On the Impact of Correlation on the Optimality of Product-type Designs in SUR Models

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

1. Institute of Process Engineering, Otto-von-Guericke University, PF 4120, D-39016 Magdeburg, Germany
2. Institute for Mathematical Stochastics, Otto-von-Guericke University, PF 4120, D-39016 Magdeburg, Germany

Abstract: For multivariate observations with seemingly unrelated variables product-type designs often turn out to be optimal which are generated by their univariate optimal counterparts. This is, in particular, the case when all variables contain an intercept term. If these intercepts are missing, the product-type designs may lose their optimality when the correlation between the components becomes stronger.

Key Words: multivariate linear model, seemingly unrelated regression, SUR, optimal design, product-type design, optimal control, pharmacokinetics, kinetic model

1 Introduction

In many applications two or more variables are observed at the same observational units. These observations will be typically correlated, even if the variables are observed at different time points and under different experimental conditions. When additionally each observational variable is influenced by a separate set of explanatory variables one might be tempted to perform univariate analyses for each variable separately. Such models of a seemingly unrelated regression (SUR) type have been introduced by Zellner (1962) in the context of economic data, who showed that a joint analysis, which accounts for the correlation of the variables, will improve the precision of estimates and tests.

Such models also occur in many scientific fields when problems or phenomena are investigated in pharmacology, toxicology, process engineering etc. at different time points or input variables. For example, when we are interested in the pharmacokinetic measurement of the concentration of two or more drugs or the toxicity with different dissolution times, then the problem can be adequately described by a SUR model.

To be more specific consider the univariate regression model of concentration of a substance, hormone, drug or of the toxicity with response function $\eta(t) = Ae^{-t\theta}$, where A and θ are the initial concentration and reaction rate (see Atkinson et al., 2007). When we observe two such relations at the same observational units but potentially at different time points, we obtain multivariate data with correlated components and marginal response functions $\eta_1(t_1) = A_1e^{-t_1\theta_1}$ and $\eta_2(t_2) = A_2e^{-t_2\theta_2}$, respectively. We will distinguish between two cases, where either the initial concentrations A_1 and A_2 are known or not depending on the experimental situation. Moreover, we assume that the data are appropriately modeled on a logarithmic scale which leads to a linear model formulation for the components

$$\begin{aligned} Y_{i1} &= \beta_{11} + \beta_{12}x_{i1} + \varepsilon_{i1}, \quad \beta_{11} = \ln A_1,\ \beta_{12} = \theta_1,\ x_{i1} = t_{i1} \\ Y_{i2} &= \beta_{21} + \beta_{22}x_{i2} + \varepsilon_{i2}, \quad \beta_{21} = \ln A_2,\ \beta_{22} = \theta_2,\ x_{i2} = t_{i2} \end{aligned} \tag{1}$$

If the initial concentrations A_1 and A_2 are known, the marginal models can be reduced to

$$\begin{aligned} Y_{i1} &= \beta_{12}x_{i1} + \varepsilon_{i1} \\ Y_{i2} &= \beta_{22}x_{i2} + \varepsilon_{i2} \end{aligned} \tag{2}$$

*Corresponding author: moudar.soumaya@gmail.com

In any case the SUR structure comes in when we admit that the observations are correlated ($\mathrm{Cov}(Y_{i1}, Y_{i2}) = \mathrm{Cov}(\varepsilon_{i1}, \varepsilon_{i2}) \neq 0$) within the observational units.

While the data analysis has been widely investigated in such models, hardly anything has been done in the design of such experiments. This is partly due to the fact that these models are mainly considered in economic applications, where there are only data available from observational studies, but this is also caused by a wide-spread opinion that everything is clear about optimal design for multivariate observations. In particular, it seems to be obvious that designs, which are optimal in the univariate settings, will also be optimal in the multivariate case. This is definitely true for MANOVA or multivariate regression settings, where the univariate models depend on the same (values of the) explanatory variables and all univariate models coincide (see e. g. Kurotschka and Schwabe, 1996). For various design criteria the corresponding proofs are based on multivariate equivalence theorems by Fedorov (1972).

In the situation of SUR models techniques concerning product-type designs have to be employed, which were developed in Schwabe (1996). Soumaya and Schwabe (2011) established that in the presence of intercept terms the D-optimal design can be generated as a product of the D-optimal counterparts in the corresponding univariate models for the single components irrespectively of the underlying covariance structure. These results were widely extended to other design criteria by Soumaya (2013). In the absence of intercept terms the product-type design may lose its optimality when the correlation becomes large.

It should be mentioned that against common believe even in the case of identical models for all components the MANOVA-type design, in which the settings of the explanatory variables are the same for all components, turns out to be not optimal, in general, when the observations are correlated.

The paper is organized as follows: In the second section we specify the model, and we characterize optimal designs in the third section. In section 4 the results are illustrated by means of an example in the bivariate case and some conclusions are given in section 5.

2 Model specification

In general for the m components the m-dimensional observations are described by m model equations. The components of the multivariate observations can be heterogeneous, which means that the response can be specified by different regression functions and different experimental settings, which may be chosen from different experimental regions.

The observation of the jth component of individual i can be described by

$$Y_{ij} = \mathbf{f}_j(x_{ij})^\top \boldsymbol{\beta}_j + \varepsilon_{ij} = f_{j1}(x_{ij})\beta_{j1} + ... + f_{jp_j}(x_{ij})\beta_{jp_j} + \varepsilon_{ij}\,, \tag{3}$$

where $\mathbf{f}_j = (f_{j1}, ..., f_{jp_j})^\top$ are known regression functions, $\boldsymbol{\beta}_j = (\beta_{j1}, ..., \beta_{jp_j})^\top$ are the unknown parameter vectors and the experimental settings x_{ij} may be chosen from experimental regions $\mathcal{X}_j$.

Denote by $\mathbf{Y}_i = (Y_{i1}, ..., Y_{im})^\top$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{im})^\top$ the multivariate vectors of observations and error terms for individual i. Accordingly the multivariate regression function is block diagonal

$$\mathbf{f}(\mathbf{x}) = \mathrm{diag}\,(\mathbf{f}_j(x_j))_{j=1,...,m}$$

for the multivariate experimental setting $\mathbf{x} = (x_1, ..., x_m) \in \mathcal{X} = \times_{j=1}^m \mathcal{X}_j$. For the examples (1) and (2) the regression functions have the following forms

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} 1 & x_1 & 0 & 0 \\ 0 & 0 & 1 & x_2 \end{pmatrix}^\top \quad \text{resp.} \quad \mathbf{f}(\mathbf{x}) = \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix}.$$

The individual observation vector can be written as

$$\mathbf{Y}_i = \mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \,, \tag{4}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, ..., \boldsymbol{\beta}_m^\top)^\top$ is the complete stacked parameter vector for all components of dimension $p = \sum_{j=1}^m p_j$. For the error vectors $\boldsymbol{\varepsilon}_i$ it is assumed that they have zero mean and have a common positive definite covariance matrix $\mathrm{Cov}\,(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma}$ within the observational units, while they are uncorrelated across the observational units.

Finally, 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 overall observation vector as

$$\mathbf{Y} = \mathbf{F}\boldsymbol{\beta} + \varepsilon \,, \tag{5}$$

where $\mathbf{F} = (\mathbf{f}(\mathbf{x}_1), ..., \mathbf{f}(\mathbf{x}_n))^\top$ is the overall design matrix. The full observational error vector $\boldsymbol{\varepsilon}$ then has the 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 we assume that the covariance matrix $\boldsymbol{\Sigma}$ and, hence, $\mathbf{V}$ is known, we can estimate the parameter $\boldsymbol{\beta}$ efficiently by the Gauss-Markov estimator

$$\hat{\boldsymbol{\beta}}_{\mathrm{GM}} = (\mathbf{F}^\top \mathbf{V}^{-1} \mathbf{F})^{-1} \mathbf{F}^\top \mathbf{V}^{-1} \mathbf{Y} \,. \tag{6}$$

Its covariance matrix is equal to the inverse of the corresponding information matrix

$$\mathbf{M} = \mathbf{F}^\top \mathbf{V}^{-1} \mathbf{F} = \sum_{i=1}^n \mathbf{f}(\mathbf{x}_i) \boldsymbol{\Sigma}^{-1} \mathbf{f}(\mathbf{x}_i)^\top \,, \tag{7}$$

which is the sum of the individual informations $\mathbf{M}(\mathbf{x}_i) = \mathbf{f}(\mathbf{x}_i)\boldsymbol{\Sigma}^{-1}\mathbf{f}(\mathbf{x}_i)^\top$.

Note that the univariate marginal models of the components are of the form

$$\mathbf{Y}^{(j)} = \mathbf{F}^{(j)} \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}^{(j)} \,, \tag{8}$$

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 for the jth component, respectively, and $\mathbf{F}^{(j)} = (\mathbf{f}_j(x_{1j}), ..., \mathbf{f}_j(x_{nj}))^\top$ is the design matrix for the jth marginal model. The corresponding 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 $\boldsymbol{\Sigma}$.

3 Optimal designs

We can define an experimental design in the multivariate case

$$\xi = \begin{pmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_k \\ w_1 & \dots & w_k \end{pmatrix} \tag{9}$$

by the set of all different experimental settings $\mathbf{x}_i = (x_{i1}, \dots, x_{im})$, $i = 1, ..., k$, which belong to the design region $\mathcal{X} = \times_{j=1}^m \mathcal{X}_j$, with the corresponding relative frequencies $w_i = \frac{n_i}{n}$, where n_i is the number of replications at $\mathbf{x}_i$. Then the corresponding standardized information matrix for the GM-estimator can be obtained as

$$\mathbf{M}(\xi) = \sum_{i=1}^k w_i \mathbf{f}(\mathbf{x}_i) \boldsymbol{\Sigma}^{-1} \mathbf{f}(\mathbf{x}_i)^\top. \tag{10}$$

For analytical purposes we consider approximate designs, see for example Kiefer (1974), for which the weights $w_i \geq 0$ need not be multiples of $\frac{1}{n}$, but only have to satisfy $\sum_{i=1}^k w_i = 1$.

As information matrices are not necessarily comparable, we have to consider some real-valued criterion function of the information matrix. In this paper we will adopt the most popular criterion $-\ln\det(\mathbf{M}(\xi))$ of D-optimality and some linear criteria $\mathrm{trace}(\mathbf{L}\mathbf{M}(\xi)^{-1})$ like A- and IMSE-optimality, where $\mathbf{L}$ is a positive definite weight matrix.

An approximate design ξ_D^* is called D-optimal if it minimizes the determinant of the variance covariance matrix, i. e. it minimizes the volume of the confidence ellipsoid under the assumption of Gaussian errors. An approximate design ξ_A^* is called A-optimal if it minimizes the trace $\mathrm{trace}(\mathbf{M}(\xi)^{-1})$ of the variance covariance matrix, i. e. it minimizes the average of the variances of the parameter estimates. Hence, the A-criterion is linear with $\mathbf{L}$ equal to the identity matrix $\mathbf{I}_p$.

The integrated mean squared error is the integrated predictive covariance with respect to the uniform measure $\mu(\mathrm{d}\mathbf{x})$ on the design region $\mathcal{X}$ and will be defined as follows

$$\mathrm{IMSE} = \int_{\mathcal{X}} \mathrm{E}\left(\|\mathbf{f}(\mathbf{x})^\top(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\|^2\right)\mu(\mathrm{d}\mathbf{x}) = \int_{\mathcal{X}} \mathrm{trace}\left(\mathrm{Cov}(\mathbf{f}(\mathbf{x})^\top\hat{\boldsymbol{\beta}})\right)\mu(\mathrm{d}\mathbf{x})\,,$$

where $\|.\|$ denotes the Euclidean norm. Then an approximate design ξ_{IMSE}^* is called IMSE-optimal in the multivariate case, if it minimizes the averaged predictive variance $\mathrm{E}\left(\|\mathbf{f}(\mathbf{x})^\top(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\|^2\right)$. After rearranging terms the IMSE-criterion is equivalent to a linear criterion with

$$\mathbf{L} = \int_{\mathcal{X}} \mathbf{f}(\mathbf{x})\mathbf{f}(\mathbf{x})^\top\,\mu(\mathrm{d}\mathbf{x})\,. \tag{11}$$

Useful tools for checking the optimality of given candidate designs are the following multivariate versions of the equivalence theorems for the above mentioned criteria (see Fedorov (1972), Theorems 5.2.1 and 5.3.1).

Theorem 3.1 *The approximate design ξ_D^* is D-optimal in the multivariate linear model if and only if*

$$\varphi_D(\mathbf{x};\xi_D^*) := \mathrm{trace}\left(\boldsymbol{\Sigma}^{-1}\mathbf{f}(\mathbf{x})^\top\mathbf{M}(\xi_D^*)^{-1}\mathbf{f}(\mathbf{x})\right) \le p\,, \tag{12}$$

for all $x\in\mathcal{X}$, where $p=\sum_{j=1}^m p_j$ is the number of parameters in the model.

Theorem 3.2 *The approximate design ξ_L^* is linear optimal in the multivariate linear model if and only if*

$$\varphi_L(\mathbf{x};\xi_L^*) := \frac{\mathrm{trace}\left(\boldsymbol{\Sigma}^{-1}\mathbf{f}(\mathbf{x})^\top\mathbf{M}(\xi_L^*)^{-1}\mathbf{L}\ \mathbf{M}(\xi_L^*)^{-1}\mathbf{f}(\mathbf{x})\right)}{\mathrm{trace}\left(\mathbf{L}\ \mathbf{M}(\xi_L^*)^{-1}\right)} \le 1\,, \tag{13}$$

for all $x\in\mathcal{X}$.

For A-optimality $\mathbf{L}=\mathbf{I}_p$ and for the IMSE-optimality $\mathbf{L}=\int_{\mathcal{X}}\mathbf{f}(\mathbf{x})\mathbf{f}(\mathbf{x})^\top\mu(\mathrm{d}\mathbf{x})$.

Note that the linear optimality criteria considered here (A and IMSE) result in a block-diagonal weight matrix $\mathbf{L}=\mathrm{diag}(\mathbf{L}_j)_{j=1,\dots,m}$, where the diagonal blocks $\mathbf{L}_j$ are the weight matrices for the corresponding linear criteria in the marginal components: $\mathbf{I}_p=\mathrm{diag}(\mathbf{I}_{p_j})_{j=1,\dots,m}$ for the A-criterion and $\int_{\mathcal{X}}\mathbf{f}(\mathbf{x})\mathbf{f}(\mathbf{x})^\top\mu(\mathrm{d}\mathbf{x}) = \mathrm{diag}(\int_{\mathcal{X}_j}\mathbf{f}(x_j)\mathbf{f}(x_j)^\top\mu_j(\mathrm{d}x_j))_{j=1,\dots,m}$ for the IMSE-criterion, where μ_j is the uniform distribution on $\mathcal{X}_j$.

In the case that all marginal models contain an intercept term the following result has been established by Soumaya (2013).

Theorem 3.3 *Let ξ_j^* be Φ-optimal for the jth marginal component (8) with an intercept on the marginal design region $\mathcal{X}_j$, then the product-type design $\xi^*=\otimes_{j=1}^m\xi_j^*$ is Φ-optimal for the multivariate SUR model (5) on the design region $\mathcal{X}=\times_{j=1}^m\mathcal{X}_j$.*

Φ can be either the D-criterion or a linear criterion with block-diagonal weight matrix $\mathbf{L}$.

For the D-criterion the sensitivity function φ_D does not depend on the covariance matrix $\boldsymbol{\Sigma}$.

This theorem states that the D-, A- or IMSE-optimal designs can be obtained as the product of their corresponding counterparts in the marginal models.

To retain the optimality of product-type designs also in the case that all or some of the marginal models are lacking an intercept term additional orthogonality conditions have to be imposed similarly to results for additive models (see Schwabe, 1996, section 5.2).

Theorem 3.4 *Let ξ_j^* be Φ-optimal for the jth marginal component (8) without intercept on the marginal design region $\mathcal{X}_j$. If the marginal components are uncorrelated ($\mathbf{\Sigma} = \mathrm{diag}(\sigma_j^2)_{j=1,\ldots,m}$) or the regression functions are orthogonal to a constant with respect to the Φ-optimal designs ξ_j^*, i. e. $\int_{\mathcal{X}_j} \mathbf{f}_j(x_j)\,\xi_j^*(\mathrm{d}x_j) = \mathbf{0}$, then the product-type design $\xi^* = \otimes_{j=1}^m \xi_j^*$ is Φ-optimal for the multivariate SUR model (5) on the design region $\mathcal{X} = \times_{j=1}^m \mathcal{X}_j$.*

Φ can be either the D-criterion or a linear criterion with block-diagonal weight matrix $\mathbf{L}$.

For the D-criterion the sensitivity function φ_D does not depend on the covariance matrix $\mathbf{\Sigma}$.

The conditions of the above theorem guarantee that the information matrix of the product-type design ξ^* is block-diagonal such that the considered criteria as well as the associated sensitivity functions decompose.

If these conditions are violated, the overall optimality of the product-type designs can fail. The sensitivity function may depend crucially on the correlation. While for small correlations the product-type designs may be still optimal, they will lose their optimality for stronger correlations. In particular, we will exhibit by an example that the product-type designs are D-optimal for SUR models without intercepts for a restricted range of correlations around 0. For large correlations the D-optimal designs have weights related to the size of the correlation. In the present example the optimal weights tend to those for an additive model without intercept as considered in Schwabe (1996, section 5.2).

For linear criteria the situation is even worse, as the optimal weights also depend on the ratio of the variances for the single components. If the variances differ by a large factor, the product-type design may not be optimal even for very small correlation.

Example 1 (without intercept). To illustrate the results for SUR models without intercepts in the marginal models we consider the SUR model (2) on the unit intervals $\mathcal{X}_1 = \mathcal{X}_2 = [0,1]$ as experimental regions. Then the setting $x_{ij} = 0$ gives a measurement at a baseline.

It is well-known that in the marginal models the optimal designs are all concentrated on the setting $x_j^* = 1$ with maximal response. The resulting product-type design $\xi_\otimes = \xi_1^* \otimes \xi_2^*$ is also a one-point design concentrated on $\mathbf{x}^* = (1,1)$. Let

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

The information matrix for a one-point design $\xi_{\mathbf{x}}$, i. e. a single observation at $\mathbf{x} = (x_1, x_2)$, equals

$$\mathbf{M}(\xi_{\mathbf{x}}) = \mathbf{M}(x_1, x_2) = \frac{1}{1-\rho^2} \begin{pmatrix} x_1/\sigma_1^2 & -\rho x_1 x_2/\sigma_1\sigma_2 \\ -\rho x_1 x_2/\sigma_1\sigma_2 & x_2/\sigma_2^2 \end{pmatrix}.$$

In particular, $\mathbf{M}(\xi_\otimes) = \mathbf{\Sigma}^{-1}$ for the product-type design $\xi_\otimes$ with optimal marginals. In this model the weight matrix $\mathbf{L}$ for the IMSE-criterion equals $\frac{1}{3}\mathbf{I}_2$. Hence, A- and IMSE-optimality coincide. For the uncorrelated case ($\rho = 0$) the product-type design $\xi_\otimes$ is D-, A- and IMSE-optimal by Theorem 3.4.

In general, for the D-criterion the sensitivity function φ_D of the product-type design $\xi_\otimes$ equals

$$\varphi_D(\mathbf{x}; \xi_\otimes) = \tfrac{1}{1-\rho^2}(x_1^2 + x_2^2 - 2\rho^2 x_1 x_2)\,, \qquad (14)$$

Table 1: Intervals for ρ for which $\xi_\otimes$ is A-optimal

τ^2	interval for ρ	length
1	$[-\sqrt{3}/3, \sqrt{3}/3]$	$2\sqrt{3}/3$
4	$[-0.40, 0.40]$	0.80
9	$[-0.30, 0.30]$	0.60
16	$[-0.23, 0.23]$	0.46
100	$[-0.09, 0.09]$	0.18
400	$[-0.04, 0.04]$	0.08
900	$[-0.03, 0.03]$	0.06
2500	$[-0.01, 0.01]$	0.02
10000	$[-0.00, 0.00]$	0.00

which is bounded by $p = 2$ as long as $|\rho| \leq 1/\sqrt{2} \approx 0,7071$. This sensitivity function is plotted in Figures 1, 2 and 3 for the values $\rho = 0$, $\rho = 0.5$ and $\rho = 1/\sqrt{2}$, respectively. Note that in the last case the sensitivity function equals $p = 2$ at the support point $\mathbf{x}^* = (1, 1)$ and additionally at the settings $(0, 1)$ and $(1, 0)$, where one of the components is observed at baseline.

For larger values of the correlation, $|\rho| > 1/\sqrt{2}$, the D-optimal design ξ^* also contains these additional points with weights increasing in $|\rho|$,

$$\xi_D^* = \begin{pmatrix} (1,1) & (1,0) & (0,1) \\ 1-2w^* & w^* & w^* \end{pmatrix} \quad \text{with} \quad w^* = \frac{2\rho^2 - 1}{4\rho^2 - 1}$$

The sensitivity function

$$\varphi_D(\mathbf{x}; \xi_D^*) = 2(x_1^2 + x_2^2 - x_1 x_2) \tag{15}$$

is bounded by $p = 2$. This sensitivity function coincides with φ_D for $\xi_\otimes$ in the boundary case $\rho = 1/\sqrt{2}$.

For the linear criteria the standardized sensitivity function $\varphi_A = \varphi_{\text{IMSE}}$ for the product-type design $\xi_\otimes$ is

$$\varphi_A(\mathbf{x}; \xi_\otimes) = \frac{(1+\rho^2)x_1^2 + \tau^2(1+\rho^2)x_2^2 - 2(1+\tau^2)\rho^2 x_1 x_2}{(1+\tau^2)(1-\rho^2)}, \tag{16}$$

where $\tau^2 = \frac{\sigma_2^2}{\sigma_1^2}$ is the variance ratio.

For $\rho = 0$, $\sigma_1 = 0.1$ and $\sigma_2 = 10$ the sensitivity function φ_A is equal to $0.0001x_1^2 + 0.9999x_2^2$ and plotted in Figure 4. For $\rho = 0.3$, $\sigma_1 = 1$ and $\sigma_2 = 3$ the sensitivity function is computed to $0.1989x_1^2 - 0.1978x_1x_2 + 0.9989x_2^2$ and is plotted in Figure 5. In these two cases the product-type design $\xi_\otimes$ appear to be optimal.

For $\sigma_1 = 0.1$, $\sigma_2 = 10$ and $\tau^2 = 10000$ the maximum of the sensitivity function φ_A is attained at $\mathbf{x} = (0, 1)$ and exceeds 1 even for ρ close to 0. Therefor the product-type designs cannot be A-optimal for extremely different variances.

Numerical calculations indicate that for every variance ratio τ^2 there is an interval for the correlations ρ centered at 0, for which the product-type design $\xi_\otimes$ is A-optimal, while it loses its olptimality for larger values of $|\rho|$. Some numerical examples are given in Table 1. If τ^2 increases, these intervals may become arbitrarily small. Note that for symmetry reasons the same intervals are valid for τ^2 replaced by $1/\tau^2$.

For equal variances ($\tau^2 = 1$) the product-type design $\xi_\otimes$ is A-optimal as long as $|\rho| \leq 1/\sqrt{3}$,

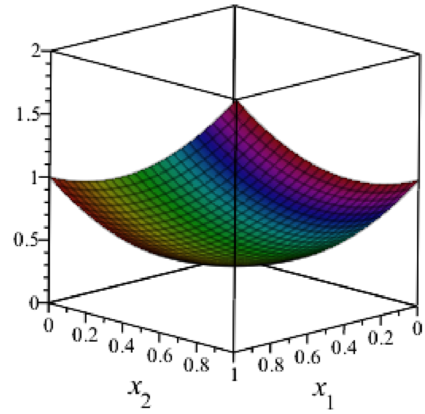

Figure 1: φ_D for $\xi_\otimes$, $\rho = 0$

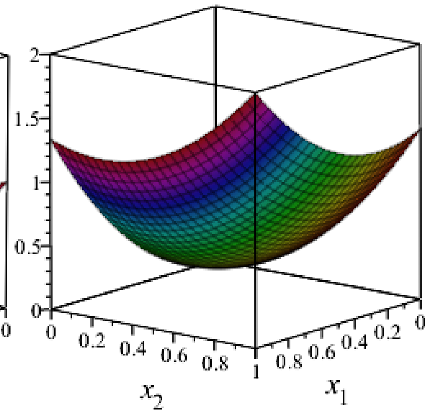

Figure 2: φ_D for $\xi_\otimes$, $\rho = 0.5$

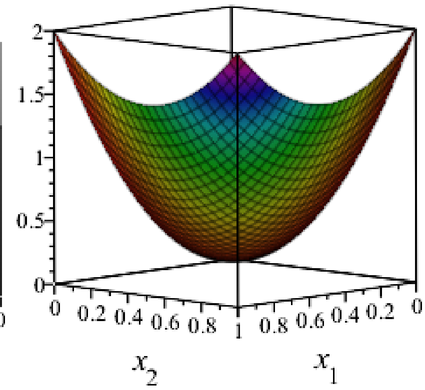

Figure 3: φ_D for ξ_D^*, $\rho \geq 1/\sqrt{2}$

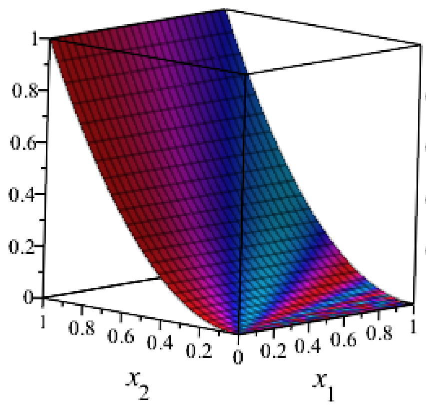

Figure 4: φ_A for $\xi_\otimes$, $\tau = 0.01$, $\rho = 0$

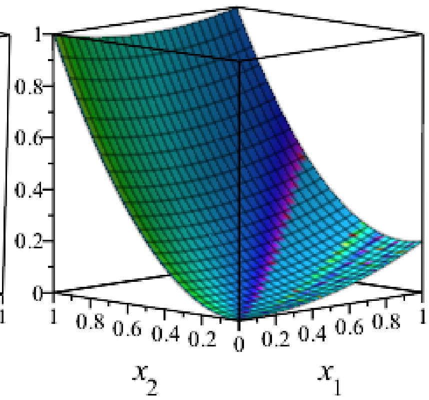

Figure 5: φ_A for $\xi_\otimes$, $\tau = 1/3$, $\rho = 0.30$

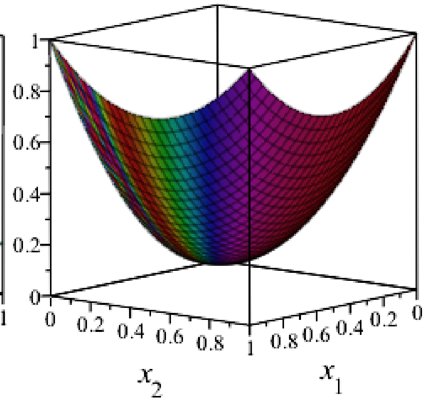

Figure 6: φ_A for ξ_A^*, $\tau = 1$, $\rho \geq 1/\sqrt{3}$

while for stronger correlations, $|\rho| > 1/\sqrt{3}$, the A-optimal design is

$$\xi_A^* = \begin{pmatrix} (1,1) & (1,0) & (0,1) \\ 1-2w^* & w^* & w^* \end{pmatrix} \quad \text{with} \quad w^* = 1 - \sqrt{\frac{\rho^2}{4\rho^2 - 1}}\,.$$

For $\rho = 1/\sqrt{3}$ the standardized sensitivity function $\varphi_A(\mathbf{x}; \xi_A^*) = x_1^2 + x_2^2 - x_1 x_2$ is plotted in Figure 6.

Example 2 (with intercept). For completeness we present the D-, IMSE- and A-optimal designs for the SUR model (1) which are generated as product type designs

$$\xi^* = \begin{pmatrix} (1,1) & (1,0) & (0,1) & (0,0) \\ w_1^2 & w_1 w_0 & w_1 w_0 & w_0^2 \end{pmatrix},$$

where $w_1 = w_0 = 1/2$ for D- and IMSE-optimality and $w_1 = \sqrt{2} - 1 \approx 0.41$ and $w_0 = 2 - \sqrt{2} \approx 0.59$ for A-optimality.

4 Conclusions

In the presence of intercept terms the variance covariance structure does not affect the D-optimality and linear optimality (block-diagonal $\mathbf{L}$) of the product design, which is generated by the corresponding optimal counterparts. But the D- and linear optimality of the product design reaches its limits for SUR models without intercepts, when the information matrix is not block-diagonal. Then product-type designs are D-optimal for this kind of SUR models only for weak to moderate correlation. Otherwise the weights of the optimal designs become dependent on the size of the correlation. Concerning linear optimality the product-type designs show a similar behavior, when the variances are equal, but perform worse, if the variances differ substantially. These results may be extended to higher dimensions, when a homogneneous correlation structure is asumed.

References

[1] Atkinson, A.C., Donev, A.N. and Tobias R.D. (2007): Optimum Experimental Design, with SAS. *Oxford University Press, Oxford.*

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

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

[4] 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, pp. 193–204.*

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

[6] Soumaya, M. (2013): Optimal designs for multivariate linear models. Dissertation, Otto-von-Guericke University Magdeburg, Faculty of Mathematics,

[7] 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: University of Natural Resources and Life Sciences, Vienna, pp. 170–174.*

[8] 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.*