THE REDUCTION OF DESIGN PROBLEMS FOR MULTIVARIATE EXPERIMENTS TO UNIVARIATE POSSIBILITIES AND THEIR LIMITATIONS

Viktor G. Kurotschka, Rainer Schwabe[1]

Freie Universität Berlin, 1. Mathematisches Institut,
Arnimallee 2–6, D-14 195 Berlin, Germany

ABSTRACT

Situations are exhibited in which optimum designs for univariate models remain optimum for multiresponse models.

1. Introduction

A multivariate experimental situation is defined by the fact, that for each of the more or less complex experimental conditions $t \in T$ the observation $X(t)$ is multivariate, i.e. one observes instead of real valued random variables r-dimensional real valued vectors

$$X(t) = (X^{(1)}(t), \ldots, X^{(r)}(t)), \quad t \in T, \tag{1}$$

with unknown location, here for simplicity expected responses

$$\mu(t) := E(X(t)) = (E(X^{(1)}(t)), \ldots, E(X^{(r)}(t))) \in \mathbb{R}^r, \quad t \in T, \tag{2}$$

and correlation structure $cov(X(t)) = \Sigma \in \mathbb{R}^{s,+}_{r \times r}$ (the set of symmetric, positive definite $r \times r$ matrices) which is generally assumed to be invariant under change of experimental conditions and replications, because it represents the interior relationship of the components defining the r-dimensional vector of observations.

Despite the different non-linear approaches to experiments under even less complex experimental conditions $t \in T$ the most appropriate setting is still apparently a linear parametrization $\beta = (\beta_1, ..., \beta_p)$ of $\mu = \sum_{\pi=1}^{p} a_\pi^T \beta_\pi$, for several reasons:

[1] research supported by grant Ku719/2-1 of the Deutsche Forschungsgemeinschaft

1) In the case of a *small* experimental region, mathematically described by a finite set T (which covers the whole area of the classically called *analysis of variance models* or *models with qualitative factors*), the practically most useful parametrizations are necessarily linear.

2) In the case of a *large* set T, which is sufficiently well described by a convex compact subset of $\mathbb{R}^K$, the transparent approach via an appropriate system of adequate regression functions $a_1, ..., a_p$ on T, approximating the unknown response μ of the experiment, is still the most attractive known, in particular, in view of the easier mathematical treatment which uses *continuous* mathematics instead of *finite* mathematics like derivatives and integrals instead of differences and sums.

3) The design problem in the non-linear approaches can be almost always adequately reduced to appropriate design problems arising from linearly parametrized responses of an experiment.

The optimality of designs for multivariate experiments are obtained, here, by a reduction to the associated univariate problems, which themselves may be of a complex structure involving qualitative and quantitative factors with different types of interactions. This means that, throughout the paper, the experimental region T is kept on the most general level including, in particular, experimental regions

$$T = \times_{k=1}^{K_1} \{1, ..., I_k\} \ \times \ C, \tag{3}$$

where C is a convex subset of $\mathbb{R}^{K_2}$. The latter is the most general representation of the most realistic experimental situation, in which K_1 qualitative factors are operating on levels $1, ..., I_k$, $k = 1, ..., K_1$, and K_2 quantitative factors are operating on levels in C. For those situations the whole optimum design theory has been developed by the authors and several other members of the research group in Berlin, and the results have been taken into account in the forthcoming *Lecture Notes* by the second author (Schwabe, 1995c) including a number of new and general results for the univariate case which, together with the present paper, opens a wide field of new applications.

Also, the main result by Krafft and Schaefer (1992) appears to be a variation of the reduction theme and can be regarded as a particular case of Theorem 5. Note that all the theorems of the present paper may be read as statements on the optimality of an *exact* design within the class of all *exact* designs and, alternatively, as statements on the optimality of a *general* design within the class of all those *general* designs. More detailed descriptions of these results are omitted because of the limited length of the paper.

2. Multivariate linear models under elementary designs

We consider a model for multivariate observations $X(t) = (X^{(1)}(t), ..., X^{(r)}(t))$ in which the mean response $\mu(t) = (\mu_1(t), ..., \mu_r(t))$ can be linearly parametrized for each component $\mu_\varrho(t) = E(X^{(\varrho)}(t))$, $\varrho = 1, ..., r$, in dependence on the setting $t \in T$ for the influential factors. More precisely,

$$X^{(\varrho)}(t) = a_\varrho(t)^T \beta_\varrho + Z^{(\varrho)}(t), \tag{4}$$

where $a_\varrho : T \to \mathbb{R}^{p_\varrho}$ and $\beta_\varrho \in \mathbb{R}^{p_\varrho}$ are the known regression functions resp. the unknown parameters for the ϱth component, and $Z(t) = (Z^{(1)}(t), ..., Z^{(r)}(t))$ is the random error associated with the observation $X(t)$, $E(Z(t)) = \mathbf{0}$. Hence, a multivariate observation $X(t)$ is described by

$$X(t) = a(t)^T \beta + Z(t), \tag{5}$$

where $\beta \in \mathbb{R}^p$, $p = \sum_{\varrho=1}^r p_\varrho$, is the vector of all unknown parameters

$$\beta = \begin{pmatrix} \beta_{(1)} \\ \beta_{(2)} \\ \vdots \\ \beta_{(r)} \end{pmatrix} \tag{6}$$

and $a(t)$ is block diagonal

$$a(t) = \begin{pmatrix} a_1(t) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & a_2(t) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & a_r(t) \end{pmatrix}. \tag{7}$$

We may obtain N different observations

$$X_n(t_n) = a(t_n)^T \beta + Z_n(t_n) \tag{8}$$

with components

$$X_n^{(\varrho)}(t_n) = a_\varrho(t_n)^T \beta_\varrho + Z_n^{(\varrho)}(t_n) \tag{9}$$

at N possibly different adjustments t_n of the factors of influence, $n = 1, ..., N$.

An elementary design d of size N is given as an N-dimensional vector $(t_1, ..., t_N)^T$ which contains the adjustments for the N different experiments.

With $X_d^{(\varrho)} = (X_1^{(\varrho)}(t_1), ..., X_N^{(\varrho)}(t_N))^T$ and $Z_d^{(\varrho)} = (Z_1^{(\varrho)}(t_1), ..., Z_N^{(\varrho)}(t_N))^T$ being the vectors of observations and errors, respectively, the observations for the ϱth component can be written in the usual matrix notation

$$X_d^{(\varrho)} = A_d^{(\varrho)}\beta_\varrho + Z_d^{(\varrho)} \tag{10}$$

where

$$A_d^{(\varrho)} = \begin{pmatrix} a_\varrho(t_1)^T \\ a_\varrho(t_2)^T \\ \vdots \\ a_\varrho(t_N)^T \end{pmatrix} \tag{11}$$

is the design matrix for the ϱth component.

In accordance with the common notations in multivariate analysis we vectorize the model and arrange its components successively

$$X_d = \begin{pmatrix} X_d^{(1)} \\ X_d^{(2)} \\ \vdots \\ X_d^{(r)} \end{pmatrix}, \qquad Z_d = \begin{pmatrix} Z_d^{(1)} \\ Z_d^{(2)} \\ \vdots \\ Z_d^{(r)} \end{pmatrix}. \tag{12}$$

Hence we obtain

$$X_d = A_d\beta + Z_d \tag{13}$$

where the design matrix A_d is block diagonal

$$A_d = \begin{pmatrix} A_d^{(1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & A_d^{(2)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & A_d^{(r)} \end{pmatrix}. \tag{14}$$

Note that a vectorization $\widetilde{X}_d = (X_1(t_1)^T, ..., X_N(t_N)^T)^T$ would be more appealing for design purposes. However, dealing with the rearranged vector $\widetilde{X}_d$ leads to considerably more complicated notations, in particular, in Section 3.

Next we assume that the underlying error structure is homogeneous, i.e. $cov(Z_n(t_n)) = \Sigma$, $n = 1, ..., N$, and that the observations are uncorrelated, i.e. $cov(Z_m(t_m), Z_n(t_n)) = \mathbf{0}$ for $m \neq n$. This results in a covariance structure $cov(Z_d) = \Sigma \otimes \mathbf{E}_N$ for the vectorized errors Z_d, where $\mathbf{E}_N$ is the $N \times N$ identity matrix and $\otimes$ denotes the Kronecker product. To avoid trivial cases we assume that $\Sigma > \mathbf{0}$ is positive definite.

In the linear model (13) the general (weighted) least squares estimator

$$\widehat{\beta}_d = (A_d^T(\Sigma \otimes \mathbf{E}_N)^{-1}A_d)^{-1}A_d^T(\Sigma \otimes \mathbf{E}_N)^{-1}X_d \tag{15}$$

is the best linear unbiased estimator of the unknown parameters β. We will consider only such designs for which β is linearly identifiable (estimable), i.e. for which A_d has full column rank. The quality of the estimator $\widehat{\beta}_d$ based on the design d and, hence, the quality of the design d itself is measured by the covariance matrix

$$cov(\widehat{\beta}_d) = (A_d^T(\Sigma \otimes \mathbf{E}_N)^{-1}A_d)^{-1}. \tag{16}$$

If we denote the entries of Σ^{-1} by $u_{\varrho\varrho'}$, $\varrho, \varrho' = 1, ..., r$, then the covariance matrix $cov(\widehat{\beta}_d)$ can be calculated more explicitly in terms of the design matrices associated with the components

$$cov(\widehat{\beta}_d) = \begin{pmatrix} u_{11}A_d^{(1)^T}A_d^{(1)} & u_{12}A_d^{(1)^T}A_d^{(2)} & \cdots & u_{1r}A_d^{(1)^T}A_d^{(r)} \\ u_{21}A_d^{(2)^T}A_d^{(1)} & u_{22}A_d^{(2)^T}A_d^{(2)} & \cdots & u_{2r}A_d^{(2)^T}A_d^{(r)} \\ \vdots & \vdots & \ddots & \vdots \\ u_{r1}A_d^{(r)^T}A_d^{(1)} & u_{r2}A_d^{(r)^T}A_d^{(2)} & \cdots & u_{rr}A_d^{(r)^T}A_d^{(r)} \end{pmatrix}^{-1} . \tag{17}$$

In general, it is not possible to minimize the covariance matrix $cov(\widehat{\beta}_d)$ uniformly, i.e. in the positive-definite sense. As a compromise we have to look at real valued functions of $cov(\widehat{\beta}_d)$, like the determinant, the trace, or the largest eigenvalue, which are isotonic. In particular, a design d^* will be called D- (A- resp. E-) optimum if it minimizes the determinant (the trace resp. the largest eigenvalue of $cov(\widehat{\beta}_d)$.

3. Homogeneous components

Additionally, in the present section, we assume that the mean responses μ_ϱ are modeled in the same way for each component, i.e. the regression functions a_ϱ for the components coincide: $a_\varrho = a_1$, $\varrho = 1, ..., r$. Note that still, in general, the unknown parameters β_ϱ and, hence, the mean responses will be different for the single components. In this situation also the componentwise design matrices $A_d^{(\varrho)}$ coincide, and both the regression function a and the design matrix A_d factorize according to (cf (7) and (14))

$$a(t) = \mathbf{E}_r \otimes a_1(t) \tag{18}$$

and

$$A_d = \mathbf{E}_r \otimes A_d^{(1)}. \tag{19}$$

Because of the equi-modelling of the components we can calculate the general least squares estimator (15) by using (19) as

$$\begin{aligned}\widehat{\beta}_d &= (\Sigma^{-1} \otimes A_d^{(1)^T} A_d^{(1)})^{-1} (\Sigma^{-1} \otimes A_d^{(1)^T}) X_d \\ &= (\mathbf{E}_r \otimes (A_d^{(1)^T} A_d^{(1)})^{-1} A_d^{(1)^T}) X_d \end{aligned} \tag{20}$$

From (20) we see that the knowledge of Σ is not required for the calculation of $\widehat{\beta}_d$ and that the general least squares estimator coincides with the ordinary least squares estimator

$$\widehat{\beta}_d^{ols} = (A_d^T A_d)^{-1} A_d^T X_d. \tag{21}$$

Moreover, the covariance matrix (16) factorizes

$$cov(\widehat{\beta}_d) = \Sigma \otimes (A_d^{(1)^T} A_d^{(1)})^{-1}. \tag{22}$$

According to the Kronecker product structure (22) the eigenvalues of the covariance matrix $cov(\widehat{\beta}_d)$ factorize, i.e. let λ_Σ be the vector of eigenvalues of Σ, $\lambda_d^{(1)}$ the vector of eigenvalues of $(A_d^{(1)^T} A_d^{(1)})^{-1}$, and λ_d the vector of eigenvalues of $cov(\widehat{\beta}_d)$, then

$$\lambda_d = \lambda_\Sigma \otimes \lambda_d^{(1)}. \tag{23}$$

In particular, we obtain for the determinant

$$\det(cov(\widehat{\beta}_d)) = \det(\Sigma)^{p_1} \det(A_d^{(1)^T} A_d^{(1)})^{-r}, \tag{24}$$

for the trace

$$\mathrm{tr}(cov(\widehat{\beta}_d)) = \mathrm{tr}(\Sigma) \cdot \mathrm{tr}((A_d^{(1)^T} A_d^{(1)})^{-1}) \tag{25}$$

and for the largest eigenvalue

$$\lambda_{\max}(cov(\widehat{\beta}_d)) = \lambda_{\max}(\Sigma) \lambda_{\max}((A_d^{(1)^T} A_d^{(1)})^{-1}). \tag{26}$$

Hence the optimization with respect to the usual D- (A- resp. E-) criterion reduces to the corresponding univariate optimization problem.

THEOREM 1. *If d^* is D- (A- resp. E-) optimum in the univariate model with mean response $\mu_1(t) = a_1(t)\beta_1$, then d^* is D- (A- resp. E-) optimum in the multivariate model with homogeneous components $\mu_\varrho(t) = a_1(t)\beta_\varrho$.*

Remark 1. Because of (23) the result of Theorem 1 can be extended to the whole class of Φ_q-criteria based on the eigenvalues of the covariance matrix (for a definition see e. g. Pázman, 1986).

Remark 2. Further generalizations are straightforward if we are interested in a linear aspect ψ of the unknown parameters β which is of the form $\psi(\beta) = (L_0 \otimes L_1)\beta$ where $\psi_1(\beta_1) = L_1\beta_1$ is the corresponding linear aspect in the univariate model, $L_1 \in \mathbb{R}^{s_1 \times p_1}$ and $L_0 \in \mathbb{R}^{s_0 \times r}$. Particular cases are covered by $L_0 = \mathbf{E}_r$, if we are interested in the same linear aspect for each component, and by L_0 with zero row sums, if we are interested in differences between the components.

Finally, we want to mention that in the particular situation of the present section the model can be written in a more appealing way

$$\mathbf{X}_d = A_d^{(1)} B + \mathbf{Z}_d, \tag{27}$$

where $\mathbf{X}_d = (X_d^{(1)}, ..., X_d^{(r)})$, $\mathbf{Z}_d = (Z_d^{(1)}, ..., Z_d^{(r)})$ and $B = (\beta_1, ..., \beta_r)$, such that each row of (27) represents an r-dimensional observation $X_n(t_n)$. However, the form (27) is not convenient for design considerations because, then, we have to deal with a matrix shape B of parameters rather than the vectorized β.

4. Uncorrelated components

If the components of the multivariate observations do not influence each other then the covariance matrix Σ of a single observation is diagonal

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r^2 \end{pmatrix}. \tag{28}$$

The induced covariance structure $\Sigma \otimes \mathbf{E}_N$ of the whole observational vector X_d is also diagonal

$$\Sigma \otimes \mathbf{E}_N = \begin{pmatrix} \sigma_1^2\mathbf{E}_N & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_2^2\mathbf{E}_N & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sigma_r^2\mathbf{E}_N \end{pmatrix}. \tag{29}$$

Using the block diagonal structure (7) of the design matrix A_d we obtain from (17) that the covariance matrix of the general least squares estimator becomes block diagonal

$$cov(\widehat{\beta}_d) = \begin{pmatrix} \sigma_1^2 (A_d^{(1)^T} A_d^{(1)})^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 (A_d^{(2)^T} A_d^{(2)})^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sigma_r^2 (A_d^{(r)^T} A_d^{(r)})^{-1} \end{pmatrix}. \tag{30}$$

Moreover, the least squares estimator $\widehat{\beta}_d$ can be written as

$$\widehat{\beta}_d = \begin{pmatrix} (A_d^{(1)^T} A_d^{(1)})^{-1} A_d^{(1)^T} X_d^{(1)} \\ (A_d^{(2)^T} A_d^{(2)})^{-1} A_d^{(2)^T} X_d^{(2)} \\ \vdots \\ (A_d^{(r)^T} A_d^{(r)})^{-1} A_d^{(r)^T} X_d^{(r)} \end{pmatrix}. \tag{31}$$

Again the calculation of $\widehat{\beta}_d$ does not require any further knowledge of the diagonal entries σ_ϱ^2 of Σ and the general least squares estimator coincides with the ordinary $\widehat{\beta}_d^{ols} = (A_d^T A_d)^{-1} A_d^T X_d$.

Denote by $\lambda_d^{(\varrho)}$ the vector of eigenvalues of $(A_d^{(\varrho)^T} A_d^{(\varrho)})^{-1}$. Then the vector λ_d of eigenvalues of $cov(\widehat{\beta}_d)$ can be partitioned according to

$$\lambda_d = \begin{pmatrix} \sigma_1^2 \lambda_d^{(1)} \\ \sigma_2^2 \lambda_d^{(2)} \\ \vdots \\ \sigma_r^2 \lambda_d^{(r)} \end{pmatrix}. \tag{32}$$

A design is D-optimum if it minimizes $\prod_{\varrho=1}^r \det(A_d^{(\varrho)^T} A_d^{(\varrho)})^{-1}$, irrespectively of the actual values of $\sigma_1^2, ..., \sigma_r^2$. Alternatively, an A-optimum design minimizes $\sum_{\varrho=1}^r \sigma_\varrho^2 \operatorname{tr}((A_d^{(\varrho)^T} A_d^{(\varrho)})^{-1})$ and an E-optimum design aims at minimizing $\max_{\varrho=1,\ldots,r} \sigma_\varrho^2 \lambda_{\max}((A_d^{(\varrho)^T} A_d^{(\varrho)})^{-1})$, respectively.

In particular, if a design is simultaneously optimum for all components then it solves also the corresponding multivariate optimization problem.

THEOREM 2. *If d^* is D- (A- resp. E-) optimum in each univariate model with mean response $\mu_\varrho(t) = a_\varrho(t)\beta_\varrho$, then d^* is D- (A- resp. E-) optimum in the multivariate model with uncorrelated components.*

5. Components with increasing complexity

In general situations the underlying models may differ for the single components, e. g. regressions of different precisions may be fitted, the components may be affected by different sets of factors of influence, or different interaction structures may occur. In the setting of increasing complexity we assume that the components can be rearranged in such a way that the regression functions a_ϱ are included in each other in ascending order

$$a_1 = f_1, \qquad a_2 = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \qquad \dots \qquad , a_r = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_r \end{pmatrix}, \tag{33}$$

$f_\varrho : T \to I\!R^{q_\varrho}$, $q_\varrho = p_\varrho - p_{\varrho-1}$, where some of the f_ϱ might be empty ($q_\varrho = 0$). Now, for a fixed design d, we can reparametrize the models for each component by orthogonalization with respect to its predecessor.

Thus we define $\widetilde{f}_1 = f_1$, $\widetilde{f}_\varrho = f_\varrho - A_{d,\varrho}^T A_d^{(\varrho-1)} (A_d^{(\varrho-1)^T} A_d^{(\varrho-1)})^{-1} a_{\varrho-1}$, where $A_{d,\varrho}^T = (f_\varrho(t_1), ..., f_\varrho(t_N))$, and $\widetilde{a}_\varrho = (\widetilde{f}_1^T, ..., \widetilde{f}_\varrho^T)^T$. Then the components of the corresponding univariate design matrices $\widetilde{A}_d^{(\varrho)}$, which are defined by (11), are orthogonal, i. e.

$$\widetilde{A}_{d,\varrho}^T \widetilde{A}_{d,\varrho'} = \mathbf{0}, \tag{34}$$

$\varrho \neq \varrho'$, where, again, $\widetilde{A}_{d,\varrho}^T = (\widetilde{f}_\varrho(t_1), ..., \widetilde{f}_\varrho(t_N))$. Moreover, the design matrices are related by linear transformations $\widetilde{A}_d^{(\varrho)^T} = L_\varrho A_d^{(\varrho)^T}$ and $\widetilde{A}_d^T = L A_d^T$, where L_ϱ and L are lower triangular matrices with all their diagonal entries equal to one (see Appendix). Next, we rearrange the regression functions in the multivariate model according to the components $\widetilde{f}_\varrho$ and, hence, permute the parameters. Let

$$\bar{a} = \begin{pmatrix} & \mathbf{E}_r \otimes \widetilde{f}_1 & & & \\ \mathbf{0} & & \mathbf{E}_{r-1} \otimes \widetilde{f}_2 & & \\ \vdots & \ddots & & \ddots & \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{E}_{r-\varrho+1} \otimes \widetilde{f}_\varrho & \\ \vdots & & & \ddots & \ddots \\ \mathbf{0} & & \cdots & \mathbf{0} & \mathbf{E}_r \otimes \widetilde{f}_1 \end{pmatrix}, \tag{35}$$

then $\bar{a} = Qa$ for some permutation matrix Q. The design matrices $\bar{A}_d$ are defined in analogy to (14)

$$\bar{A}_d = \begin{pmatrix} \bar{a}_1(t_1)^T \\ \vdots \\ \bar{a}_1(t_N)^T \\ \bar{a}_2(t_1)^T \\ \vdots \\ \bar{a}_r(t_N)^T \end{pmatrix}, \tag{36}$$

where $\bar{a}_\varrho$ is the ϱth component of $\bar{a}$. As $\det(L) = |\det(Q)| = 1$ we obtain

$$\det(cov(\widehat{\beta}_d)) = \det(\widetilde{A}_d^T(\Sigma^{-1} \otimes \mathbf{E}_N)\widetilde{A}_d)^{-1} = \det(\bar{A}_d^T(\Sigma^{-1} \otimes \mathbf{E}_N)\bar{A}_d)^{-1} \tag{37}$$

by (16). Denote by $U_\varrho = (u_{\varrho'\varrho''})_{\varrho'=\varrho,\dots,r}^{\varrho''=\varrho,\dots,r}$ that part of the inverse covariance matrix Σ^{-1} associated with the components $\varrho, ..., r$ in which the regression functions f_ϱ are present, $U_\varrho = (\mathbf{0}, \mathbf{E}_{r-\varrho+1})\Sigma^{-1}(\mathbf{0}, \mathbf{E}_{r-\varrho+1})^T$, $U_1 = \Sigma^{-1}$. Then

$$\bar{A}_d^T(\Sigma^{-1} \otimes \mathbf{E}_N)\bar{A}_d = \begin{pmatrix} U_1 \otimes \widetilde{A}_{d,1}^T\widetilde{A}_{d,1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & U_2 \otimes \widetilde{A}_{d,2}^T\widetilde{A}_{d,2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & U_r \otimes \widetilde{A}_{d,r}^T\widetilde{A}_{d,r} \end{pmatrix} \tag{38}$$

is block diagonal and, hence,

$$\det(cov(\widehat{\beta}_d)) = \left(\prod_{\varrho=1}^{r} \det(U_\varrho)^{-q_\varrho}\right)\left(\prod_{\varrho=1}^{r} \det(\widetilde{A}_{d,\varrho}^T\widetilde{A}_{d,\varrho})^{-(r-\varrho+1)}\right) \tag{39}$$

in view of (37). Note that, by convention, $\det(\widetilde{A}_{d,\varrho}^T\widetilde{A}_{d,\varrho}) = 1$ if $q_\varrho = 0$. By the orthogonality (34) of $\widetilde{A}_{d,\varrho}$ and as $\det(L_\varrho) = 1$ we have $\det(A_d^{(\varrho)^T} A_d^{(\varrho)}) = \det(\widetilde{A}_d^{(\varrho)^T} \widetilde{A}_d^{(\varrho)}) = \prod_{\varrho'=1}^{\varrho} \det(\widetilde{A}_{d,\varrho'}^T\widetilde{A}_{d,\varrho'})$ so that, finally,

$$\det(cov(\widehat{\beta}_d)) = \left(\prod_{\varrho=1}^{r} \det(U_\varrho)^{-q_\varrho}\right)\left(\prod_{\varrho=1}^{r} \det(A_d^{(\varrho)^T} A_d^{(\varrho)})^{-1}\right). \tag{40}$$

Note that formula (40) is the main result obtained by Krafft and Schaefer (1992). Now, the determinant of the covariance matrix is minimized if the determiant of the covariance matrices $(A_d^{(\varrho)^T} A_d^{(\varrho)})^{-1}$ can be minimized simultaneously for each component.

As reparametrizations do not affect the criterion of D-optimality the concept of increasing complexity can be paraphrased as $a_\varrho = C_\varrho a_{\varrho+1}$ for some $p_\varrho \times p_{\varrho+1}$ matrix C_ϱ, $\varrho = 1, ..., r-1$.

THEOREM 3. *If d^* is D-optimum in each univariate model with mean response $\mu_\varrho(t) = a_\varrho(t)\beta_\varrho$, $a_\varrho = C_\varrho a_{\varrho+1}$, then d^* is D-optimum in the multivariate model with increasing complexity.*

For models with increasing complexity, in which a design d^* exists which is simultaneously D-optimum for all components, particular examples are given by (i) polynomial regression of degrees zero and one, (ii) trigonometric regression of different degrees up to degree $(N-1)/2$, and (iii) K-factor models with different depth of interactions (see Schwabe, 1995a).

6. Components with mutually exclusive complexity

In the general setting, that the regression functions are not necessarily included in each other, further assumptions have to be made on the optimum design d^*. Essentially, optimality is also required for augmented models which are formed by the regression functions associated with any pair of components. This situation will be exhibited for a bivariate model, first.

In general, a pair of regression functions a_1 and a_2 can be reparametrized in such a way that there is a common regression vector f_0 and there are mutually different components f_1 and f_2 such that the set of entries in $(f_0^T, f_1^T, f_2^T)^T$ is linearly independent on T,

$$a_1 = \begin{pmatrix} f_0 \\ f_1 \end{pmatrix}, \qquad a_2 = \begin{pmatrix} f_0 \\ f_2 \end{pmatrix}. \tag{41}$$

As in the previous section we orthogonalize f_1 and f_2 with respect to $f_0 : T \to \mathbb{R}^{p_0}$ and a fixed design d, $\widetilde{f}_\varrho = f_\varrho - A_{d,\varrho}^T A_{d,0}(A_{d,0}^T A_{d,0})^{-1} f_0$, $\varrho = 1, 2$, where $A_{d,\varrho'}^T = (f_{\varrho'}(t_1), ..., f_{\varrho'}(t_N))$, $\varrho' = 0, 1, 2$. Again, the corresponding transformation matrices are lower diagonal with all diagonal elements equal to one. Next we rearrange the regression functions similar to (35)

$$\bar{a} = \begin{pmatrix} \multicolumn{2}{c}{\mathbf{E}_2 \otimes f_0} \\ \widetilde{f}_1 & \mathbf{0} \\ \mathbf{0} & \widetilde{f}_2 \end{pmatrix}. \tag{42}$$

We define $\bar{A}_d$ as in (36) and obtain $\det(cov(\widehat{\beta}_d)) = \det(\bar{A}_d^T(\Sigma^{-1} \otimes \mathbf{E}_N)\bar{A}_d)^{-1}$

(cf (39)) and

$$\bar{A}_d^T(\Sigma^{-1} \otimes \mathbf{E}_N)\bar{A}_d = \begin{pmatrix} \Sigma^{-1} \otimes A_{d,0}^T A_{d,0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & u_{11}\widetilde{A}_{d,1}^T \widetilde{A}_{d,1} & u_{12}\widetilde{A}_{d,1}^T \widetilde{A}_{d,2} \\ \mathbf{0} & u_{12}\widetilde{A}_{d,2}^T \widetilde{A}_{d,1} & u_{22}\widetilde{A}_{d,2}^T \widetilde{A}_{d,2} \end{pmatrix} \quad (43)$$

similar to (38). Combining these results we get

$$\det(cov(\widehat{\beta}_d)) \geq \det(\Sigma)^{p_0} u_{11}^{-(p_1-p_0)} u_{22}^{-(p_2-p_0)} \prod_{\varrho=1}^{2} \det(A_d^{(\varrho)^T} A_d^{(\varrho)})^{-1} \quad (44)$$

with equality if, and only if, either the components are uncorrelated, $u_{12} = 0$, or $\widetilde{f}_1$ and $\widetilde{f}_2$ are orthogonal with respect to the design d, $\widetilde{A}_{d,1}^T \widetilde{A}_{d,2} = \mathbf{0}$. The latter condition can equivalently be formulated as a factorization for f_1 and f_2 adjusted for f_0, $A_{d,1}^T A_{d,2} = A_{d,1}^T A_{d,0}(A_{d,0}^T A_{d,0})^{-1} A_{d,0}^T A_{d,2}$.

THEOREM 4. *If d^* is D-optimum in each univariate model with mean response $\mu_\varrho(t) = a_\varrho(t)\beta_\varrho$, $a_\varrho = (f_0^T, f_\varrho^T)^T$, $(f_0^T, f_1^T, f_2^T)^T$ linearly independent, and if $A_{d^*,1}^T A_{d^*,2} = A_{d^*,1}^T A_{d^*,0}(A_{d^*,0}^T A_{d^*,0})^{-1} A_{d^*,0}^T A_{d^*,2}$, then d^* is D-optimum in the bivariate model with mutual different complexity.*

If f_1 is omitted in Theorem 4 the factorization condition vanishes and we recover the bivariate version of Theorem 3.

Additional models which are covered by Theorem 4 include trigonometric regression, in case not all lower order frequencies are present, and K-factor models, in which the components are influenced by different factors or in which interactions between different factors are active.

For the special case that only the constant term is in common, $f_0 = 1$, the factorization condition of Theorem 4 simplifies to $\frac{1}{N}\sum_{n=1}^{N} f_1(t_n) f_2(t_n)^T = (\frac{1}{N}\sum_{n=1}^{N} f_1(t_n))(\frac{1}{N}\sum_{n=1}^{N} f_2(t_n))^T$. Moreover, in this case, the factorization condition implies that d^* is also D-optimum for the augmented univariate model with mean response $\mu_{1,2}(t) = \beta_0 + f_1(t)^T\beta_1 + f_2(t)^T\beta_2$.

Finally, if f_0 is omitted in Theorem 4, then the orthogonality of f_1 and f_2, themselves, is required, $A_{d^*,1}^T A_{d^*,2} = \mathbf{0}$. Also in this case the factorization condition implies the D-optimality of d^* for the corresponding augmented univariate model with mean response $\mu_{1,2}(t) = f_1(t)^T\beta_1 + f_2(t)^T\beta_2$.

For components with simultaneously common regression part f_0 the derivation of Theorem 4 can be extended to more than two components in a straightforward manner to obtain a result as follows.

COROLLARY 1. *If d^* is D-optimum in each univariate model with mean response $\mu_\varrho(t) = a_\varrho(t)\beta_\varrho$, $a_\varrho = (f_0^T, f_\varrho^T)^T$, $(f_0^T, f_1^T, ..., f_r^T)^T$ linearly indepen-*

dent, and if $A_{d^*,\varrho}^T A_{d^*,\varrho'} = A_{d^*,\varrho}^T A_{d^*,0}(A_{d^*,0}^T A_{d^*,0})^{-1} A_{d^*,0}^T A_{d^*,\varrho'}$, $\varrho \neq \varrho'$, *then* d^* *is D-optimum in the multivariate model with mutual different complexity.*

We note that the factorization condition can be equivalently rewritten as $A_{d^*}^{(\varrho)^T} A_{d^*}^{(\varrho')} = A_{d^*}^{(\varrho)^T} A_{d^*,0}(A_{d^*,0}^T A_{d^*,0})^{-1} A_{d^*,0}^T A_{d^*}^{(\varrho')}$. In the general case factorization conditions are requested for the regression functions adjusted to their respective common part $f_{\varrho,\varrho'}$. To simplify notations we write (a_ϱ) for the set $\{a_{\varrho 1}, ..., a_{\varrho p_\varrho}\}$ of all entries (functions) in a_ϱ, and $(f_{\varrho,\varrho'})$ analogously.

THEOREM 5. *If* d^* *is D-optimum in each univariate model with mean response* $\mu_\varrho(t) = a_\varrho(t)\beta_\varrho$, *and if for every pair* ϱ *and* ϱ', *the sets of functions* $(f_{\varrho,\varrho'}) = (a_\varrho) \cap (a_{\varrho'})$, $(a_\varrho) \setminus (f_{\varrho,\varrho'})$ *and* $(a_{\varrho'}) \setminus (f_{\varrho,\varrho'})$ *are linearly independent and the factorization* $A_{d^*}^{(\varrho)^T} A_{d^*}^{(\varrho')} = A_{d^*}^{(\varrho)^T} A_{d^*,\varrho,\varrho'}(A_{d^*,\varrho,\varrho'}^T A_{d^*,\varrho,\varrho'})^{-1} A_{d^*,\varrho,\varrho'}^T A_{d^*}^{(\varrho')}$ *holds, then* d^* *is D-optimum in the multivariate model with mutual different complexity.*

The proof can be performed by a more sophisticated rearrangement of the regression functions. We mention that, as in the bivariate case, Theorem 3 follows from Theorem 5 because the factorization condition vanishes for components with increasing complexity.

7. Applications

Within the classical terrain of experimental design the so-called *analysis of variance* models, i.e. experiments influenced by a number of qualitative factors, play an important role. For such experiments the most general region of experimental conditions is given by $T = \times_{k=1}^K \{1, ..., I_k\}$ defining K qualitative factors, with the kth one operating at I_k possible levels $1, ..., I_k$, $k = 1, ..., K$, and with different types and orders of interactions between subsets of the K factors.

The second classical approach in experimental design is concerned with the so-called *regression* models, i.e. experiments influenced by a number of quantitative factors. For such experiments the most general region of experimental conditions is given by a convex subset $T = C$ of $\mathbb{R}^K$ defining K quantitative factors operating at all possible levels $t \in T$ and, again, with different types and orders of interactions between subsets of the K factors.

The specification of these two types of experiments gives rise to a whole class of possible linear parametrizations for univariate linear models in which both kinds of factors, qualitative and quantitative ones, are active and may interact with each other in various kinds.

In the multivariate case each component μ_ϱ of the multivariate response μ may have, in general, a different type of parametrization according to the number of factors actually affecting the components and different types of interactions within the classes of factors involved in the individual components. So the main Theorem 5 provides the adequate tool for such realistic cases to solve the design problem for multivariate experiments by solving the associated design problems in the setup of the individual components.

To show only part of the huge variety of parametrizations which fall under the conditions of Theorem 5 and its specialization to hierarchical models (Theorem 3) we present the following general classes of examples:

EXAMPLE 1. If different factors are affecting different sets of factors this can be generally described by responses $\mu_\varrho(t_1, ..., t_r) = \beta_{\varrho,0} + f_\varrho(t_\varrho)^T \beta_{\varrho,\varrho}$, $\varrho = 1, ..., r$, where $t = (t_1, ..., t_r) \in T = \times_{\varrho=1}^r T_\varrho$, and T_ϱ itself can be a K_ϱ-dimensional set defining K_ϱ factors of influence.

To be more specific, in an analysis of variance setting f_ϱ may describe any analysis of variance model in K_ϱ factors operating on $T_\varrho = \times_{k=1}^{K_\varrho} \{1, ..., I_{\varrho,k}\}$. In a regression setting f_ϱ may describe polynomial or trigonometric regression in K_ϱ factors operating on a convex subset C_ϱ of $\mathbb{R}^{K_\varrho}$.

EXAMPLE 2. The situation that different interaction structures are present for the individual components leads for the particular case of two factors of influence to a hierarchical model. In the analysis of variance setting there are essentially two types of different responses $\mu_1(i,j) = \alpha_{1,i} + \beta_{1,j}$ without interactions and $\mu_2(i,j) = \alpha_{2,i} + \beta_{2,j} + \gamma_{2,ij}$ with interactions, respectively. For more than two factors more complicated interaction structures may arise, e. g. in the simplest case responses $\mu_1(i_1, i_2, i_3) = \alpha_{1,i_1}^{(1)} + \alpha_{1,i_2}^{(2)} + \alpha_{1,i_3}^{(3)} + \gamma_{1,i_1 i_2}^{(1,2)}$ and $\mu_2(i_1, i_2, i_3) = \alpha_{2,i_1}^{(1)} + \alpha_{2,i_2}^{(2)} + \alpha_{2,i_3}^{(3)} + \gamma_{2,i_1,i_3}^{(1,3)}$ with interactions between the first and second resp. the first and third factor only.

EXAMPLE 3. Another typical example for hierarchical model is obtained in trigonometric regression, where the response is modeled by a Fourier expansion $\mu_\varrho(t) = \beta_{\varrho,0} + \sum_{m=1}^{M_\varrho} (\beta_{\varrho,2m-1} \sin(2\pi m t) + \beta_{\varrho,2m} \cos(2\pi m t))$ up to order M_ϱ on $t \in T = [0, 1)$. In more than one dimension also partial interaction structures can be considered like complete and partial first and second order expansions.

Also the still larger class of more complex, realistic multivariate experiments where both kinds of factors, qualitative and quantitative ones, are simultaneously affecting the response can be treated by Theorem 5 using our most recent results for the univariate case, as exhibited in the papers by Kurotschka (1984, 1988) and Schwabe (1995a, 1995b) and the forthcoming lecture notes (Schwabe, 1995c).

8. General designs

The information matrix $\mathbf{I}(d)$ of an elementary design d is defined as the inverse of the corresponding covariance matrix $cov(\widehat{\beta}_d)$

$$\mathbf{I}(d) = A_d^T(\Sigma \otimes \mathbf{E}_N)^{-1}A_d. \tag{45}$$

By rearranging terms it can be seen that the information matrix is the sum of the *information* contained in each single setting

$$\mathbf{I}(d) = \sum_{n=1}^{N} a(t_n)\Sigma^{-1}a(t_n)^T. \tag{46}$$

If there are only few different settings and, hence, many replications in an elementary design d it is more convenient to write

$$d = \begin{pmatrix} t_1 & t_2 & \cdots & t_J \\ N_1 & N_2 & \cdots & N_J \end{pmatrix} \tag{47}$$

where the settings t_j are mutually different and N_j is the number of replications at the setting t_j, $\sum_{j=1}^{J} N_j = N$. Hence we obtain for the information matrix

$$\mathbf{I}(d) = \sum_{j=1}^{J} N_j a(t_j)\Sigma^{-1}a(t_j)^T. \tag{48}$$

As the sample size N is fixed, an elementary design d is determined by its relative frequencies $w_j = \frac{N_j}{N}$ at the actual settings. Thus d may be identified with a dicrete measure δ_d on a finite support, which assigns weights w_j to its support points t_j.

To solve the integer optimization problem for N_j is often a difficult task. Therefore it is reasonable to embed the optimization problem in a continuous setting and to drop the requirement of w_j to be an integer multiple of $\frac{1}{N}$. Hence, as a *general design* δ we define any measure on T, normalized to one, which is concentrated on a finite number of supporting points. In accordance with (48) we define the information matrix

$$\mathbf{I}(\delta) = \int a(t)\Sigma^{-1}a(t)^T\,\delta(dt), \tag{49}$$

such that, for any elementary design d of sample size N, $\mathbf{I}(\delta_d) = \frac{1}{N}\mathbf{I}(d)$ is a normalized version of the information matrix $\mathbf{I}(d)$.

Besides this traditional extension of the design problem to general designs, in order to embed the optimization problem into a richer set which is convex, a more substantial reason for considering general designs is evident. Because the supposition of a prespecified number of units with prespecified variation is sometimes rather artificial a more general idea should be associated with the concept of general designs (see Kurotschka, 1988, or Schwabe, 1995b). Assume that we may observe at certain design points $t_1, \dots, t_J$ with intensity proportional to $w_j > 0$, $\sum_{j=1}^J w_j = 1$, i.e. $cov(X_j(t_j)) = w_j^{-1}\Sigma$, $j = 1, \dots, J$, then in this general multivariate linear model the best linear unbiased estimator is given by the weighted least squares estimator $\widehat{\beta}_\delta = (A_\delta^T(\Sigma^{-1} \otimes W_\delta)A_\delta)^{-1}A_\delta^T(\Sigma^{-1} \otimes W_\delta)X_\delta$ where

$$X_\delta = \begin{pmatrix} X_\delta^{(1)} \\ X_\delta^{(2)} \\ \vdots \\ X_\delta^{(r)} \end{pmatrix}, \tag{50}$$

$X_\delta^{(\varrho)} = (X_j^{(\varrho)}(t_j))_{j=1,\dots,J}$, are the vectorized observations, $cov(X_\delta) = \Sigma \otimes W_\delta^{-1}$,

$$A_\delta = \begin{pmatrix} A_\delta^{(1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & A_\delta^{(2)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & A_\delta^{(r)} \end{pmatrix}, \tag{51}$$

$A_\delta^{(\varrho)} = (a_\varrho(t_1), ..., a_\varrho(t_J))^T$, is the design matrix associated with the supporting points t_j of the design δ and the diagonal matrix

$$W_\delta = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_J \end{pmatrix} \tag{52}$$

is the intensity matrix of the observations. We notice that

$$\mathbf{I}(\delta) = A_\delta^T(\Sigma^{-1} \otimes W_\delta)A_\delta = \sum_{j=1}^J w_j a(t_j)\Sigma^{-1}a(t_j)^T, \tag{53}$$

where $cov(\widehat{\beta}_\delta) = \mathbf{I}(\delta)^{-1}$. Hence a general design δ which assigns weights w_j to the supporting points t_j, $j = 1, \ldots, J$, has to be understood as an instruction which demands that J different observations are to be made at the design points t_j with intensity proportional to w_j, $j = 1, \ldots, J$.

A general design δ^* is called D- (A- resp. E-) optimum if it minimizes the determinant (the trace resp. the largest eigenvalue of the inverse $\mathbf{I}(\delta)^{-1}$ of the information matrix. Note that $\mathbf{I}(\delta_d)^{-1}$ coincides with the covariance matrix $cov(\widehat{\beta}_d)$ for an elementary design d with sample size N, up to a multiplicative constant $\frac{1}{N}$, and hence the generalizations of the criteria are straightforward.

All the results obtained for elementary fixed sample size designs d in the preceding sections are also valid for general designs δ^* and are formulated next.

THEOREM 1*. *If δ^* is D- (A- resp. E-) optimum in the univariate model with mean response $\mu_1(t) = a_1(t)\beta_1$, then δ^* is D- (A- resp. E-) optimum in the multivariate model with homogeneous components $\mu_\varrho(t) = a_1(t)\beta_\varrho$.*

Note that this simple theorem, restricted to D-optimality, is the main result of Chang (1994) obtained by using such demanding tools as the equivalence theorem (see Theorem 6 below).

THEOREM 2*. *If δ^* is D- (A- resp. E-) optimum in each univariate model with mean response $\mu_\varrho(t) = a_\varrho(t)\beta_\varrho$, then δ^* is D- (A- resp. E-) optimum in the multivariate model with uncorrelated components.*

THEOREM 3*. *If δ^* is D-optimum in each univariate model with mean response $\mu_\varrho(t) = a_\varrho(t)\beta_\varrho$, $a_\varrho = C_\varrho a_{\varrho+1}$, then δ^* is D-optimum in the multivariate model with increasing complexity.*

THEOREM 4*. *If δ^* is D-optimum in each univariate model with mean response $\mu_\varrho(t) = a_\varrho(t)\beta_\varrho$, $a_\varrho = (f_0^T, f_\varrho^T)^T$, $(f_0^T, f_1^T, f_2^T)^T$ linearly independent, and if $\int f_1 f_2^T \, d\delta^* = \int f_1 f_0^T \, d\delta^* (\int f_0 f_0^T \, d\delta^*)^{-1} \int f_0 f_2^T \, d\delta^*$, then δ^* is D-optimum in the bivariate model with mutual different complexity.*

For the bivariate case we consider the augmented univariate model with mean response $\mu_{1,2}(t) = f_0(t)^T\beta_0 + f_1(t)^T\beta_1 + f_2(t)^T\beta_2$, such that the corresponding regression function $a_{1,2} = (f_0^T, f_1^T, f_2^T)^T$ collects all functions occuring in the univariate components. If $f_0 = 1$ or if f_0 is not present in the model then for the information matrix $\mathbf{I}_{1,2}(\delta) = \int a_{1,2} a_{1,2}^T \, d\delta$ it is straightforward by a refinement argument that its determinant is dominated by the information contained in the univariate component models

$$\begin{aligned} \det(\mathbf{I}_{1,2}(\delta)) &\leq \det(\textstyle\int \widetilde{f}_1 \widetilde{f}_1^T \, d\delta)^{-1} \det(\textstyle\int \widetilde{f}_2 \widetilde{f}_2^T \, d\delta)^{-1} \\ &\leq \det(\textstyle\int \widetilde{f}_1 \widetilde{f}_1^T \, d\delta^*)^{-1} \det(\textstyle\int \widetilde{f}_2 \widetilde{f}_2^T \, d\delta^*)^{-1}. \end{aligned} \tag{54}$$

Hence, the factorization condition of Theorem 4* implies the D-optimality of δ^* in the augmented model. In the next section we will establish the result that the additional D-optimality of δ^* in the augmented univariate model is, indeed, necessary for the D-optimality in the bivariate model for general covariance structure Σ^{-1}.

THEOREM 5*. *If δ^* is D-optimum in each univariate model with mean response $\mu_\varrho(t) = a_\varrho(t)\beta_\varrho$, and if for every pair ϱ and ϱ', the sets of functions $(f_{\varrho,\varrho'}) = (a_\varrho) \cap (a_{\varrho'})$, $(a_\varrho) \setminus (f_{\varrho,\varrho'})$ and $(a_{\varrho'}) \setminus (f_{\varrho,\varrho'})$ are linearly independent and the factorization $\int a_\varrho a_{\varrho'}^T \, d\delta^* = \int a_\varrho f_{\varrho,\varrho'}^T \, d\delta^* (\int f_{\varrho,\varrho'} f_{\varrho,\varrho'}^T \, d\delta^*)^{-1} \int f_{\varrho,\varrho'} a_{\varrho'}^T \, d\delta^*$ holds, then δ^* is D-optimum in the multivariate model with mutual different complexity.*

Finally, we mention that a useful tool for checking a general design δ^* to be D-optimum in a multivariate model is the following equivalence theorem (see Fedorov, 1972, p. 212) of which we will make use in the subsequent section.

THEOREM 6.
(i) *A design δ^* is D-optimum if, and only if,* $\mathrm{tr}(\Sigma^{-1} a(t) \mathbf{I}(\delta^*)^{-1} a(t)^T) \leq p$, *for every $t \in T$.*
(ii) *If δ^* is D-optimum, then* $\delta^*(\mathrm{tr}(\Sigma^{-1} a(t) \mathbf{I}(\delta^*)^{-1} a(t)^T) = p) = 1$.

9. A counterexample and its theoretical background

Although the results of the preceding sections seem to include all practically relevant general parametrizations for multivariate linear experiments the following example, which is more of mathematical interest, shows that the generally accepted opinion is misleading that factorization is always possible. This is caused by the fact that the concept of orthogonalization, which is a universally useful tool in the design of experiments, may conflict with the meaning of the parameters if applied across the components of a multivariate experiment:

To keep notations as simple as possible we consider a bivariate model with different one-parameter univariate components for which the same design $\delta_1^* = \delta_2^*$ is optimum. In particular, we will deal with a purely linear component,

$$X^{(1)}(t) = \beta_1 t + Z^{(1)}, \tag{55}$$

and with a purely quadratic component,

$$X^{(2)}(t) = \beta_2 t^2 + Z^{(2)}, \tag{56}$$

on the design region $T = [\tau, 1]$, for various $-1 \leq \tau < 1$. For $\tau > -1$ the unique optimum design for both univariate models is given by the one point design

$\delta_\varrho^* = \varepsilon_1$ concentrated at the location 1 of the largest modulus of both regression functions t and t^2 on T.

The criterion under consideration for the bivariate model will be D-optimality, which is not affected by scale transformations of the componentwise variances. Hence we may assume without loss of generality

$$\Sigma(c) = \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix} \tag{57}$$

for various correlations c, $-1 < c < 1$.

As δ_ϱ^* is not D-optimum for the augmented univariate model $\mu_{1,2}(t) = \beta_1 t + \beta_2 t^2$, it is to be expected that δ_ϱ^* is not D-optimum for the bivariate model as well, at least, for large correlations $|c|$.

First we compare δ_ϱ^* with the best two-point design $\delta(\tau, c) = w^*(\tau, c)\varepsilon_\tau + (1 - w^*(\tau, c))\varepsilon_1$ concentrated on the endpoints of the design region $[\tau, 1]$, $\tau \neq 0$. That two-point design dominates δ_ϱ^* if

$$w(\tau, c) = \frac{1}{2}\,\frac{\tau^2(1-\tau)^2 - 2(1-\tau^3)(1-\varrho^2)}{\tau^2(1-\tau)^2 - (1-\tau^3)^2(1-\varrho^2)} > 0, \tag{58}$$

in which case $w^*(\tau, c) = w(\tau, c)$. Note that $w(\tau, c) = w(\tau, -c)$. Hence, for every τ there is a critical value

$$c_{krit}(\tau) = \left(1 - \frac{1}{2}\,\frac{\tau^2(1-\tau)^2}{1-\tau^3}\right)^{1/2} \tag{59}$$

for the correlation beyond which ($|c| > c_{krit}(\tau)$) the design δ_ϱ^* is dominated by $\delta(\tau, c)$. Note that $\delta(\tau, c)$ is not necessarily the D-optimum design for the bivariate model, although the Equivalence Theorem 6 ensures that the optimum design is concentrated on, at most, two supporting points including 1, at least, for $\tau \geq 0$. Note also that $c_{krit}(\tau) \to 0$ for $\tau \to -1$.

For the standard design region $T = [0, 1]$ the D-optimum design $\delta^*(c)$ is supported by, at most, two points including the right endpoint $t_1 = 1$. For $|c| \leq \min_{0<\tau<1} c_{krit}(\tau) = 0.981$ the one-point design δ_ϱ^* is D-optimum for the bivariate model. However, for stronger correlations $|c| \to 1$, it can be checked that a second supporting point t_2 is required for the D-optimum design and that t_2 approaches 0.5 and the corresponding optimal weight $w_2^* = w^*(t_2, c)$ tends to $\frac{1}{2}$. Hence $\delta^*(c)$ converges to the design $\delta^*(1)$ which is D-optimum in the augmented univariate model, as $|c| \to 1$.

In general, we consider

$$\begin{aligned}\varphi_c(\delta) &= \det(\textstyle\int f_0 f_0^T \, d\delta)^2 \det(\int \widetilde{f}_1 \widetilde{f}_1^T \, d\delta) \\ &\quad \times \det\left(\textstyle\int \widetilde{f}_2 \widetilde{f}_2^T \, d\delta - c^2 \int \widetilde{f}_2 \widetilde{f}_1^T \, d\delta (\int \widetilde{f}_1 \widetilde{f}_1^T \, d\delta)^{-1} \int \widetilde{f}_1 \widetilde{f}_2^T \, d\delta\right). \quad (60)\end{aligned}$$

Then, given $\Sigma(c)$, a design $\delta^*(c)$ is D-optimum for the bivariate model if, and only if, it maximizes $\varphi_c(\delta)$, $|c| < 1$. If $\delta^*(c) \to \delta(1)$ as $|c| \to 1$, then $\varphi_c(\delta^*(c)) \to \varphi_1(\delta(1))$, at least, for continuous regression functions on a compact design region T. Now, for $\delta^*(1)$ which maximizes $\varphi_1(\delta)$, we have $\varphi_c(\delta^*(1)) \le \varphi_c(\delta^*(c))$ and $\varphi_c(\delta^*(1)) \to \varphi_1(\delta^*(1))$, as $|c| \to 1$. Hence $\varphi_1(\delta^*(1)) \le \varphi_1(\delta(1))$ and $\delta(1)$ is also optimum with repsect to φ_1. Thus we can formulate the following result

THEOREM 7. *Let $\delta^*(c)$ be D-optimum in the bivariate model with mutual different complexity and underlying covariance structure $\Sigma(c)$, let a be continuous and T be compact. Then every limiting design $\delta(1)$ of $(\delta^*(c))$, as $|c| \to 1$, maximizes $\varphi_1(\delta)$.*

In the particular case that either $f_0 = 1$ or $(f_0) = \emptyset$ the design $\delta^*(1)$ maximizes $\varphi_1(\delta)$ if, and only if, $\delta^*(1)$ is D-optimum in the augmented univariate model. Moreover, if $\delta^*(1)$ is the unique D-optimum design in the augmented univariate model, then $\delta^*(c) \to \delta^*(1)$ as $|c| \to 1$. Hence, in this case, it is necessary for the simultaneous D-optimality of δ^* in the bivariate model for every covariance structure Σ that δ^* is also D-optimum in the augmented univariate model.

For the specific example, above, this situation occurs if $\tau = 1$ and the simultaneously optimum design δ^* assigns equal weights $\frac{1}{2}$ to both endpoints -1 and 1 of the design region $T = [-1, 1]$. Note that in this case also the factorization conditions are satisfied.

10. Appendix

By $\ell_\varrho = A_{d,\varrho}^T A_d^{(\varrho-1)} (A_d^{(\varrho-1)^T} A_d^{(\varrho-1)})^{-1}$ we denote the componentwise orthogonalization matrix of f_ϱ with respect to $a_{\varrho-1}$ and d. Then the transformation matrix

$$L_\varrho = \begin{pmatrix} \mathbf{E}_{q_1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \ell_2 & \mathbf{E}_{q_2} & \mathbf{0} & \cdots & \mathbf{0} \\ & \ell_3 & \mathbf{E}_{q_3} & \cdots & \mathbf{0} \\ & & \ddots & \ddots & \vdots \\ & & \ell_\varrho & & \mathbf{E}_{q_\varrho} \end{pmatrix} \quad (61)$$

for the univariate model, $\widetilde{a}_\varrho = L_\varrho a_\varrho$, is lower triangular with ones on the diagonal ($L_1 = \mathbf{E}_{q_1}$, $q_1 = p_1$). Finally, the transformation matrix L for the multivariate model, $\widetilde{a} = La$, is block diagonal

$$L = \begin{pmatrix} L_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & L_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & L_r \end{pmatrix} \tag{62}$$

and, hence, by (61), lower triangular with all diagonal entries one.

REFERENCES

Chang, S.I. (1994). Some properties of multiresponse D-optimal designs. *J. Math. Anal. Appl.* **184**, 256–262.

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

Krafft, O. and Schaefer, M. (1992). D-optimal designs for a multivariate regression model. *J. Multivariate Anal.* **42**, 130–140.

Kurotschka, V.G. (1984). A general approach to optimum design of experiments with qualitative and quantitative factors. In: *Statistics: Applications and New Directions: Proceedings of the Indian Statistical Institute Golden Jubilee International Conference Calcutta 1981.* (J.K. Ghosh and J. Roy (eds.)). Indian Statistical Institute, Calcutta. 353–368.

Kurotschka, V.G. (1988). Characterizations and examples of optimal experiments with qualitative and quantitative factors. In: *Model-Oriented Data Analysis, Proceedings Eisenach 1987.* (V. Fedorov and H. Läuter (eds.)). Springer, Berlin. 53–71.

Linde, A. van der (1977). Versuchsplanung für multivariate lineare Modelle. *Diplomarbeit, Freie Universität Berlin, Fachbereich Mathematik.* 175 pages.

Pázman, A. (1986). *Foundations of Optimum Experimental Design.* Reidel, Dordrecht.

Schwabe, R. (1995a). Experimental design for linear models with higher order interaction terms. In: *Symposia Gaussiana. Proceedings of the 2nd Gauss Symposium, München 1993. Conference B: Statistical Sciences* (V. Mammitzsch and H. Schneeweiß (eds.)). DeGruyter, Berlin. 281–288.

Schwabe, R. (1995b). Optimal designs for additive linear models. *Statistics* (in press).

Schwabe, R. (1995c). *Optimum Designs for Multi-factor Models. Lecture Notes in Statistics.* Springer, New York (to appear).

Wegscheider, K. (1977). Optimale Versuchsplanung für multivariate Beobachtungsprozesse mit kontinuierlich variierenden Versuchsbedingungen. *Diplomarbeit, Freie Universität Berlin, Fachbereich Mathematik.* 282 pages.