Optimal Design for Multivariate Observations in Seemingly Unrelated Linear Models

Moudar Soumaya[1*], Norbert Gaffke[2], Rainer Schwabe[2]

[1] Process Systems Engineering, Otto-von-Guericke University Magdeburg, PF 4120, 39016 Magdeburg, Germany
[2] Institute for Mathematical Stochastics, Otto-von-Guericke University Magdeburg, PF 4120, 39016 Magdeburg, Germany

Abstract The concept of seemingly unrelated models is used for multivariate observations when the components of the multivariate dependent variable are governed by mutually different sets of explanatory variables and the only relation between the components is given by a fixed covariance within the observational units. A multivariate weighted least squares estimator is employed which takes the within units covariance matrix into account. In an experimental setup, where the settings of the explanatory variables may be chosen freely by an experimenter, it might be thus tempting to choose the same settings for all components to end up with a multivariate regression model, in which the ordinary and the least squares estimators coincide. However, we will show that under quite natural conditions the optimal choice of the settings will be a product type design which is generated from the optimal counterparts in the univariate models of the single components. This result holds even when the univariate models may change form component to component.

Keywords: multivariate linear model, seemingly unrelated regression, optimal design, product type design.

1 Introduction

In an experiment often more than one dependent variable is observed for each observational unit. Sometimes for these dependent components the explanatory variables may be adjusted separately. For example, one might be interested in some processes over time like pharmacokinetics and pharmacodynamics, where observations can be made at the same subjects, but where the time points need not be identical for the measurements of the different quantities within one subject. As typically observations are correlated within units, the data are properly described by a multivariate model with separate sets of explanatory variables.

Such models have been introduced by Zellner (1962) in econometrics and have been called seemingly unrelated regression (SUR) models, because the corresponding univariate models for the components do not seem to have anything in common at a first glance. However, it has been pointed out that the correlation between the variables could be employed to transfer useful information from one component to another. Since its introduction various modifications have been considered in observational studies, and the corresponding statistical analysis has been well developed.

In experimental situations these seemingly unrelated regression models have been used less frequently, and, to the best knowledge of the authors, no explicit result is available for the construction of optimal designs in such experiments. There is a certain belief that it is sufficient to choose optimal marginal designs for the single components. However, we will show in this paper that additionally the joint distribution of the marginal designs play an important role, because information may be transferred between the components. In particular, we will establish that product type designs are optimal if all univariate models related to the components contain a

*Corresponding author: moudar.soumaya@gmail.com

constant term (intercept or grand mean). The optimal designs for the multivariate model will be constructed as products of there corresponding counterparts which are optimal in the univariate models of the components. We will also show that these product type designs outperform competitors which do not share a full factorial structure.

Proofs will be based on Fedorov's (1972) multivariate versions of the equivalence theorems for optimal designs. Some techniques concerning product type designs are adopted from Schwabe (1996), and there seems to be a relation between univariate additive models and the seemingly unrelated linear models treated here. However, in contrast to the univariate additive case the optimality of product type designs is not restricted to the commonly used D-criterion, but carries over to linear criteria and, in particular, to the A- and $IMSE$-criterion.

The paper is organized as follows: In the second section we specify the model and collect some relevant issues of optimal designs in the third section. In section 4 we present the optimality of product type designs and illustrate their performance by an example in the bivariate case in section 5. Section 6 contains some discussion of the results. Technical proofs are deferred to an appendix.

2 Model specification

We consider multivariate linear models in which m-dimensional observations $\mathbf{Y}_i$ depend on some explanatory variables for n experimental units $i = 1, ..., n$. The components (variables) Y_{ij} of the multivariate observations $\mathbf{Y}_i = (Y_{i1}, ..., Y_{im})^\top$ are assumed to be seemingly unrelated. This means that the settings x_{ij} of the explanatory variables may differ across the components. More generally, we even allow for different univariate linear models for the components, i. e. different explanatory variables, different regression functions and different experimental regions. This model approach covers and generalizes the concept of seemingly unrelated regression (SUR) by Zellner (1962) and may also contain components of the analysis of variance type or with both qualitative and quantitative factors of influence.

For each component j the observation Y_{ij} of unit i is specified by a linear model

$$Y_{ij} = \sum_{\ell=1}^{p_j} f_{j\ell}(x_{ij})\beta_{j\ell} + \varepsilon_{ij} = \mathbf{f}_j(x_{ij})^\top \boldsymbol{\beta}_j + \varepsilon_{ij} \,, \tag{1}$$

where $\mathbf{f}_j = (f_{j1}, ..., f_{jp_j})^\top$ are known regression functions of the experimental setting x_{ij}, $\boldsymbol{\beta}_j = (\beta_{j1}, ..., \beta_{jp_j})^\top$ are unknown parameters and p_j is the dimension for the jth component. The experimental setting x_{ij} may be chosen from an experimental region $\mathcal{X}_j$.

The combined observational vector $\mathbf{Y}_i$ can then be written as a multivariate linear model

$$\mathbf{Y}_i = \mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \,, \tag{2}$$

where $\mathbf{f}$ is the block diagonal multivariate regression function

$$\mathbf{f}(\mathbf{x}) = \operatorname{diag}\left(\mathbf{f}_j(x_j)\right)_{j=1,\ldots,m} = \begin{pmatrix} \mathbf{f}_1(x_1) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{f}_m(x_m) \end{pmatrix} \tag{3}$$

for the multivariate experimental setting $\mathbf{x} = (x_1, ..., x_m)$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, ..., \boldsymbol{\beta}_m^\top)^\top$ is the stacked parameter vector of dimension $p = \sum_{j=1}^m p_j$ for all components and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{im})^\top$ the multivariate vector error terms for unit i. We assume additionally that the components x_{ij} of the multivariate experimental setting $\mathbf{x}$ may be chosen independently for each component resulting in a rectangular form of the experimental region, $\mathbf{x} \in \mathcal{X} = \times_{j=1}^m \mathcal{X}_j$. To assure estimability it is further assumed that the components of $\mathbf{f}_j$ are linearly independent functions on $\mathcal{X}_j$ for each $j = 1, ..., m$, which implies that the components of $\mathbf{f}$ are linearly independent functions on $\mathcal{X}$.

To complete the model the $\boldsymbol{\varepsilon}_i$ are assumed to be zero mean error vectors with homogeneous non-singular covariance matrix $\boldsymbol{\Sigma} = \mathrm{Cov}\,(\boldsymbol{\varepsilon}_i)$ which are uncorrelated across the units. Hence, the observational vectors $\mathbf{Y}_i$ inherit the covariance structure from the error terms, $\mathrm{Cov}\,(\mathbf{Y}_i) = \boldsymbol{\Sigma}$, and the covariance structure does not depend on the experimental settings $\mathbf{x}_i$.

For the whole experiment denote by $\mathbf{Y} = (\mathbf{Y}_1^\top, ..., \mathbf{Y}_n^\top)^\top$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, ..., \boldsymbol{\varepsilon}_n^\top)^\top$ the stacked vectors of all observations and all error terms, respectively. Then we can write the complete vector of observations as

$$\mathbf{Y} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \tag{4}$$

where $\mathbf{F} = (\mathbf{f}(\mathbf{x}_1), ..., \mathbf{f}(\mathbf{x}_n))^\top$ is the stacked design matrix of dimension $(nm) \times p$.

The complete observational error $\boldsymbol{\varepsilon}$ and, hence, $\mathbf{Y}$ has covariance matrix $\mathbf{V} = \mathrm{Cov}(\boldsymbol{\varepsilon}) = \mathbf{I}_n \otimes \boldsymbol{\Sigma}$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix and "$\otimes$" denotes the Kronecker product.

If the covariance matrix $\boldsymbol{\Sigma}$ and, hence, $\mathbf{V}$ is known, and if the matrix $\mathbf{F}$ has rank p, then we can estimate the parameter vector $\boldsymbol{\beta}$ by the weighted least squares estimator

$$\hat{\boldsymbol{\beta}}_{\mathrm{WLS}} = (\mathbf{F}^\top \mathbf{V}^{-1} \mathbf{F})^{-1} \mathbf{F}^\top \mathbf{V}^{-1} \mathbf{Y}\,. \tag{5}$$

Note that under the above moment assumptions this estimator is best linear unbiased, and it becomes best unbiased under the additional assumption of Gaussian errors which is not imposed here.

The covariance matrix of the weighted least squares estimator equals the inverse of the corresponding information matrix

$$\mathbf{M} = \mathbf{F}^\top \mathbf{V}^{-1} \mathbf{F} = \sum_{i=1}^{n} \mathbf{f}(\mathbf{x}_i) \boldsymbol{\Sigma}^{-1} \mathbf{f}(\mathbf{x}_i)^\top \tag{6}$$

which is equal to the sum of the information $\mathbf{M}(\mathbf{x}_i) = \mathbf{f}(\mathbf{x}_i)\boldsymbol{\Sigma}^{-1}\mathbf{f}(\mathbf{x}_i)^\top$ for the single units at their corresponding experimental settings $\mathbf{x}_1, ..., \mathbf{x}_n$.

By (3) the information $\mathbf{M}(\mathbf{x}) = \mathbf{f}(\mathbf{x})\boldsymbol{\Sigma}^{-1}\mathbf{f}(\mathbf{x})^\top$ at $\mathbf{x} = (x_1, ..., x_m)$ can be represented as

$$\mathbf{M}(\mathbf{x}) = \left(\sigma^{(jh)} \mathbf{f}_j(x_j) \mathbf{f}_h(x_h)^\top \right)_{j=1,\dots,m}^{h=1,\dots,m}, \tag{7}$$

where $\sigma^{(jh)}$ is the entry in the jth row and hth column of the inverse $\boldsymbol{\Sigma}^{-1}$ of the covariance matrix $\boldsymbol{\Sigma}$ for a single unit.

Thus, for the whole experiment we obtain a block representation of the information matrix

$$\mathbf{M} = \left(\sigma^{(jh)} \sum_{i=1}^{n} \mathbf{f}_j(x_{ij}) \mathbf{f}_h(x_{ih})^\top \right)_{j=1,\dots,m}^{h=1,\dots,m}. \tag{8}$$

As we will relate the multivariate model to its univariate counterparts for the single components, we will have a closer look at them. For the jth component the univariate marginal model has the form

$$\mathbf{Y}^{(j)} = \mathbf{F}^{(j)} \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}^{(j)}\,, \tag{9}$$

where $\mathbf{Y}^{(j)} = (Y_{1j}, ..., Y_{nj})^\top$ and $\boldsymbol{\varepsilon}^{(j)} = (\varepsilon_{1j}, ..., \varepsilon_{nj})^\top$ are the vectors of observations and errors, respectively, and $\mathbf{F}^{(j)} = (\mathbf{f}_j(x_{1j}), ..., \mathbf{f}_j(x_{nj}))^\top$ is the design matrix in the univariate model for the jth component. The error terms are uncorrelated and homoscedastic, $\mathrm{Cov}\,(\boldsymbol{\varepsilon}^{(j)}) = \sigma_j^2 \mathbf{I}_n$, where $\sigma_j^2 = \sigma_{jj}$ is the jth diagonal entry of $\boldsymbol{\Sigma}$ resulting in a standard univariate linear model.

Note that the diagonal blocks in the representation (8) of the multivariate information matrix are equal to the information matrix ${\mathbf{F}^{(j)}}^\top \mathbf{F}^{(j)} = \sum_{i=1}^{n} \mathbf{f}_j(x_{ij}) \mathbf{f}_j(x_{ij})^\top$ in the corresponding univariate model multiplied by the factor $\sigma^{(jj)}$.

3 Optimal design

The design of an experiment consists of the choice of the experimental settings $\mathbf{x}_1, ..., \mathbf{x}_n$, and its quality will be measured in terms of the information matrrix $\mathbf{M} = \sum_{i=1}^n \mathbf{f}(\mathbf{x}_i)\boldsymbol{\Sigma}^{-1}\mathbf{f}(\mathbf{x}_i)^\top$. As the information matrix only depends on the values of the settings and does not depend on their order, an experimental design

$$\xi = \begin{pmatrix} \mathbf{x}_1 & \ldots & \mathbf{x}_k \\ w_1 & \ldots & w_k \end{pmatrix} \tag{10}$$

can be characterized by the set of all different experimental settings $\mathbf{x}_i = (x_{i1}, \ldots, x_{im}) \in \mathcal{X}$, $i = 1, ..., k$, with corresponding relative frequencies $w_i = \frac{n_i}{n}$, where n_i is the number of replications at $\mathbf{x}_i$ in the experiment. With this notation the corresponding standardized information matrix can be written as

$$\mathbf{M}(\xi) = \sum_{i=1}^k w_i \mathbf{f}(\mathbf{x}_i)\boldsymbol{\Sigma}^{-1}\mathbf{f}(\mathbf{x}_i)^\top = \left(\sigma^{(jh)} \sum_{i=1}^k w_i \mathbf{f}_j(x_{ij})\mathbf{f}_h(x_{ih})^\top\right)_{j=1,\ldots,m}^{h=1,\ldots,m} \tag{11}$$

which equals the information matrix (6) divided by the sample size n.

For analytical ease we consider approximate designs in the spirit of Kiefer (1974) for which the weights $w_i \geq 0$ need not be multiples of $\frac{1}{n}$, but only have to satisfy $w_i \geq 0$ and $\sum_{i=1}^k w_i = 1$. As information matrices are not necessarily comparable, we have to consider some real-valued criterion function of the information matrix. Here we will make use of the most popular criterion $-\ln\det(\mathbf{M}(\xi))$ of D-optimality and the general form of the linear criterion $\mathrm{trace}(\mathbf{L}\mathbf{M}(\xi)^{-1})$ of L-optimality, where $\mathbf{L}$ is a positive definite weight matrix. Note that each criterion aims at an optimal design minimizing that criterion over all competing designs. Note further that A-optimality and *IMSE*-optimality are special cases of L-optimality.

The *IMSE*-criterion averages the predictive variance of the design over the design region $\mathcal{X}$ with respect to the uniform measure μ on $\mathcal{X}$ (discrete or continuous or a product thereof). The multivariate version of the predictive variance at $\mathbf{x} = (x_1, ..., x_m)$ is defined as

$$\mathrm{trace}(\mathrm{Cov}(\mathbf{f}(\mathbf{x})^\top\hat{\boldsymbol{\beta}})) = \mathrm{trace}(\mathbf{f}(\mathbf{x})^\top\mathbf{M}^{-1}\mathbf{f}(\mathbf{x})) = \tfrac{1}{n}\mathrm{trace}(\mathbf{f}(\mathbf{x})^\top\mathbf{M}(\xi)^{-1}\mathbf{f}(\mathbf{x}))\,. \tag{12}$$

Then the *IMSE*-criterion is given by

$$\int_{\mathcal{X}} \mathrm{trace}(\mathbf{f}(\mathbf{x})^\top\mathbf{M}(\xi)^{-1}\mathbf{f}(\mathbf{x}))\mu(\mathrm{d}\mathbf{x}) = \mathrm{trace}(\mathbf{L}\mathbf{M}(\xi)^{-1})\,,\ \text{where } \mathbf{L} = \int_{\mathcal{X}} \mathbf{f}(\mathbf{x})\mathbf{f}(\mathbf{x})^\top\mu(\mathrm{d}\mathbf{x})\,.$$

Note that the uniform measure $\mu = \otimes_{j=1}^m \mu_j$ on $\mathcal{X} = \times_{j=1}^m \mathcal{X}_j$ factorizes into the uniform measures μ_j on $\mathcal{X}_j$, where "$\otimes$" denotes here the measure theoretic product. Thus the weight matrix $\mathbf{L} = \mathrm{diag}(\mathbf{L}_j)_{j=1,\ldots,m}$ is block diagonal with diagonal blocks $\mathbf{L}_j = \int_{\mathcal{X}_j} \mathbf{f}_j(x_j)\mathbf{f}_j(x_j)^\top\mu_j(\mathrm{d}x_j)$.

Next we present Fedorov's (1972, Theorems 5.2.1 and 5.3.1) multivariate versions of the equivalence theorems as useful tools for checking the optimality of a given candidate design.

Theorem 3.1 *A design ξ^* with nonsingular information matrix $\mathbf{M}(\xi^*)$ is D-optimal in the multivariate linear model if and only if*

$$\mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{f}(\mathbf{x})^\top\mathbf{M}(\xi^*)^{-1}\mathbf{f}(\mathbf{x})) \leq p \qquad \text{for all } x \in \mathcal{X}. \tag{13}$$

Theorem 3.2 *A design ξ^* with nonsingular information matrix $\mathbf{M}(\xi^*)$ is L-optimal with respect to $\mathbf{L}$ in the multivariate linear model if and only if*

$$\mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{f}(\mathbf{x})^\top\mathbf{M}(\xi^*)^{-1}\mathbf{L}\ \mathbf{M}(\xi^*)^{-1}\mathbf{f}(\mathbf{x})) \leq \mathrm{trace}(\mathbf{L}\mathbf{M}(\xi^*)^{-1}) \qquad \text{for all } x \in \mathcal{X}. \tag{14}$$

The function $\varphi_D(\mathbf{x};\xi) = \mathrm{trace}(\mathbf{\Sigma}^{-1}\mathbf{f}(\mathbf{x})^\top\mathbf{M}(\xi)^{-1}\mathbf{f}(\mathbf{x}))$ on the left hand side of the inequality (13) is called the sensitivity function of the design ξ with respect to the D-criterion. Similarly, $\varphi_L(\mathbf{x};\xi) = \mathrm{trace}(\mathbf{\Sigma}^{-1}\mathbf{f}(\mathbf{x})^\top\mathbf{M}(\xi)^{-1}\mathbf{L}\mathbf{M}(\xi)^{-1}\mathbf{f}(\mathbf{x}))$ is the sensitivity function in (14) for the L-criterion. These functions are essentially the directional derivatives of the corresponding criterion functions and indicate which experimental settings are "most informative".

4 Optimality of product type design

Product type designs $\xi = \otimes_{j=1}^m \xi_j$ are defined on $\mathcal{X} = \times_{j=1}^m \mathcal{X}_j$ as the measure theoretic product of their univariate marginals ξ_j on $\mathcal{X}_j$. To be more specific, let

$$\xi_j = \begin{pmatrix} x_{j1} & \ldots & x_{jk_j} \\ w_{j1} & \ldots & w_{jk_j} \end{pmatrix}$$

be a design on $\mathcal{X}_j$ with k_j settings x_{ji} and corresponding weights w_{ji}, $i = 1, ..., k_j$, respectively. Then the product type design $\xi = \otimes_{j=1}^m \xi_j$ has $k = \prod_{j=1}^m k_j$ settings $\mathbf{x}_{i_1,...,i_m} = (x_{1i_1}, ..., x_{mi_m})$ with corresponding weights $w_{i_1,...,i_m} = \prod_{j=1}^m w_{ji_j}$, $i_j = 1, ..., k_j$, $j = 1, ..., m$.

For product type designs $\xi = \otimes_{j=1}^m \xi_j$ the representation (11) of the information matrix $\mathbf{M}(\xi)$ simplifies, because the off-diagonal blocks factorize and the diagonal blocks can be directly related to the information matrices of the marginal designs in the univariate models.

Lemma 4.1 *Let $\xi = \otimes_{j=1}^m \xi_j$ be a product type design on the design region $\mathcal{X} = \times_{j=1}^m \mathcal{X}_j$ and denote by $\mathbf{M}_j(\xi_j) = \int_{\mathcal{X}_j} \mathbf{f}_j(x_j)\mathbf{f}_j(x_j)^\top \xi_j(\mathrm{d}x_j)$ the information matrix and by $\mathbf{m}_j(\xi_j) = \int_{\mathcal{X}_j} \mathbf{f}_j(x_j)\xi_j(\mathrm{d}x_j)$ the moment vector of the marginal design ξ_j. Then the information matrix for the seemingly unrelated linear model (4) has the form*

$$\mathbf{M}(\xi) = \mathrm{diag}\left(\sigma^{(jj)}(\mathbf{M}_j(\xi_j) - \mathbf{m}_j(\xi_j)\mathbf{m}_j(\xi_j)^\top)\right) + \mathbf{m}(\xi)\mathbf{\Sigma}^{-1}\mathbf{m}(\xi)^\top \,, \tag{15}$$

where $\mathbf{m}(\xi) = \int_{\mathcal{X}} \mathbf{f}(\mathbf{x})\xi(\mathrm{d}\mathbf{x}) = \mathrm{diag}(\mathbf{m}_j(\xi_j))_{j=1,...,m}$ is the block diagonal matrix of the marginal moments.

Proof. By the product structure of ξ the blocks of $\mathbf{M}(\xi)$ in (11) are equal to $\sigma^{(jj)}\mathbf{M}_j(\xi_j)$ if $j = h$, and equal to $\sigma^{(jh)}\mathbf{m}_j(\xi_j)\mathbf{m}_h(\xi_h)^\top$ if $j \neq h$. Now (15) follows, observing a block representation

$$\mathbf{m}(\xi)\mathbf{\Sigma}^{-1}\mathbf{m}(\xi)^\top = \left(\sigma^{(jh)}\mathbf{m}_j(\xi_j)\mathbf{m}_h(\xi_h)^\top\right)_{j=1,...,m}^{h=1,...,m}$$

of the last term. □

To establish the optimality of product type designs $\xi^* = \otimes_{j=1}^m \xi_j^*$, which are products of the optimal designs ξ_j^* in the corresponding univariate models for the components, we have to require the quite natural additional assumption that all univariate models contain a constant term. This means that for every component j there is a vector $\mathbf{c}_j$ of dimension p_j such that $\mathbf{c}^\top\mathbf{f}_j(x_j) = 1$ for all $x_j \in \mathcal{X}_j$. If there is an explicit constant, $f_{j1}(x_j) \equiv 1$ say, like in regression models, then $\mathbf{c}_j = (1, 0, ..., 0)^\top$. Another example is an ANOVA model under means parameterization which contains a constant term with $\mathbf{c}_j = (1/p_j, ..., 1/p_j)^\top$.

Under this model assumption the inverse of the information matrix $\mathbf{M}(\xi)$ of a product type design $\xi = \otimes_{j=1}^m \xi_j$ can be expressed explicitly as given in the next lemma.

Lemma 4.2 *Assume that each univariate model contains a constant term, ($\mathbf{c}_j^\top\mathbf{f}_j = 1$ for all $j = 1, ..., m$). Let ξ be a product type design, $\xi = \otimes_{j=1}^m \xi_j$, such that for each j the information* *matrix* $\mathbf{M}_j(\xi_j)$ *of* ξ_j *in the* j*th marginal model is nonsingular. Then the information matrix* $\mathbf{M}(\xi)$ *of* ξ *in the multivariate model is nonsingular and*

$$\mathbf{M}^{-1}(\xi) = \mathrm{diag}\left(\frac{1}{\sigma^{(jj)}}(\mathbf{M}_j(\xi_j)^{-1} - \mathbf{c}_j\mathbf{c}_j^\top)\right) + \mathbf{c}\boldsymbol{\Sigma}\mathbf{c}^\top\,, \tag{16}$$

where $\mathbf{c} = \mathrm{diag}(\mathbf{c}_j)_{j=1,...,m}$ *is the block diagonal matrix of the constant vectors* $\mathbf{c}_j$*.*

The proof is given in the Appendix. With this representation of the inverse of the information matrix we can establish the optimality of product type designs in seemingly unrelated linear models, in which every component contains a constant term.

Theorem 4.1 *Assume that each univariate model contains a constant term. For each* $j = 1, ..., m$ *let* ξ_j^* *be a D-optimal design for the* j*th univariate model on the design region* $\mathcal{X}_j$*. Then the product type design* $\xi^* = \otimes_{j=1}^m \xi_j^*$ *is a D-optimal design for the multivariate model on the design region* $\mathcal{X} = \times_{j=1}^m \mathcal{X}_j$*.*

Proof. By Lemma A.1 in the appendix the sensitivity function $\varphi_D(\mathbf{x}; \xi^*)$ is the sum of the marginal sensitivity functions $\varphi_{D,j}(x_j; \xi_j^*)$. Because of the D-optimality of the marginal designs ξ_j^* in the univariate models the univariate version of the Equivalence Theorem 3.1 establishes that for each component the marginal sensitivity function is bounded by the dimension of the univariate model, $\varphi_{D,j}(x_j; \xi_j^*) \le p_j$ for all $x_j \in \mathcal{X}_j$. Combining these results we obtain

$$\varphi_D(\mathbf{x}; \xi^*) = \sum_{j=1}^m \varphi_{D,j}(x_j; \xi_j^*) \le \sum_{j=1}^m p_j = p \tag{17}$$

for all $\mathbf{x} = (x_1, ..., x_m) \in \mathcal{X} = \times_{j=1}^m \mathcal{X}_j$, which proves the D-optimality of the product type design ξ^* by means of the multivariate Equivalence Theorem 3.1. □

For L-optimality it will be additionally required that the weight matrix $\mathbf{L}$ is a block diagonal matrix of marginal weight matrices, i. e. $\mathbf{L} = \mathrm{diag}(\mathbf{L}_j)_{j=1,...,m}$, where $\mathbf{L}_j$ is a positive definite $p_j \times p_j$ matrix for each $j = 1, ..., m$.

Theorem 4.2 *Assume that each univariate model contains a constant term. For each* $j = 1, ..., m$ *let* ξ_j^* *be an L-optimal design with respect to the weight matrix* $\mathbf{L}_j$ *for the* j*th univariate model on the design region* $\mathcal{X}_j$*. Then the product type design* $\xi^* = \otimes_{j=1}^m \xi_j^*$ *is an L-optimal design with respect to the weight matrix* $\mathbf{L} = \mathrm{diag}(\mathbf{L}_j)_{j=1,...,m}$ *for the multivariate model on the design region* $\mathcal{X} = \times_{j=1}^m \mathcal{X}_j$*.*

Proof. Because of the L-optimality of the design ξ_j^* the marginal sensitivity function $\varphi_{L,j}(x_j; \xi_j^*)$ is bounded by $\mathrm{trace}(\mathbf{L}_j\mathbf{M}_j(\xi_j^*)^{-1})$ for all $x_j \in \mathcal{X}_j$ in view of the univariate version of the Equivalence Theorem 3.2. By Lemma A.2 in the appendix the sensitivity function $\varphi_L(\mathbf{x}; \xi^*)$ can be related to a weighted sum of the marginal sensitivity functions $\varphi_{L,j}(x_j; \xi_j^*)$,

$$\varphi_L(\mathbf{x}; \xi^*) = \sum_{j=1}^m \frac{1}{\sigma^{(jj)}} \varphi_{L,j}(x_j; \xi_j^*) + \sum_{j=1}^m \left(\sigma_j^2 - \frac{1}{\sigma^{(jj)}}\right) \mathbf{c}_j^\top \mathbf{L}_j \mathbf{c}_j\,.$$

As $\mathrm{trace}(\mathbf{L}\mathbf{M}(\xi^*)^{-1}) = \int_{\mathcal{X}} \varphi_L(\mathbf{x}; \xi^*)\xi^*(\mathrm{d}\mathbf{x})$ and $\mathrm{trace}(\mathbf{L}_j\mathbf{M}_j(\xi_j^*)^{-1}) = \int_{\mathcal{X}_j} \varphi_{L,j}(x_j; \xi_j^*)\xi_j^*(\mathrm{d}x_j)$ the same relation holds for the right hand sides of the multivariate equivalence theorem 3.2 and its univariate counterparts, respectively,

$$\mathrm{trace}(\mathbf{L}\mathbf{M}(\xi^*)^{-1}) = \sum_{j=1}^m \frac{1}{\sigma^{(jj)}} \mathrm{trace}(\mathbf{L}_j\mathbf{M}_j(\xi_j^*)^{-1}) + \sum_{j=1}^m \left(\sigma_j^2 - \frac{1}{\sigma^{(jj)}}\right) \mathbf{c}_j^\top \mathbf{L}_j \mathbf{c}_j\,.$$

Combining these results we obtain $\varphi_L(\mathbf{x};\xi^*) \leq \mathrm{trace}(\mathbf{LM}(\xi^*)^{-1})$ for all $\mathbf{x} = (x_1, ..., x_m) \in \mathcal{X} = \times_{j=1}^m \mathcal{X}_j$, which proves the L-optimality of the product type design ξ^* by means of the multivariate Equivalence Theorem 3.2. □

The A-criterion is a special L-criterion with weight matrix $\mathbf{L} = \mathbf{I}_p$, the identity matrix of dimension $p = \sum_{j=1}^m p_j$. Clearly, $\mathbf{I}_p = \mathrm{diag}(\mathbf{I}_{p_j})_{j=1,...,m}$, and Theorem 4.2 gives the following result on A-optimality.

Corollary 4.1 *Assume that each univariate model contains a constant term. For each* $j = 1, ..., m$ *let* ξ_j^* *be an* A*-optimal design for the* j*th univariate model on the design region* $\mathcal{X}_j$*. Then the product type design* $\xi^* = \otimes_{j=1}^m \xi_j^*$ *is an* A*-optimal design for the multivariate model on the design region* $\mathcal{X} = \times_{j=1}^m \mathcal{X}_j$*.*

Also, as pointed out earlier, the *IMSE*-criterion is a special L-criterion with weight matrix $\mathbf{L} = \mathrm{diag}(\mathbf{L}_j)_{j=...,m}$, where $\mathbf{L}_j = \int_{\mathcal{X}_j} \mathbf{f}_j(x_j)\mathbf{f}_j(x_j)^\top \mu_j(\mathrm{d}x_j)$, $j = 1, ..., m$. Since the diagonal blocks $\mathbf{L}_j$ are related to the *IMSE*-criterion for the univariate component models Theorem 4.2 implies the following result on *IMSE*-optimality.

Corollary 4.2 *Assume that each univariate model contains a constant term. For each* $j = 1, ..., m$ *let* ξ_j^* *be an IMSE-optimal design for the* j*th univariate model on the design region* $\mathcal{X}_j$*. Then the product type design* $\xi^* = \otimes_{j=1}^m \xi_j^*$ *is an IMSE-optimal design for the multivariate model on the design region* $\mathcal{X} = \times_{j=1}^m \mathcal{X}_j$*.*

5 Example: SUR model with two regression lines

To illustrate the results and to provide an impression on the gain of efficiency we consider the simple example of a SUR model with straight line regression in $m = 2$ components,

$$Y_{ij} = \beta_{j1} + \beta_{j2} x_{ij} + \varepsilon_{ij} \tag{18}$$

on the unit intervals $\mathcal{X}_j = [0, 1]$, $j = 1, 2$, with covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

So the multivariate regression function $\mathbf{f}$ is given by

$$\mathbf{f}(x_1, x_2) = \begin{pmatrix} 1 & x_1 & 0 & 0 \\ 0 & 0 & 1 & x_2 \end{pmatrix}^\top$$

and by (8) the information matrix at $\mathbf{x} = (x_1, x_2)$ reads as

$$\mathbf{M}(x_1, x_2) = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & x_1 & -\rho & -\rho x_2 \\ x_1 & x_1^2 & -\rho x_1 & -\rho x_1 x_2 \\ -\rho & -\rho x_1 & 1 & x_2 \\ -\rho x_2 & -\rho x_1 x_2 & x_2 & x_2^2 \end{pmatrix}. \tag{19}$$

It is well-known that for the univariate models the D-, A- and *IMSE*-optimal designs are of the form

$$\xi_j^* = \begin{pmatrix} 0 & 1 \\ 1 - w^* & w^* \end{pmatrix},$$

where $w^* = 1/2$ for the D- and *IMSE*-optimal design and $w^* = \sqrt{2} - 1$ for the A-optimal design.

By Theorem 4.1 and Theorem 4.2 the product type design

$$\xi_D^* = \begin{pmatrix} (0,0) & (0,1) & (1,0) & (1,1) \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$$

is D- and *IMSE*-optimal, and the product type design

$$\xi_A^* = \begin{pmatrix} (0,0) & (0,1) & (1,0) & (1,1) \\ 6-4\sqrt{2} & 3\sqrt{2}-4 & 3\sqrt{2}-4 & 3-2\sqrt{2} \end{pmatrix}$$

is A-optimal for the SUR model on the unit square $\mathcal{X} = [0,1]^2$.

To simplify the analysis and to reduce the number of different design points it might be tempting in the present situation of equal regression functions $\mathbf{f}_1 = \mathbf{f}_2$ and equal design regions $\mathcal{X}_1 = \mathcal{X}_2$ to use a multivariate regression (MANOVA type) design ξ_{MV}, in which the settings x_{ij} for the components coincide ($x_{i1} = x_{i2}$). First note that in the case of uncorrelated components ($\rho = 0$) it is clear that for any design ξ the information matrix $\mathbf{M}(\xi) = \mathrm{diag}(\mathbf{M}_j(\xi_j))$ is block diagonal and depends only on the marginals ξ_j of ξ. Thus for $\rho = 0$ also the multivariate regression design

$$\xi_{\mathrm{MV}}^* = \begin{pmatrix} (0,0) & (1,1) \\ 1-w^* & w^* \end{pmatrix}$$

generated from the optimal marginal design $\xi_1^* = \xi_2^*$ is D- and *IMSE*-optimal ($w^* = w_D^* = 1/2$) resp. A-optimal ($w^* = w_A^* = \sqrt{2} - 1$) for the SUR model (18) on the unit square $\mathcal{X} = [0,1]^2$.

For arbitrary correlation ρ the information matrix of a multivariate regression design ξ_{MV} factorizes, $\mathbf{M}(\xi_{\mathrm{MV}}) = \mathbf{\Sigma}^{-1} \otimes \mathbf{M}(\xi_1)$, in general, where ξ_1 is the marginal of ξ_{MV}. Hence, for the D-, A- and *IMSE*-criterion ξ_{MV}^* generated from the optimal marginals ξ_j^* is optimal for the SUR model within the class of multivariate regression designs (cf Kurotschka and Schwabe, 1996). Moreover, it is a commonly used fact that for a multivariate regression design the ordinary and the weighted least squares estimator coincide, and no knowledge of the covariance matrix $\mathbf{\Sigma}$ is required for the estimation of the location parameters $\boldsymbol{\beta}$.

In the particular case of the present example we get from (19)

$$\mathbf{M}(\xi_{\mathrm{MV}}) = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & w^* & -\rho & -\rho w^* \\ w^* & w^* & -\rho w^* & -\rho w^* \\ -\rho & -\rho w^* & 1 & w^* \\ -\rho w^* & -\rho w^* & w^* & w^* \end{pmatrix},$$

which differs only in the entries $(2,4)$ and $(4,2)$ from the information matrix of the product type design whose $(2,4)$th ($(4,2)$th) entry equals $-\rho(w^*)^2$.

The determinant of the information matrix of the multivariate regression design equals $\det(\mathbf{M}(\xi_{\mathrm{MV}})) = (1-\rho^2)^{-2}/16$ compared to $\det(\mathbf{M}(\xi^*)) = (1-\rho^2)^{-3}/16$ for the optimal product type design. This results in a D-efficiency $\mathrm{eff}_D(\xi_{\mathrm{MV}}) = (\det(\mathbf{M}(\xi_{\mathrm{MV}}))/\det(\mathbf{M}(\xi^*)))^{1/4} = (1-\rho^2)^{1/4}$ of the multivariate regression design, which goes to zero when $|\rho|$ approaches 1 (cf Soumaya and Schwabe, 2011).

In the case of the *IMSE*-criterion the loss in efficiency is less pronounced: For the multivariate regression design its value equals $\mathrm{IMSE}(\xi_{\mathrm{MV}})) = 8/3$ independently of the correlation, while $\mathrm{IMSE}(\xi^*) = 8/3 - 2\rho^2/3$. This results in an *IMSE*-efficiency $\mathrm{eff}_{IMSE}(\xi_{\mathrm{MV}}) = \mathrm{IMSE}(\xi^*)/\mathrm{IMSE}(\xi_{\mathrm{MV}}) = 1 - \rho^2/4$ of the multivariate regression design, which goes down to 75% when $|\rho|$ approaches 1.

Similarly, the A-criterion equals $\mathrm{trace}(\mathbf{M}(\xi_{\mathrm{MV}})^{-1}) = 6 + 4\sqrt{2}$ for the multivariate regression design and $\mathrm{trace}(\mathbf{M}(\xi^*)^{-1}) = (6+4\sqrt{2})(1 - 2(\sqrt{2}-1)\rho^2)$ for the optimal product type design. This results in an A-efficiency $\mathrm{eff}_A(\xi_{\mathrm{MV}}) = \mathrm{trace}(\mathbf{M}(\xi^*)^{-1})/\mathrm{trace}(\mathbf{M}(\xi_{\mathrm{MV}})^{-1}) = 1 - 2(\sqrt{2}-1)\rho^2$ of the multivariate regression design, which drops down to 17,16% when $|\rho|$ approaches 1.

6 Discussion

For the present situation of seemingly unrelated linear models with constant terms in each component the optimal designs do not depend on the covariance matrix $\mathbf{\Sigma}$. If $\mathbf{\Sigma}$ is unknown and has to be estimated itself, the Fisher information matrix for both the location parameters $\boldsymbol{\beta}$ and the covariance matrix $\mathbf{\Sigma}$ is block-diagonal in the case of multivariate normal errors, where the block related to the location parameters $\boldsymbol{\beta}$ coincides with the information matrix (6) of the weighted least squares estimator $\hat{\boldsymbol{\beta}}_{\mathrm{WLS}}$ considered so far. Hence, on the design stage no prior knowledge is required about $\mathbf{\Sigma}$ to find an optimal design for the estimation of the location parameters $\boldsymbol{\beta}$.

However, on the estimation stage the weighted least squares estimator $\hat{\boldsymbol{\beta}}_{\mathrm{WLS}}$ specified in (5) involves the covariance matrix $\mathbf{\Sigma}$. Then $\mathbf{\Sigma}$ has to be replaced in $\hat{\boldsymbol{\beta}}_{\mathrm{WLS}}$ by a suitable estimate $\hat{\mathbf{\Sigma}}$ to obtain a feasible general least squares estimator $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ for $\boldsymbol{\beta}$. This can be achieved, for example, by maximum likelihood or by iterative weighted least squares methods starting with the ordinary least square estimator $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} = (\mathbf{F}^\top\mathbf{F})^{-1}\mathbf{F}^\top\mathbf{Y}$. Under certain regularity conditions these estimators $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ are asymptotically normal with asymptotic covariance matrix equal to the inverse of the information matrix $\mathbf{M}$ for $\boldsymbol{\beta}$ (cf Fedorov et al., 2002). So, the optimal designs obtained in the present paper are asymptotically efficient even when the covariance matrix $\mathbf{\Sigma}$ is not known.

To circumvent the dependence of the estimator on $\mathbf{\Sigma}$ another tempting approach might be to use the ordinary least squares estimator $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$, which consists of the univariate estimators $\hat{\boldsymbol{\beta}}_{j,\mathrm{OLS}} = (\mathbf{F}^{(j)\top}\mathbf{F}^{(j)})^{-1}\mathbf{F}^{(j)\top}\mathbf{Y}^{(j)}$ based on the single components. As indicated in the example of Section 5 the weighted and the ordinary least squares estimator coincide when a multivariate regression design is used. However, for the optimal product type designs the estimators may substantially differ. Moreover, when the optimal product type design is used, the performance of the ordinary least squares estimator may become even worse than in the case of a multivariate regression design. For example, in Section 5 the relative D-efficiency of the ordinary least squares estimator under the product type design ξ^{*}‘ compared to the best multivariate regression design ξ^*_{MV} equals $(1-\rho^2)^{1/4}$, which may become arbitrarily small when $|\rho|$ approaches 1.

Finally, it should be mentioned that the product type designs may be no longer optimal when there is no constant term present in the marginal models for the single components (see Soumaya and Schwabe, 2015).

A Proofs and auxiliary results

To simplify notations we suppress the dependence of the information matrices and moment vectors on the corresponding designs: $\mathbf{M}_j = \mathbf{M}_j(\xi_j)$, $\mathbf{m}_j = \mathbf{m}_j(\xi_j)$, $\mathbf{M} = \mathbf{M}(\xi)$ and $\mathbf{m} = \mathbf{m}(\xi)$.

Proof of Lemma 4.2. The proof is carried out by straightforward multiplication of the representation $\mathrm{diag}(\sigma^{(jj)}(\mathbf{M}_j - \mathbf{m}_j\mathbf{m}_j^\top)) + \mathbf{m}\mathbf{\Sigma}^{-1}\mathbf{m}^\top$ of the information matrix $\mathbf{M}$ in (15) by the expression $\mathrm{diag}((\mathbf{M}_j^{-1} - \mathbf{c}_j\mathbf{c}_j^\top)/\sigma^{(jj)}) + \mathbf{c}\mathbf{\Sigma}\mathbf{c}^\top$ on the right hand side of (16). To do so we first note that because of $\mathbf{f}_j^\top\mathbf{c}_j \equiv 1$ we get $\mathbf{m}_j^\top\mathbf{c}_j = 1$, $\mathbf{M}_j\mathbf{c}_j = \mathbf{m}_j$ and, hence, $\mathbf{M}_j^{-1}\mathbf{m}_j = \mathbf{c}_j$. It follows that $(\mathbf{M}_j - \mathbf{m}_j\mathbf{m}_j^\top)\mathbf{c}_j = \mathbf{0}$ which implies

$$\mathrm{diag}(\sigma^{(jj)}(\mathbf{M}_j - \mathbf{m}_j\mathbf{m}_j^\top))\mathrm{diag}((\mathbf{M}_j^{-1} - \mathbf{c}_j\mathbf{c}_j^\top)/\sigma^{(jj)}) = \mathrm{diag}(\mathbf{I}_{p_j} - \mathbf{m}_j\mathbf{c}_j^\top) = \mathbf{I}_p - \mathbf{m}\mathbf{c}^\top \qquad (20)$$

and

$$\mathrm{diag}(\sigma^{(jj)}(\mathbf{M}_j - \mathbf{m}_j\mathbf{m}_j^\top))\mathbf{c}\mathbf{\Sigma}\mathbf{c}^\top = \mathbf{0}\,. \qquad (21)$$

Similarly, $(\mathbf{M}_j^{-1} - \mathbf{c}_j\mathbf{c}_j^\top)\mathbf{m}_j = \mathbf{0}$ yields

$$\mathbf{m}\boldsymbol{\Sigma}^{-1}\mathbf{m}^\top \mathrm{diag}((\mathbf{M}_j^{-1} - \mathbf{c}_j\mathbf{c}_j^\top)/\sigma^{(jj)}) = \mathbf{0}\,. \tag{22}$$

Finally, because of $\mathbf{m}^\top\mathbf{c} = \mathbf{I}_m$ we get

$$\mathbf{m}\boldsymbol{\Sigma}^{-1}\mathbf{m}^\top\mathbf{c}\boldsymbol{\Sigma}\mathbf{c}^\top = \mathbf{m}\mathbf{c}^\top\,. \tag{23}$$

Combining (20) to (23) we obtain

$$\left(\mathrm{diag}(\sigma^{(jj)}(\mathbf{M}_j - \mathbf{m}_j\mathbf{m}_j^\top)) + \mathbf{m}\boldsymbol{\Sigma}^{-1}\mathbf{m}^\top\right)\left(\mathrm{diag}((\mathbf{M}_j^{-1} - \mathbf{c}_j\mathbf{c}_j^\top)/\sigma^{(jj)}) + \mathbf{c}\boldsymbol{\Sigma}\mathbf{c}^\top\right) = \mathbf{I}_p \tag{24}$$

which proves the lemma. □

For Theorem 4.1 we show that the sensitivity function $\varphi_D(\mathbf{x};\xi) = \mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{f}(\mathbf{x})^\top\mathbf{M}^{-1}\mathbf{f}(\mathbf{x}))$ of a product type design can be represented as the sum of the sensitivity functions $\varphi_{D,j}(x_j;\xi_j) = \mathbf{f}_j(x_j)^\top\mathbf{M}_j^{-1}\mathbf{f}_j(x_j)$ in the univariate models (9) of the components. In particular, the sensitivity function of a product ype design with respect to the D-criterion does not depend on the covariance matrix $\boldsymbol{\Sigma}$ of the observations within the units.

Lemma A.1 *Assume that each univariate model contains a constant term. Let ξ be a product type design, $\xi = \otimes_{j=1}^m\xi_j$, and such that for each j the information matrix $\mathbf{M}_j(\xi_j)$ of ξ_j in the jth marginal model is nonsingular. Then $\varphi_D(\mathbf{x};\xi) = \sum_{j=1}^m \varphi_{D,j}(x_j;\xi_j)$ for all $\mathbf{x} = (x_1,...,x_m)$.*

Proof. By the representation of the inverse of the information matrix in Lemma 4.2 the sensitivity function φ_D can be written as

$$\varphi_D(\mathbf{x}) = \mathrm{trace}\left(\boldsymbol{\Sigma}^{-1}\mathbf{f}(\mathbf{x})^\top\mathrm{diag}\left(\frac{1}{\sigma^{(jj)}}(\mathbf{M}_j^{-1} - \mathbf{c}_j\mathbf{c}_j^\top) + \mathbf{c}\boldsymbol{\Sigma}\mathbf{c}^\top\right)\mathbf{f}(\mathbf{x})\right)\,. \tag{25}$$

Again we make use of the identity $\mathbf{f}_j(x_j)^\top\mathbf{c}_j = 1$ and, hence, $\mathbf{f}(\mathbf{x})^\top\mathbf{c} = \mathbf{I}_m$ to obtain

$$\varphi_D(\mathbf{x}) = \mathrm{trace}\left(\boldsymbol{\Sigma}^{-1}\left(\mathrm{diag}\left((\mathbf{f}_j(x_j)^\top\mathbf{M}_j^{-1}\mathbf{f}_j(x_j) - 1)/\sigma^{(jj)}\right) + \boldsymbol{\Sigma}\right)\right)\,. \tag{26}$$

Because $\mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\alpha_j)) = \sum_{j=1}^m \sigma^{(jj)}\alpha_j$ for any α_j we can derive from (26) that

$$\varphi_D(\mathbf{x}) = \sum_{j=1}^m(\mathbf{f}_j(x_j)^\top\mathbf{M}_j^{-1}\mathbf{f}_j(x_j) - 1) + \mathrm{trace}(\mathbf{I}_m) = \sum_{j=1}^m\varphi_{D,j}(x_j) \tag{27}$$

which concludes the proof. □

For Theorem 4.2 the sensitivity function $\varphi_L(\mathbf{x};\xi) = \mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{f}(\mathbf{x})^\top\mathbf{M}^{-1}\mathbf{L}\mathbf{M}^{-1}\mathbf{f}(\mathbf{x}))$ can be seen to be a weighted sum of the sensitivity functions $\varphi_{L,j}j(x_j;\xi_j) = \mathbf{f}_j(x_j)^\top\mathbf{M}_j^{-1}\mathbf{L}_j\mathbf{M}_j^{-1}\mathbf{f}_j(x_j))$ in the univariate models (9) of the components plus a correction term adjusting for correlations which does not depend on the design.

Lemma A.2 *Assume that each univariate model contains a constant term. Let ξ be a product type design, $\xi = \otimes_{j=1}^m\xi_j$, and such that for each j the information matrix $\mathbf{M}_j(\xi_j)$ of ξ_j in the jth marginal model is nonsingular. Then*

$$\varphi_L(\mathbf{x};\xi) = \sum_{j=1}^m\frac{1}{\sigma^{(jj)}}\varphi_{L,j}(x_j;\xi_j) + \sum_{j=1}^m\left(\sigma_j^2 - \frac{1}{\sigma^{(jj)}}\right)\mathbf{c}_j^\top\mathbf{L}_j\mathbf{c}_j$$

for all $\mathbf{x} = (x_1,...,x_m)$.

Proof. Because of $\mathbf{f}_j(x_j)^\top \mathbf{c}_j = 1$ we obtain from the representation (16) of the inverse of the information matrix of a product type design that

$$\mathbf{M}^{-1}\mathbf{f}(\mathbf{x}) = \mathrm{diag}\left(\frac{1}{\sigma^{(jj)}}(\mathbf{M}_j^{-1}\mathbf{f}_j(x_j) - \mathbf{c}_j)\right) + \mathbf{c}\boldsymbol{\Sigma} = \mathrm{diag}\left(\frac{1}{\sigma^{(jj)}}\mathbf{M}_j^{-1}\mathbf{f}_j(x_j)\right) + \mathbf{c}\boldsymbol{\Sigma}_0\,, \qquad (28)$$

where $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma} - \mathrm{diag}(1/\sigma^{(jj)})$ accounts for the correlation. Note that $\boldsymbol{\Sigma}_0 = \mathbf{0}$ if the components are uncorrelated ($\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_j^2)$).

Then the sensitivity function φ_L of a product type design can be written as

$$\begin{aligned}\varphi_L(\mathbf{x}) \;=\; & \mathrm{trace}\left(\boldsymbol{\Sigma}^{-1}\mathrm{diag}\left(\mathbf{f}_j(x_j)^\top\mathbf{M}_j^{-1}\mathbf{L}_j\mathbf{M}_j^{-1}\mathbf{f}_j(x_j)/(\sigma^{(jj)})^2\right)\right)\\ & + 2\,\mathrm{trace}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0\,\mathrm{diag}\left(\mathbf{c}_j^\top\mathbf{L}_j\mathbf{M}_j^{-1}\mathbf{f}_j(x_j)/\sigma^{(jj)}\right)\right)\\ & + \mathrm{trace}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0\,\mathbf{c}^\top\mathbf{L}\mathbf{c}\,\boldsymbol{\Sigma}_0) \qquad (29)\end{aligned}$$

The first term on the right hand side of (29) is equal to $\sum_{j=1}^m \varphi_{L,j}(x_j)/\sigma^{(jj)}$. The second term vanishes because all diagonal entries of $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0$ are equal to 0. For the last term note that $\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma} - 2\,\mathrm{diag}(1/\sigma^{(jj)}) + \mathrm{diag}(1/\sigma^{(jj)})\boldsymbol{\Sigma}^{-1}\mathrm{diag}(1/\sigma^{(jj)})$ with diagonal entries equal to $\sigma_j^2 - 1/\sigma^{(jj)}$. Hence this term reduces to $\sum_{j=1}^m(\sigma_j^2 - 1/\sigma^{(jj)})\mathbf{c}_j^\top\mathbf{L}_j\mathbf{c}_j$ which completes the proof of the lemma. □

References

Fedorov , V.V. (1972): *Theory of Optimal Experiments.* Academic Press, New York.

Fedorov, V.V., Gagnon, R.C. and Leonov, S.L. (2002): Design of experiments with unknown parameters in variance. *Applied Stochastic Models in Business and Industry* **18**, 207–218.

Kiefer, J. (1974): General equivalence theory for optimum designs (approximate theory). *Annals of Statistics* **2**, 849–879.

Kurotschka, V.G. and Schwabe, R. (1996): The reduction of design problems for multivariate experiments to univariate possibilities and their limitations . In E. Brunner and M. Denker (eds.): *Research Developments in Probability and Statistics.* VSP, Utrecht, 193–204.

Schwabe, R. (1996): *Optimum Designs for Multi-factor Models.* Springer, New York.

Soumaya, M. and Schwabe, R. (2011): *D*-optimal design for a seemingly unrelated linear model. In: V. Melas, G. Nachtmann and D. Rasch (eds.): *Optimal Design of Experiments - Theory and Application.* Center of Experimental Design, University of Natural Resources and Life Sciences, Vienna, 170–174.

Soumaya, M. and Schwabe, R. (2015): On the impact of correlation on the optimality of product-type designs in SUR models. In: A. Steland, E. Rafajłowicz and K. Szajowski (eds.): *Stochastic Models, Statistics and Their Applications.* Springer, New York. 159–167.

Zellner, A. (1962): An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *Journal of the American Statistical Association* **57**, 348–368.