of more than 99%. From a plot of the directional derivative of the objective function we can see, however, that they are not Q-optimal. Figure 4 shows a plot of the directional derivative of the objective function for the Q-optimal design and the Q-optimal product design, both for the quadratic two-factor model with knots $\lambda_{1,1} = -0.5$, $\lambda_{1,2} = 0.5$ and $\lambda_{2,1} = 0.5$.

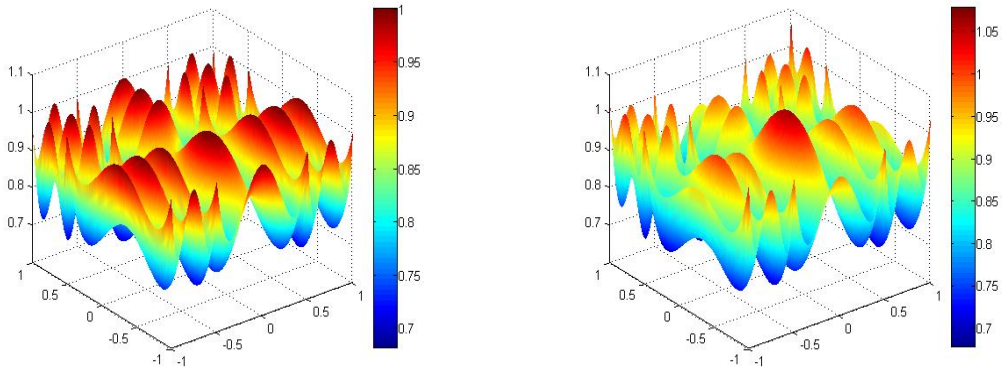


Figure 4: *The directional derivative of the criterion function evaluated at the Q-optimal design (left) and the Q-optimal product design (right), both for the quadratic two-factor model with knots* $\lambda_{1,1} = -0.5$, $\lambda_{1,2} = 0.5$ *and* $\lambda_{2,1} = 0.5$.

Figure 4 indicates that the product design does not put enough weight on the vertices and one point towards the centre of the design range, resulting in values of the directional derivative greater than 1 at these points.
We also assessed the performance of the product of single factor Bayesian Q-optimal designs with marginal priors π_k relative to the optimal multifactor design for the prior π. It turned out that in this situation the support of the Q-optimal multifactor design was usually slightly larger than (but still similar to) the support of the optimal product design. The efficiencies of the optimal product designs are still high, but not quite as impressive as in the local case. For example, the Q-optimal product design for the quadratic two-factor model (3) with one knot in each direction from Example 1 has Bayesian Q-efficiency of 0.8974 for the four point prior with equal weights on the knot locations $\{(-0.5, -0.5), (-0.5, 0.5), (0.5, -0.5), (0.5, 0.5)\}$.

5 Other spline bases - a generalisation

The following result, which is proven in Appendix A.6, relates D- and Q-optimal designs for the splines generated by the truncated power basis to splines from other bases. Truncated power bases can suffer from severe ill-conditioning (De Boor, 1978) whereas for example B-spline bases are well-conditioned and have local support, thus facilitating computation.

Theorem 4 *The results presented in Theorems 1 and 3 for the truncated power basis are also valid for any regression spline basis spanning the same space, i.e. a change in the spline basis does not affect the optimal designs with respect to the local and robust D- and Q-optimality criteria.*

6 Conclusion/Discussion

We have shown that D-optimal designs for the full parameter vector and D_s-optimal designs for the knots λ in additive multivariable spline models of the form (2) with unknown knot locations can be found as the products of the D- and D_s-optimal designs in the corresponding single factor models (1), thus reducing computational effort for calculating optimal multi-factor designs, as it is sufficient to compute the corresponding optimal designs in one variable. Since model (2) is partially nonlinear these designs depend on the unknown knots λ, and misspecifications of these parameters can lead to poor designs. Hence we have generalised our results to parameter robust optimality criteria, namely Bayesian and standardised maximin D- and D_s-optimality.
In many situations, the main goal of an experiment is to predict the response at unobserved locations. We have considered Bayesian/minimax Q-optimal designs and shown that the products of Bayesian/minimax Q-optimal designs in the single factor models are optimal in the additive model within the class of product designs. In a numerical study, we have found that the Q-optimal product designs are - if not Q-optimal among all designs - extremely efficient.
We have finally shown that all results on local and robust D- and Q-optimality presented above are valid regardless of the choice of regression spline basis.
We have illustrated our results throughout the paper through examples and applied them to our motivating example on engine mapping (Grove et al., 2004, Woods et al., 2003). We hope that our work will facilitate the utilisation of optimal designs either directly or indirectly in upcoming experiments in the automotive industry.
We finally note that for some applications interactions between the explanatory variables might be present. We have briefly discussed the full interaction model in Section 3.4, but the true model might lie somewhere in between the strictly additive and the full interaction model. Future work is planned on optimal design for model selection to find the best fit. Another issue to be pursued in future work is designing experiments efficiently in the situation when the number of knots is unknown.

Acknowledgements: The support of the British Council, the Deutscher Akademischer Austausch Dienst (DAAD) and the Deutsche Forschungsgemeinschaft (SFB 823, "Statistik nichtlinearer dynamischer Prozesse") is gratefully acknowledged. D.C. Woods was partially supported by a grant from the Defence Threat Reduction Agency. The authors would also like to thank M. Trampisch for his expert computational assistance.

A Proofs

A.1 Proof of Theorem 1

For clarity of presentation, in what follows we restrict ourselves to proving the assertion of the Theorem for $K = 2$. The general case $K \geq 2$ follows by defining meta-factors consisting of more than one single factor and applying the result for $K = 2$.

(a) Let ξ_1, ξ_2 denote the marginals of the design ξ. The special form of the information matrices permits application of Lemma 5.1 in Schwabe (1996), which is stated for information matrices of linear models, and we find that

$$|I(\xi, \lambda)| \;\leq\; |I_1(\xi_1, \lambda_1)|\, |I_2(\xi_2, \lambda_2)| \;=\; |I(\xi_1 \otimes \xi_2, \lambda)|. \tag{15}$$

Using this inequality, the following holds:

$$\begin{aligned}
\int_\Lambda \log|I(\xi^*_{D,\pi_1} \otimes \xi^*_{D,\pi_2}, \lambda)|\, d\pi(\lambda) \;&\leq\; \max_{\xi} \int_\Lambda \log|I(\xi, \lambda)|\, d\pi(\lambda) \\
&\leq\; \max_{\xi_1,\xi_2} \int_\Lambda \log(|I_1(\xi_1, \lambda_1)|\; |I_2(\xi_2, \lambda_2)|)\, d\pi(\lambda) \\
&=\; \max_{\xi_1} \int_{\Lambda_1} \log|I_1(\xi_1, \lambda_1)|\, d\pi_1(\lambda_1) \\
&\quad + \max_{\xi_2} \int_{\Lambda_2} \log|I_2(\xi_2, \lambda_2)|\, d\pi_2(\lambda_2) \qquad (16) \\
&=\; \int_{\Lambda_1} \log|I_1(\xi^*_{D,\pi_1}, \lambda_1)|\, d\pi_1(\lambda_1) + \int_{\Lambda_2} \log|I_2(\xi^*_{D,\pi_2}, \lambda_2)|\, d\pi_2(\lambda_2)
\end{aligned}$$

So substituting the Bayesian D-optimal designs $\xi^*_{D,\pi_1}, \xi^*_{D,\pi_2}$ for the single factor models into (16) gives an upper bound for the maximal determinant for the two-factor model. Moreover, this upper bound is attained at $\xi^*_{D,\pi} = \xi^*_{D,\pi_1} \otimes \xi^*_{D,\pi_2}$ since following (15) we obtain

$$\int_\Lambda \log|I(\xi^*_{D,\pi_1} \otimes \xi^*_{D,\pi_2}, \lambda)|\, d\pi(\lambda) = \int_{\Lambda_1} \log|I_1(\xi^*_{D,\pi_1}, \lambda_1)|\, d\pi_1(\lambda_1) + \int_{\Lambda_2} \log|I_2(\xi^*_{D,\pi_2}, \lambda_2)|\, d\pi_2(\lambda_2)$$

So all inequalities turn into equalities and ξ^*_π is optimal.

(b) Let the set $\mathcal{N}(\xi)$ be the subset of Λ_M at which for a given design ξ the minimum of $\Phi(\xi, \lambda)$ over λ is attained, i.e.

$$\mathcal{N}(\xi) = \{\lambda \in \Lambda_M \,|\, \Phi(\xi, \lambda) = \min_{\lambda \in \Lambda_M} \Phi(\xi, \lambda)\}.$$

As we consider compact sets $\Lambda_{M,k}$, $k = 1, 2$, the product set Λ_M is also compact, and the infimum in the definition of the standardised maximin D-optimality criterion will be attained, so the minimum exists for all ξ. The set $\mathcal{N}(\xi)$ will therefore not be empty for any design ξ. Let Ψ_0 be the optimal value of the standardised maximin D-optimality criterion, i.e. $\Psi_0 = \min_{\lambda \in \Lambda_M} \Phi(\xi^*, \lambda)$. The proof will be divided into two steps, the first of which will show that the product design characterised by the marginals ξ^*_k, $k = 1, 2$, of the standardised maximin D-optimal design ξ^* for Λ_M is also standardised maximin D-optimal for Λ_M. In step 2, we will show that the product of the standardised maximin D-optimal designs $\xi^*_{D,\Lambda_{M,k}}$ with respect to $\Lambda_{M,k}$ in the single factor models will produce the same value as the product of the marginals ξ^*_k, $k = 1, 2$, when substituted into the criterion function $\Psi_{D,\Lambda_M}(\cdot)$.

Applying (15) and the result for locally D-optimal designs from part (a) of this Theorem, we obtain the following inequality for all $\lambda = (\lambda_1^T, \lambda_2^T)^T \in \Lambda_M$:

$$\Phi(\xi^*, \lambda) \;\leq\; \Phi(\xi_1^*, \lambda_1)\Phi(\xi_2^*, \lambda_2) = \Phi(\xi_1^* \otimes \xi_2^*, \lambda) \tag{17}$$

For all $\lambda \in \mathcal{N}(\xi_1^* \otimes \xi_2^*)$, we have

$$\Phi(\xi_1^* \otimes \xi_2^*, \lambda) \;\leq\; \Psi_0, \tag{18}$$

because of the optimality of Ψ_0. Combining (17) and (18), yields that for all $\lambda \in \mathcal{N}(\xi_1^* \otimes \xi_2^*)$

$$\Psi_0 \leq \Phi(\xi^*, \lambda) \leq \Phi(\xi_1^* \otimes \xi_2^*, \lambda) \leq \Psi_0,$$

hence

$$\min_{\lambda \in \Lambda_M} \Phi(\xi_1^* \otimes \xi_2^*, \lambda) = \Psi_0 = \max_{\xi} \min_{\lambda \in \Lambda_M} \Phi(\xi, \lambda),$$

so the product design $\xi_1^* \otimes \xi_2^*$ is also standardised maximin D-optimal with respect to Λ_M.

Let $\Psi_k = \min_{\lambda_k \in \Lambda_{M,k}} \Phi(\xi^*_{D,\Lambda_{M,k}}, \lambda_k)$, $k = 1, 2$. Applying (15) and the result for locally D-optimal designs from part (a) again, it follows that for all $\lambda \in \mathcal{N}(\xi^*_{D,\Lambda_{M,1}} \otimes \xi^*_{D,\Lambda_{M,2}})$

$$\Psi_1\, \Psi_2 \;\leq\; \Phi(\xi^*_{D,\Lambda_{M,1}}, \lambda_1)\Phi(\xi^*_{D,\Lambda_{M,2}}, \lambda_2) = \Phi(\xi^*_{D,\Lambda_{M,1}} \otimes \xi^*_{D,\Lambda_{M,2}}, \lambda) \leq \Psi_0. \tag{19}$$

Since neither of the values Ψ_0, Ψ_1 or Ψ_2 depend on λ, we obtain from (19) that $\Psi_1\, \Psi_2 \leq \Psi_0$. Now, for all $\lambda \in \mathcal{N}(\xi_1^*) \times \mathcal{N}(\xi_2^*)$ we have

$$\Phi(\xi_1^*, \lambda_1)\Phi(\xi_2^*, \lambda_2) \;=\; \Phi(\xi_1^* \otimes \xi_2^*, \lambda) \geq \Psi_0. \tag{20}$$

Combining (19) and (20), we find that

$$\min_{\lambda_1 \in \Lambda_{M,1}} \Phi(\xi_1^*, \lambda_1) \min_{\lambda_2 \in \Lambda_{M,2}} \Phi(\xi_2^*, \lambda_2) \;\geq\; \Psi_1\, \Psi_2, \tag{21}$$

so to avoid a contradiction to the optimality of $\xi^*_{D,\Lambda_{M,1}}$ and $\xi^*_{D,\Lambda_{M,2}}$ in the single factor models there must be equality in (21), and using (19) again we obtain that

$$\min_{\lambda \in \Lambda_M} \Phi(\xi^*_{D,\Lambda_{M,1}} \otimes \xi^*_{D,\Lambda_{M,2}}, \lambda) = \Psi_0,$$

which completes the proof of Theorem 1. □

A.2 Proof of Remark 1

From Corollary 5.4 in Schwabe (1996) we obtain that a necessary condition for local D-optimality of a design ξ in the additive model (2) is local D-optimality of the marginals of ξ in the corresponding single factor models (1) which proves the assertion. □

A.3 Proof of Lemma 1

The multiplication of a matrix B with A^T from the left and A from the right yields the top left corner B_{11} of B of size $s \times s$. We therefore show that the determinant of $M_{11}^-(\xi, \theta, \lambda)$ is proportional to the determinant of $I_{11}^-(\xi, \lambda)$ in what follows. We define $g_s(x, \theta, \lambda)$ as the gradient of the regression function with respect to λ and θ, i.e. the first s entries corresponding to the derivatives with respect to λ will be linear combinations (involving θ) of terms of the form $(x_k - \lambda_{k,i})_+^j$, $j = m_k - l_{ik}, \ldots, m_k$, where m_k and l_{ik} are defined in (1) and $m_k - l_{ik} + 1$ and m_k are the lowest and the highest exponent of $(x_k - \lambda_{k,i})_+$ appearing in the model, respectively. Similarly, define $f_s(x, \lambda)$ as the vector with the first s entries equal to $(x_k - \lambda_{k,i})_+^{m_k - l_{ik}}$, and the last $(p - s)$ entries equal to the last $(p - s)$ entries in $g_s(x, \theta, \lambda)$. We therefore have

$$M(\xi, \theta, \lambda) = \int g_s(x, \theta, \lambda) g_s^T(x, \theta, \lambda)\, d\xi(x) \quad \text{and} \quad I(\xi, \lambda) = \int f_s(x, \lambda) f_s^T(x, \lambda)\, d\xi(x).$$

We can express $g_s(x, \theta, \lambda)$ in the form $g_s(x, \theta, \lambda) = C_\theta f_s(x, \lambda)$ where C_θ is a non-singular $p \times p$-matrix with the lower $(p - s)$ rows equal to $(0_{(p-s)\times s} \mid J_{p-s})$, and $C_{\theta,11}$, the top left corner of C_θ of size $s \times s$, is a diagonal matrix where the entry $c_{k,i}$ corresponding to the knot $\lambda_{k,i}$ is given by $-(m_k - l_{ik} + 1)$ times the coefficient $\theta_{k,i,l_{ik}-1}$ of the term $(x_k - \lambda_{k,i})_+^{m_k - l_{ik}+1}$, i.e. C_θ is of the form

$$C_\theta \;=\; \left(\begin{array}{c|c} C_{\theta,11} & C_{\theta,12} \\ \hline 0_{(p-s)\times s} & J_{p-s} \end{array}\right) \;=\; \left(\begin{array}{c|c} \mathrm{diag}(c_{k,i}) & C_{\theta,12} \\ \hline 0_{(p-s)\times s} & J_{p-s} \end{array}\right).$$

Since both factors of $c_{k,i}$ are non-zero by the definition of the model, so is their product $-(m_k - l_{ik} + 1)\, \theta_{k,i,l_{ik}-1}$ and both the top left and the bottom right corner of C_θ are non-singular. By the formula for inverting block matrices, we obtain that C_θ^{-1} is a block matrix of the form

$$C_\theta^{-1} \;=\; \left(\begin{array}{c|c} C_{\theta,11}^{-1} & C_{\theta,11}^{-1} C_{\theta,12} \\ \hline 0_{(p-s)\times s} & J_{p-s} \end{array}\right) \;=\; \left(\begin{array}{c|c} \mathrm{diag}(1/c_{k,i}) & C_{\theta,11}^{-1} C_{\theta,12} \\ \hline 0_{(p-s)\times s} & J_{p-s} \end{array}\right). \tag{22}$$

Now $A^T (C_\theta^T)^{-1} = (\mathrm{diag}(1/c_{k,i}) \mid 0_{s\times(p-s)})$, so

$$|A^T M^-(\xi, \theta, \lambda) A| \;=\; |A^T (C_\theta^T)^{-1} I^-(\xi, \lambda) C_\theta^{-1} A| = \prod_{k,i} \left(\frac{1}{c_{k,i}}\right)^2 |A^T I^-(\xi, \lambda) A|.$$

Therefore the assertion of Lemma 1 follows with $c_\theta = \prod_{k,i} c_{k,i}^2$. □

A.4 Proof of Theorem 2

From the proof of Theorem 5.13 in Schwabe (1996), we obtain the inequality

$$\psi_s(I(\xi, \lambda)) \le \psi_s(I_1(\xi_1, \lambda_1))\ \psi_s(I_2(\xi_2, \lambda_2)) = \psi_s(I(\xi_1 \otimes \xi_2, \lambda))$$

where ξ_1 and ξ_2 are the marginals of the design ξ, and $\lambda = (\lambda_1^T, \lambda_2^T)^T$. The rest of the proof now follows exactly along the same lines as the proof of Theorem 1 and is therefore omitted. □

A.5 Proof of Theorem 3

We first establish a result on the lower Cholesky factor G for the matrix Q.

Lemma 2 *If $H(x)$ is a product measure of the marginals $H_1(x_1), \ldots, H_K(x_K)$ there exists a Cholesky decomposition GG^T of Q where the lower Cholesky factor G is block-diagonal except for the first column, which may consist of non-zero entries. The blocks are lower triangular matrices, and have sizes $\mathbb{R}^{p_k \times p_k}$, $k = 1, \ldots, K$, where $p_k + 1$ is the number of parameters in the k^{th} marginal model (1). If the choice of the distribution $H(x)$ makes Q positive definite the unique lower Cholesky factor G is of the form described above.*

Proof of Lemma 2: For clarity of presentation, we show the proof for the special case $K = 2$. The proof for general $K \geq 2$ then follows by defining meta-factors of more than one variable. Let $f(x_1, x_2, \lambda) = (1, \tilde{f}_1(x_1, \lambda_1), \tilde{f}_2(x_2, \lambda_2))^T$. Then the matrix Q is given by

$$
\begin{aligned}
Q &= \left(\begin{array}{c|c|c} 1 & Q_{10}^T & Q_{20}^T \\ \hline Q_{10} & Q_{11} & Q_{21}^T \\ \hline Q_{20} & Q_{21} & Q_{22} \end{array}\right), \quad \text{where} \quad Q_{k0} = \int \tilde{f}_k(x_k, \lambda_k)\, dH_k(x_k), \qquad (23) \\
Q_{kk} &= \int \tilde{f}_k(x_k, \lambda_k) \tilde{f}_k^T(x_k, \lambda_k)\, dH_k(x_k), \quad k = 1, 2, \quad \text{and} \\
Q_{21} &= \int \tilde{f}_2(x_2, \lambda_2) \tilde{f}_1^T(x_1, \lambda_1)\, dH(x_1, x_2) = \int \tilde{f}_2(x_2, \lambda_2)\, dH_2(x_2) \int \tilde{f}_1^T(x_1, \lambda_1)\, dH_1(x_1).
\end{aligned}
$$

The last equality in (23) follows from Fubini's Theorem and the assumption that $H(x)$ is a product measure. The lower triangular matrix G in the Cholesky decomposition of Q is of the form

$$
G = \left(\begin{array}{c|c|c} 1 & 0_{1\times p_1} & 0_{1\times p_2} \\ \hline G_{10} & G_{11} & 0_{p_1\times p_2} \\ \hline G_{20} & G_{21} & G_{22} \end{array}\right) \qquad (24)
$$

where the blocks G_{kk} are lower triangular matrices of size $p_k \times p_k$, respectively, $k = 1, 2$, and $(1, G_{10}, G_{20})^T$ is a column vector. To prove the assertion of Lemma 2, we show in the following that there exists a Cholesky decomposition G with the block $G_{21} = 0_{p_2\times p_1}$. We note that from (24) we obtain

$$
GG^T = \left(\begin{array}{c|c|c} 1 & G_{10}^T & G_{20}^T \\ \hline G_{10} & G_{11}G_{11}^T + G_{22}G_{22}^T & G_{10}G_{20}^T \\ \hline G_{20} & G_{20}G_{10}^T & G_{20}G_{20}^T + G_{22}G_{22}^T \end{array}\right) \qquad (25)
$$

for $G_{21} = 0_{p_2\times p_1}$. We are now looking for a matrix G such that GG^T of the form (25) is equal to Q. Equating the corresponding blocks in GG^T and Q and using that both matrices are

symmetric we obtain

$$\begin{aligned} G_{10} &= Q_{10}, \quad G_{20} = Q_{20} && (26)\\ (26),(23) &\Rightarrow G_{10}G_{10}^T + G_{11}G_{11}^T = Q_{11} \Rightarrow G_{11}G_{11}^T = Q_{11} - Q_{10}Q_{10}^T \\ &= \int \tilde{f}_1(x_1)\tilde{f}_1^T(x_1)\, dH_1(x_1) - \int \tilde{f}_1(x_1)\, dH_1(x_1) \int \tilde{f}_1^T(x_1)\, dH_1(x_1) && (27) \end{aligned}$$

where the expression in (27) is obviously non-negative definite, and therefore a Cholesky decomposition $G_{11}G_{11}^T$ exists. Similarly, we find that

$$G_{22}G_{22}^T = Q_{22} - G_{20}G_{20}^T = \int \tilde{f}_2(x_2)\tilde{f}_2^T(x_2)\, dH_2(x_2) - \int \tilde{f}_2(x_2)\, dH_2(x_2) \int \tilde{f}_2^T(x_2)\, dH_2(x_2)$$

which is also non-negative definite, so, again, there exists a Cholesky decomposition $G_{22}G_{22}^T$ for this matrix. Finally, we consider the term

$$G_{20}G_{10}^T = Q_{20}Q_{10}^T = \int \tilde{f}_2(x_2)\, dH_2(x_2)(\int \tilde{f}_1(x_1)\, dH_1(x_1))^T = Q_{21}$$

from (26) and (23). The last assertion is obvious because if Q is positive definite then the Cholesky decomposition is unique, and we have just shown that there exists a lower Cholesky factor G which satisfies the condition $G_{21} = 0_{p_2 \times p_1}$. □

From Lemma 2 we obtain that there exists a lower Cholesky factor G for Q, which is a block-diagonal matrix except for the first column. Applying the rules for block-wise matrix inversion (repeatedly if $K > 2$) we find that G^{-1} is of the same form as G. We note that the matrix $G^T I^{-1}(\xi, \lambda)G$ is at the same time (asymptotically) proportional to the covariance matrix of the maximum likelihood estimator for β in the linear regression model with expectation $\breve{\eta}(x) = \breve{f}^T(x, \lambda)\beta = (G^{-1}f(x, \lambda))^T\beta$ and i.i.d. normal errors. Therefore $\breve{f}(x, \lambda)$ is of the form $(1, \breve{f}_1(x_1, \lambda_1), \ldots, \breve{f}_K(x_K, \lambda_K))^T$, i.e. each single vector $\breve{f}_k(x_k, \lambda_k)$ depends only on the variable x_k, $k = 1, \ldots, K$. We can therefore apply Lemma 5.5 (ii) in Schwabe (1996) to the weighted regression model $\breve{\eta}(x) = (G^{-1}f(x, \lambda))^T\beta$, and we obtain the form of the covariance matrix $C_G(\xi, \lambda) = G^T I^{-1}(\xi, \lambda)G$ in the additive model if ξ is a product design, and of the covariance matrices $C_{G,k}(\xi_k, \lambda_k)$ in the corresponding marginal models $\breve{\eta}_k(x_k) = (1, \breve{f}_k(x_k, \lambda_k)^T)\beta_k$ where ξ_k are the marginals of ξ. For the marginal model we define the matrices Q_k with

$$\begin{aligned} Q_k &= \int \begin{pmatrix} 1 \\ \tilde{f}_k(x_k, \lambda_k) \end{pmatrix} (1 \quad \tilde{f}_k^T(x_k, \lambda_k))\, dH_k(x_k) \\ &= \left(\begin{array}{c|c} 1 & \int \tilde{f}_k^T(x_k, \lambda_k)\, dH_k(x_k) \\ \hline \int \tilde{f}_k(x_k, \lambda_k)\, dH_k(x_k) & \int \tilde{f}_k(x_k, \lambda_k) \int \tilde{f}_k^T(x_k, \lambda_k)\, dH_k(x_k) \end{array}\right) \\ &= \left(\begin{array}{c|c} 1 & Q_{k0}^T \\ \hline Q_{k0} & Q_{kk} \end{array}\right), \quad k = 1, \ldots, K, \end{aligned}$$

using the notation from the proof of Lemma 2. The lower Cholesky factor G_k for each Q_k is therefore given by

$$G_k = \left(\begin{array}{c|c} 1 & 0_{1\times p_k} \\ \hline G_{k0} & G_{kk} \end{array} \right), \quad k = 1, \ldots, K.$$

The block-wise inverse G_k^{-1} of G_k is

$$G_k^{-1} = \left(\begin{array}{c|c} 1 & 0_{1\times p_k} \\ \hline -G_{kk}^{-1}G_{k0} & G_{kk}^{-1} \end{array} \right).$$

Now G^{-1} has the blocks $1, G_{11}^{-1}, \ldots, G_{KK}^{-1}$ on its main diagonal, some possibly non-zero entries in the first column and all other entries are zero. The model $(G_k^{-1} f_k(x_k, \lambda_k))^T \beta_k$ is therefore equivalent to the marginal model $\breve{\eta}_k(x_k) = (1, \breve{f}_k(x_k, \lambda_k)^T)\beta_k$ of $\breve{\eta}(x)$, $k = 1, \ldots, K$. We write $C_{G,k}(\xi_k, \lambda_k)$ in diagonal form

$$C_{G,k}(\xi_k, \lambda_k) = \left(\begin{array}{c|c} C_{k,\beta_0} & \gamma^T \\ \hline \gamma & \tilde{C}_{G,k}(\xi_k, \lambda_k) \end{array} \right),$$

where γ is some vector, and use the representation of the covariance matrix $C_G(\xi, \lambda)$ of a product design $\xi = \xi_1 \otimes \ldots \otimes \xi_K$ from Lemma 5.5 (ii) in Schwabe (1996)

$$C_G(\xi, \lambda) = \left(\begin{array}{c|c|c|c} C_{\beta_0} & & & \\ \hline & \tilde{C}_{G,1}(\xi_1, \lambda_1) & & \\ \hline & & \ldots & \\ \hline & & & C_{G,K}(\xi_K, \lambda_K) \end{array} \right),$$

where the off-diagonal blocks in $C_G(\xi, \lambda)$ have been omitted since they do not contribute to the trace and $C_{\beta_0} = \sum_{k=1}^K C_{k,\beta_0} - (K-1)$. From this representation it is obvious that the local Q-criterion for ξ in the additive model with respect to $H(x)$ and λ can be expressed (apart from an additive constant) as the sum of the local Q-criteria for ξ_k in the single factor models with respect to H_k and λ_k. So interchanging the integration with respect to $\pi(\lambda)$ and the summation of the Q-criteria in the marginal models yields the desired result for Bayesian Q-optimality. Equivalently, the maximisation with respect to $\lambda = (\lambda_1^T, \ldots, \lambda_K^T)^T \in \Lambda_{M,1} \times \ldots \times \Lambda_{M,K}$ and the summation can be interchanged, so the result for minimax Q-optimality follows. □

A.6 Proof of Theorem 4:

Let $b_i(x, \lambda)$, $i = 1, \ldots, r$, be the truncated power basis for the spline space R with given knots λ and given smoothnesses at the knots where x can be multi-dimensional, i.e. $x = (x_1, \ldots, x_K)$. Then for any system of splines (e.g. B-splines) there exists a basis $\breve{b}_i(x, \lambda)$, $i = 1, \ldots, r$, for the same space R. So the original regression function $\mu(x)$ (an element of R) can be written as

$$\mu(x) \;=\; \sum_{i=1}^r \theta_i b_i(x, \lambda) = \sum_{i=1}^r \breve{\theta}_i \breve{b}_i(x, \lambda) \tag{28}$$

for some parameters θ_i and $\check{\theta}_i$, $i = 1, \ldots, r$. From (4) and (5), we obtain that the vector of derivatives of $\sum_{i=1}^{r} \theta_i b_i(x, \lambda)$ with respect to the parameters θ_i and λ (when multiplied by some non-singular matrix C_θ^{-1}, which depends on θ but neither on the knots λ nor on the design) forms the basis of another spline space S with the same knots as for R, the original one, but smaller smoothnesses. This basis depends on λ and x, but not on θ. Since each element of R can also be expressed in terms of $\check{b}_i(x, \lambda)$, there exists another non-singular matrix $D_{\check{\theta}}$, which may depend on $\check{\theta}$ and the knots but not on x, so that the vector of derivatives multiplied by $D_{\check{\theta}}^{-1}$ is a basis of S, which only depends on λ and x. Now the two new bases for S are related to each other by multiplication with a non-singular matrix $E_{\theta,\lambda}$, which does not depend on x (basis transformation). As a result, the information matrices $I(\xi, \lambda)$, $\check{I}(\xi, \lambda)$ for the models generated by using the different spline bases are related by $I(\xi, \lambda) = E_{\theta,\lambda}\, \check{I}(\xi, \lambda)\, E_{\theta,\lambda}^T$.
For the D- and Q-optimality criteria we obtain:

- Bayesian D-optimality: $\int \log |I(\xi, \lambda)|\, d\pi(\lambda) = \int \log |E_{\theta,\lambda}|^2\, d\pi(\lambda) + \int \log |\check{I}(\xi, \lambda)|\, d\pi(\lambda)$. Since $E_{\theta,\lambda}$ does not depend on x and therefore not on the design ξ, the criterion functions for both matrices are maximised by the same design.

- Standardised Maximin optimality: Using that the locally D-optimal designs are equal, and that the expression $|E_{\theta,\lambda}|^2$ cancels, we find that

$$\min_{\lambda \in \Lambda_M} \frac{|I(\xi, \lambda)|}{|I(\xi_{D,\lambda}^*, \lambda)|} = \min_{\lambda \in \Lambda_M} \frac{|E_{\theta,\lambda}|^2\, |\check{I}(\xi, \lambda)|}{|E_{\theta,\lambda}|^2\, |\check{I}(\xi_{D,\lambda}^*, \lambda)|},$$

  so the criterion functions are equal and therefore maximised by the same design.

- Q-optimality: From $f(x, \lambda) = E_{\theta,\lambda} \check{f}(x, \lambda)$ it follows that

$$\begin{aligned} Q(\xi, \lambda) &= \int f^T(x, \lambda) I^{-1}(\xi, \lambda) f(x, \lambda)\, dH(x) \\ &= \int \check{f}^T(x, \lambda) E_{\theta,\lambda}^T (E_{\theta,\lambda}^T)^{-1}(x, \lambda) \check{I}^{-1}(\xi, \lambda) E_{\theta,\lambda}^{-1} E_{\theta,\lambda} \check{f}(x, \lambda)\, dH(x) \\ &= \int \check{f}^T(x, \lambda) \check{I}^{-1}(\xi, \lambda) \check{f}(x, \lambda)\, dH(x) \end{aligned}$$

  So the Q-criterion function is equal for both bases, and therefore minimised by the same design. The assertion for Bayesian and minimax Q-optimality is now obvious. □

References

Chaloner, K., Verdinelli, I. (1995). Bayesian experimental design: A review. Statistical Science, 10, 273-304.

Chen, V. C. P., Ruppert, D. and Shoemaker, C. A. (1999). Applying experimental design and regression splines to high dimensional continues state stochastic dynamic programming, Operations Research, 47(1), 38-53.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. Ann. Math. Statist. 24, 586-602.

De Boor, C. (1978). A practical guide to splines. Applied Mathematical Sciences 27. Springer, New York.

Dette, H., Melas, V.B. and Pepelyshev, A. (2009). Optimal design for smoothing splines. To appear in Ann. Inst. Stat. Math.

Dette, H., Melas, V.B. and Pepelyshev, A. (2008). Optimal designs for free knot least squares splines. Statistica Sinica, 18, 1047-1062.

Dierckx, P. (1995). Curve and surface fitting with splines. Monographs on Numerical Analysis. Clarendon Press, Oxford.

Eubank, R.L. (1999). Nonparametric regression and spline smoothing. 2nd ed. Statistics: Textbooks and Monographs 157. Marcel Dekker, New York.

Fang, K-T., Li, R. and Sudjianto, A. (2006). Design and Modeling for Computer Experiments. Chapman and Hall. (p. 139-145)

Friedman, J.H. (1991). Multivariate Adaptive Regression Splines (with discussion). Annals of Statistics 19, 1-141.

Gallant, A.R. and Fuller, W.A. (1973). Fitting segmented polynomial regression models whose join points have to be estimated. Journal of the American Statistical Association 68, 144-147.

Grove, D.M., Woods, D.C. and Lewis, S.M. (2004). Multifactor B-spline mixed models in designed experiments for the engine mapping problem. Journal of Quality Technology, 36, No.4, 380-391.

Hartley, H.O. (1961). The modified Gauss-Newton method for the fitting of nonlinear regression functions by least squares. Technometrics 3, 269-280.

Heiligers, B. (1999). Experimental design for polynomial spline regression. Tatra Mt. Math. Publ. 17, 157-165.

Heiligers, B. (1998). E-optimal designs for polynomial spline regression. J. Stat. Plann. Inference 75, 159-172.

Hill, Peter D.H. (1980). D-optimal designs for partially nonlinear regression models. Technometrics 22, No. 2, 275-276.

Imhof, L. (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. Annals of Statistics, 29, No. 2, 561-576.

Jupp, D.L.B. (1978). Approximation to data by splines with free knots. SIAM J. Numer. Anal. 15, 328-343.

Kaishev, V.K. (1989). Optimal experimental designs for the B-spline regression. Comput. Stat. Data Anal. 8, 39-47.

Khuri, A.I. (1984). A note on D-optimal designs for partially nonlinear regression models. Technometrics 26, No. 1, 59-61.

Mao, W. and Zhao, L.H. (2003). Free-knot polynomial splines with confidence intervals. Journal of the Royal Statistical Society, Series B, 65, 901-919.

Murty, V.N. (1971a). Optimal designs with a Tchebycheffian spline regression function. Ann. Math. Stat. 42, 643-649.

Murty, V.N. (1971b). Optimal designs with a polynomial spline regression with a single multiple knot at the center. Ann. Math. Stat. 42, 952-960.

Park, S.H. (1978). Experimental designs for fitting segmented polynomial regression models. Technometrics 20, 151-154.

Put, R., Xu, Q.S., Massart, D.L. and Vander Heyden, Y. (2004). Multivariate adaptive regression splines (MARS) in chromatographic quantitative structureretention relationship studies. Journal of Chromatography A, 1055, 11-19.

Schwabe, R. (1996). Optimum designs for multi-factor models. Lecture Notes in Statistics, Springer, New York.

Siddappa, S., Günther, D., Rosenberger, J.M. and Chen, V.C.P. (2007). Refined experimental design and regression splines method for network revenue management. Journal of Revenue and Pricing Management, 6(3), 188-199.

Stone, C.J., Hansen, M.H., Kooperberg, C. and Truong, Y.K. (1997). Polynomial splines and their tensor products in extended linear modeling: 1994 Wald memorial lecture. The Annals of Statistics 25, No. 4, 1371-1470.

Studden, W.J. (1971). Optimal designs and spline regression. Optimizing Meth. Statist., Proc. Sympos. Ohio State Univ.

Studden, W.J. and VanArman, D.J. (1969). Admissible designs for polynomial spline regression. Ann. Math. Stat. 40, 1557-1569.

Yue, R. and Hickernell, F. (2002). Robust designs for smoothing spline ANOVA models. Metrika 55, 161-176

Woods, D.C., Lewis, S.M. and Dewynne, J.N. (2003). Designing experiments for multi-variable B-spline models. Sankhya, 65, 660-677.