Optimal population design for uncorrelated mixed-effects models

M. Patan and B. Bogacka

August 29, 2006

1 Introduction

This is the draft of something what might become a paper for publication. At the moment, it needs writing an introduction with motivation, perhaps some simulation studies, conclusions and further work.

This paper is our response to the paper of Mentré et al. (1997). The authors define a population design in a way which simply restricts groups to those of fixed size (design for group one has one support point, for group two - two support points and so on, up to a predefined maximum number of support points). The example is taken from their paper, for comparison.

2 Regression model

We consider a nonlinear regression model

$$y = \eta(t, \theta) + \epsilon,$$

where $t \in T = [0, t_{max}]$, $t_{max} < \infty$, denotes an explanatory variable (time instant in our examples), $\theta \in \Theta$ is a p-dimensional vector of unknown parameters, Θ is a set of admissible values of θ, $\eta(t, \theta)$ is the expected response at t and ϵ denotes the random error of observations.

In what follows, we suppose that there is a population of N individuals (patients, units, systems etc.) for each of which n_i measurements are gathered, possibly according to different time schedules, that is, the model for each observation can be written as

$$y_i^k = \eta(t_i^k; \theta_i^k) + \varepsilon_i^k, \quad i = 1, \ldots, n_k, \quad k = 1, \ldots, N \tag{1}$$

where y_i^k is an observation at time $t_i^k \in T$, and ε_i^k are i.i.d. random errors with a known density f.

We denote by $g(y|\theta)$ the conditional probability density of observation y given the value of θ, i.e.

$$y_i^k|\theta^k \sim g(y_i^k|\theta^k).$$

Vectors $\theta^k \in \Theta$ are assumed to be realizations of the random vector $\theta = (\theta_1, \ldots, \theta_p)^{\mathrm{T}}$ with probability density $h(\theta;\psi)$, i.e.,

$$\theta \sim h(\theta;\psi).$$

The function h is entirely determined by the population parameter vector $\psi = (\psi_1, \ldots, \psi_{\widetilde{p}})^{\mathrm{T}}$. Efficient estimation of this vector of constant parameters ψ_i is of our primary interest.

Because of the iid assumption made for the errors, the log-likelihood function for ψ given a set of observations Y takes the form

$$\ell(\psi|Y) = \log \prod_{k=1}^{N} \prod_{i=1}^{n_k} \int_{\Theta} g(y_i^k|\theta) h(\theta;\psi)\, \mathrm{d}\theta \tag{2}$$

where $Y = [(y^1)^T, \ldots, (y^N)^T]^T$ and $y^k = [y_1^k, \ldots, y_{n_k}^k]^T$.

3 Experimental design

We assume that a population of N patients consists of G groups of size N_j each and the individuals in same group follow the same schedule of measurements (design). We construct the population experimental design in two stages:

Individual level The response variable y (e.g. concentration of a drug) is observed for each individual at specified time instants $t_i^j,\ i = 1, \ldots, n_j,\ j = 1, \ldots, G$. One experiment (e.g. visit at a clinic) yields one observation y_i^j. The experimental design for an individual in j-th group is a list of n_j time instants, at which measurements are to be made, namely

$$\xi_j = \{t_1^j, \ldots, t_{n_j}^j\}.$$

Since some of the time instants may be the same, it is convenient to write the design using proportions of the numbers of repeated observations, r_i^j, to the number of all observations taken for the individual at t_i^j, that is

$$\xi_j = \left\{ \begin{array}{ccc} t_1^j & \ldots & t_{s_j}^j \\ w_1^j & \ldots & w_{s_j}^j \end{array} \right. ; \quad w_i^j = r_i^j/n_j, \quad \left. \sum_{i=1}^{s_j} r_i^j = n_j \right\} \tag{3}$$

where s_j is the number of distinct time instants. We further relax the constraints in (3) and allow the weights w_i^j be any real numbers from the interval (0,1], and we write the design as

$$\xi_j = \left\{ \begin{array}{ccc} t_1^j & \dots & t_{s_j}^j \\ w_1^j & \dots & w_{s_j}^j \end{array} \; ; \quad w_i^j \in (0,1], \quad \sum_{i=1}^{s_j} w_i^j = 1 \right\}. \tag{4}$$

Design $\xi_j \in \Xi$, where Ξ denotes a set of admissible designs defined by (4), is a continuous measure on a set of s_j distinct (support) points in a design region T. Such definition of a design allows us to follow the convex design theory, see for example Fedorov (1972). It is also very useful from the numerical calculations point of view.

This form of a design does not preserve the information about the number of measurements n_j and the same design can be realized for different individuals with different experimental costs (since the number of measurements may vary). Hence, the whole experimental system per individual is described by the pair (ξ_j, n_j).

Population level We assume that all N_j individuals in group j, $j = 1, \dots, G$, follow the same sampling schedule. We define the population design as

$$\zeta = \left\{ \begin{array}{ccc} (\xi_1, n_1) & \dots & (\xi_G, n_G) \\ \alpha_1 & \dots & \alpha_G \end{array} \; ; \quad \sum_{j=1}^{G} \alpha_j = 1 \right\}, \tag{5}$$

where $\alpha_j = \frac{N_j}{N}$ is the proportion of individuals in the whole population who follow plan (ξ_j, n_j) (also relaxed later to be a set of positive values summing up to one).

This definition is a generalization of the definition proposed by Mentré et al. (1997) since individual designs ξ_j are continuous in contrast to those in Mentré et al. (1997) and the number s_j of support points for an individual in a group is not predetermined, and may, or may not, be the same as in other groups.

4 Fisher Information Matrix

The Fisher Information Matrix (FIM) is an argument of many design optimality criteria. In the fixed non-linear models (θ not random) it depends on the model parameters θ and on the individual experimental design. In the mixed effects non-linear models (θ random) it depends on the population parameters ψ, also called hyperparameters, and on the population experimental design.

The assumption of independent observations allows us to sum up the FIMs for all single observations. Here we have

$$M(\zeta, N) = N \sum_{j=1}^{G} \alpha_j M(\xi_j, n_j) = N \sum_{j=1}^{G} \alpha_j n_j \sum_{i=1}^{s_j} w_i^j M(t_i^j), \tag{6}$$

where

$$M(t_i^j) = \mathrm{E}\left\{ -\frac{\partial^2 \ell(\psi|y_i^j)}{\partial\psi \partial\psi^{\mathrm{T}}} \right\}$$

is the elementary FIM for the observation made at time instant t_i^j and

$$\ell(\psi|y_i^j) = \log \int_{\Theta} g(y_i^j|\theta) h(\theta;\psi)\, \mathrm{d}\theta.$$

Since for nonlinear response models, in general, the integral above is analytically intractable, in order to evaluate the FIM some approximation procedures are required. In the statistics literature there exist a variety of methods such as numerical integration or stochastic approximation. In the special case where h and f are normal density functions the linearisation of the model around the expected value of random-effect is most commonly used (Jones and Wang, 1999; Retout et al., 2001, 2002; Retout and Mentré, 2003) as it leads to a relatively simple closed form of the FIM.

5 Problem formulation

The design problem we are interested in is to optimize a functional Ψ operating on $\mathcal{M}$, a set of FIMs defined in (6),

$$\Psi : \mathcal{M} \longrightarrow \mathcal{R}.$$

We seek a FIM for which Ψ attains minimum, hence we can state our design problem as an optimization one:

$$\Psi[M(\zeta, N)] \longrightarrow \min. \tag{7}$$

In other words we look for a design $\zeta^\star$ which gives the optimum FIM for some a priori chosen values of the population parameters.

Some of the most common forms of Ψ used in the applications are (Walter and Pronzato, 1997; Fedorov and Hackl, 1997):

- D-optimality criterion:

$$\Psi(M) = -\det[M(\zeta, N)]], \tag{8}$$

- A-optimality criterion:

$$\Psi(M) = \text{trace}[M(\zeta, N)^{-1}], \tag{9}$$

- sensitivity criterion:

$$\Psi(M) = -\,\text{trace}[M(\zeta, N)]. \tag{10}$$

Further, we shall make the following assumptions:

(A1) T is compact,

(A2) $M(t)$ is continuous on T,

(A3) Ψ is convex,

(A4) If $M_1 \leq M_2$, then $\Psi(M_1) \geq \Psi(M_2)$.

The optimization problem (7) requires some additional constraints to have a bounded solution. This is due to the assumption of independent observations. Every new measurement carries some amount of new information and if the number of measurements is unlimited, the expected solution to the problem is unbounded. Here, we constrain the total number of observations to be not greater than N_0:

$$N \sum_{j=1}^{G} \alpha_j n_j \leq N_0. \tag{11}$$

In general, we introduce a nonnegative cost function $c(\xi_j, n_j)$, to bound the total cost, C_0, of gathering data:

$$N \sum_{j=1}^{G} \alpha_j c(\xi_j, n_j) \leq C_0. \tag{12}$$

If the number of all individuals in population is not predetermined *a priori* and has to be estimated, it is convenient to relax the restriction of N being a positive integer and allow it to take any positive real value. Then, we can formulate the following useful result:

Proposition 1. *Let the assumption (A4) hold, then the optimal solution* $(\zeta^\star, N^\star)$ *satisfies* (12) *on the boundary, i.e., the inequality becomes an equality at* $(\zeta^\star, N^\star)$.

Proof. First, note that since $M(t)$ is, by its definition, a nonnegative definite matrix, then for any ζ and positive N matrix $M(\zeta, N)$ being a sum of nonnegative definite matrices is also nonnegative definite matrix.

Further, let us assume that the optimal solution $N^\star$ and $\zeta^\star = \{(\xi_j^\star, n_j^\star, \alpha_j^\star)\}_{j=1}^{G}$ satisfies (12) strictly, i.e $N^\star \sum_{j=1}^{G} \alpha_j^\star c(\xi_j^\star, n_j^\star) < C_0$. Then, there exists $\lambda = \frac{N}{N^\star} > 1$ such

that $M(\zeta^\star, N) = \lambda M(\zeta^\star, N^\star)$ and $N\sum_{j=1}^{G} \alpha_j^\star c(\xi_j^\star, n_j^\star) = C_0$. But then, according to the monotonicity of Ψ (A4) we have

$$\Psi[M(\zeta^\star, N)] = \Psi[\lambda M(\zeta^\star, N^\star)] = \Psi[M(\zeta^\star, N^\star)+(\lambda-1)M(\zeta^\star, N^\star)] \leq \Psi[M(\zeta^\star, N^\star)]. \quad (13)$$

Hence, the pair $(\zeta^\star, N^\star)$ cannot be optimal solution. The obtained contradiction finishes the proof. ■

Remark 1. Although the property established in Proposition 1 is rather easy to show, it is not obvious, since the constraint (12) is not convex with respect to the set of parameters N, α_j and n_j. Its significance, however, cannot be underestimated. It allows to restate the problem as a convex optimization one, what will be shown in the following.

Most of common criteria are homogeneous, so in order to obtain the independence of the solution on the total cost of experiment (i.e. obtain the solution in terms of optimal proportions of the cost or numbers of repeated observations) we introduce, without any loss of generality, the so-called *average* per total cost (normalized) FIM

$$M(\upsilon) = \frac{N}{C_0}\sum_{j=1}^{G} \alpha_j c(\xi_j, n_j) \sum_{i=1}^{s_j} w_i^j M(t_i^j) = \sum_{j=1}^{G} \beta_j M(\xi_j), \quad (14)$$

where

$$\beta_j = \frac{N}{C_0}\alpha_j c(\xi_j, n_j)\ ; \quad M(\xi_j) = \sum_{i=1}^{s_j} w_i^j M(t_i^j)$$

and

$$\upsilon = \left\{ \begin{array}{ccc} \xi_1 & \dots & \xi_G \\ \beta_1 & \dots & \beta_G \end{array} ; \quad \beta_j \in (0,1], \quad \sum_{j=1}^{G} \beta_j = 1 \right\} \quad (15)$$

Due to Proposition 1, instead of solving the problem of minimizing $\Psi(M(\zeta, N))$ subject to (12) we can equivalently solve the problem of minimization of $\Psi(\upsilon)$ subject to $\sum_{j=1}^{G} \beta_j = 1$. This significantly simplifies the very complex procedure of finding an optimum population design, what will be shown in the following.

6 Characterizations of the optimal designs

In this section we study properties of the information matrices and optimal designs for mixed non-linear models considered in this paper.

At this point it is convenient to further generalize the considered designs to the absolutely continuous Lebesgue measures satisfying

$$\int_{\Xi} \upsilon(\mathrm{d}\xi) = 1. \quad (16)$$

Then we have

$$M(\upsilon) = \int_{\Xi} M(\xi)\upsilon(\mathrm{d}\xi) \tag{17}$$

Let Υ be a set of admissible designs defined by (15) with the normalization condition (16). Further, we shall make the additional assumptions:

(A5) There exists a finite real value a such that $\{\upsilon : \Psi[M(\upsilon)] \leq a < \infty\} = \Upsilon_a \neq \emptyset$,

(A6) For any $\upsilon \in \Upsilon_a$ and $\bar{\upsilon} \in \Upsilon$, functional Ψ is Fréchet differentiable at $M(\upsilon)$ in direction of $M(\bar{\upsilon})$.

First, we prove some properties of the FIM given by (17):

Lemma 2. *For any $\upsilon \in \Upsilon$, the matrix $M(\upsilon)$ is symmetric and non-negative definite.*

Proof. Since matrix $M(t)$ by definition is symmetric and nonnegative definite, then the $M(\upsilon)$ being an integral of a sum of nonnegative definite matrices with nonnegative weights is also a nonnegative matrix. ■

Lemma 3. *Let $\mathfrak{M}(\Xi)$ denote a set of all admissible information matrices for the designs in Υ, i.e.,*

$$\mathfrak{M}(\Xi) = \{M(\upsilon) : \upsilon \in \Upsilon\}. \tag{18}$$

The set $\mathfrak{M}(\Xi)$ is compact and convex.

Proof. See Appendix A.1. ■

Now we are able to further characterize the optimal designs.

Theorem 4. *Suppose that Assumptions (A1)–(A6) hold. Then an optimal design $\upsilon^\star$ exists and it consists of no more than $\widetilde{p}(\widetilde{p}+1)/2$ support points. Moreover, the set of optimal designs is convex.*

Proof. See the Appendix A.2. ■

Analytical determination of optimal designs is only possible in very simple cases and in practical problems some iterative design procedure are usually required. Then the Equivalence Theorem we present next, first proved by Kiefer and Wolfowitz (1960) for linear models, can be used to check optimality of numerically obtained designs.

In order to establish a form of the general equivalence theorem for the population design we introduce the so called *population sensitivity function* ψ_P, which exists due to the assumption (A6)

$$\psi_P(\xi, \upsilon) = \varsigma_P(\upsilon) - \phi_P(\xi, \upsilon), \tag{19}$$

$$\varsigma_P(\upsilon) = -\operatorname{trace}\left[\overset{\circ}{\Psi}[M(\upsilon)]M(\upsilon)\right], \tag{20}$$

$$\phi_P(\xi, \upsilon) = -\operatorname{trace}\left[\overset{\circ}{\Psi}[M(\upsilon)]M(\upsilon)\right], \tag{21}$$

where

$$\overset{\circ}{\Psi}[M(\upsilon)] = \left.\frac{\partial \Psi(M)}{\partial M}\right|_{M=M(\upsilon)}. \tag{22}$$

Theorem 5 (Generalized Equivalence Theorem for population design)**.** *Suppose that Assumptions (A1)–(A6) hold. The following conditions are equivalent:*

(i) the design $\upsilon^\star$ *minimizes* $\Psi[M(\upsilon)]$,

(ii) the design $\upsilon^\star$ *minimizes* $\max_{\xi \in \Xi} \phi_P(\xi, \upsilon) - \varsigma_P(\upsilon)$,

(iii) $\max_{\xi \in \Xi} \phi_P(\xi, \upsilon^\star) = \varsigma_P(\upsilon^\star)$

All the designs which satisfy (i)–(iii) and their convex combinations have the same information matrices equal to $M(\upsilon^\star)$*, provided that the criterion* $\Psi[\,\cdot\,]$ *is strictly convex.*

Proof. See Appendix A.3. ■

The properties of optimum designs $\upsilon^\star$ were derived assuming only the independence of observations among different groups. Therefore, direct application of Theorem (5) to build an efficient algorithm calculating such optimal designs is not straightforward and, in fact, there is lack of such procedures in the related literature.

However, taking advantage of the independence of the observations on the individual level, it is possible to strengthen this result. First, observe that the average FIM (14) may be rewritten in the form

$$M(\upsilon) = \sum_{j=1}^{G} \sum_{i=1}^{s_j} \frac{N}{C_0} \alpha_j c(\xi_j, n_j) w_i^j M(t_i^j) = \sum_{j=1}^{G} \sum_{i=1}^{s_j} q_i^j M(t_i^j) \tag{23}$$

where

$$q_i^j = \frac{N}{C_0} \alpha_j c(\xi_j, n_j) w_i^j = \beta_j w_i^j \; ; \quad \sum_{j=1}^{G} \sum_{i=1}^{s_j} q_i^j = 1.$$

Different groups do not have to have all support points different, that is, some points t_i^j may be the same for different j's. Consequently, it is sensible to reformulate the problem so as to operate on the locations $t_1, \ldots, t_s$ (relabelled different time instants) in lieu of t_i^j's. Here, we introduce weights $q_1, \ldots, q_s$ which are the sums of q_i^j's for the repeated time instants. This allows us to rewrite (23) as

$$M(\upsilon) = \sum_{k=1}^{s} q_k M(t_k) = M(\omega), \tag{24}$$

where

$$\omega = \left\{ \begin{array}{ccc} t_1 & \ldots & t_s \\ q_1 & \ldots & q_s \end{array} ; \quad \sum_{k=1}^{s} q_k = 1 \right\}. \tag{25}$$

Such reformulation makes it possible to solve the problem of finding the two level hierarchical optimal population design in terms of finding the equivalent one level design. Note, that $\omega \in \Xi$. We call it a *global design*.

The information about groups is included in q_i^j and so in q_k. This information is later recovered after an optimum design ω has been found. Also, such formulation of the population design significantly reduced the problem of dimensionality.

For every $\upsilon \in \Upsilon$ there exists such $\omega \in \Xi$ that $M(\upsilon) = M(\omega)$ and we have the analogous properties of those designs, in particular we have the following corollary, which can be proven in the same way as Theorem 4.

Corollary 6. *Suppose that Assumptions (A1)–(A6) hold. Then an optimal individual design $\omega^\star$ exists and it consists of no more than $\widetilde{p}(\widetilde{p}+1)/2$ support points. Moreover, the set of individual optimal designs is convex.*

Furthermore, we introduce, by analogy to (19)–(22), the sensitivity function ψ_I for the global design,

$$\psi_I(t, \omega) = \varsigma_I(\omega) - \phi_I(t, \omega), \tag{26}$$

$$\varsigma_I(\omega) = -\operatorname{trace}\left[\overset{\circ}{\Psi}[M(\omega)]M(\omega)\right], \tag{27}$$

$$\phi_I(t, \omega) = -\operatorname{trace}\left[\overset{\circ}{\Psi}[M(\omega)]M(t)\right], \tag{28}$$

where

$$\overset{\circ}{\Psi}[M(\omega)] = \left.\frac{\partial \Psi(M)}{\partial M}\right|_{M=M(\omega)}. \tag{29}$$

and we restate the Generalized Equivalence Theorem as follows.

Corollary 7 (Generalized Equivalence Theorem for global design)**.** *The following conditions are equivalent:*

(i) the design $\omega^\star$ *minimizes* $\Psi[M(\omega)]$*,*

(ii) the design $\omega^\star$ *minimizes* $\max_{\omega\in\Upsilon} \phi_I(t,\omega) - \varsigma_I(\omega)$*,*

(iii) $\max_{\omega\in\Upsilon} \phi_I(t,\omega^\star) = \varsigma_I(\omega^\star)$

All the designs which satisfy (i)–(iii) and their convex combinations have the same information matrices equal to $M(\omega^\star)$*, provided that the criterion* $\Psi[\,\cdot\,]$ *is strictly convex.*

This theorem makes it possible to solve the problem of determination of optimal population design through solving the conventional problem of approximate non hierarchical design. It allows to apply existing numerical procedures and also leads to other interesting features of optimal solutions what constitutes the subject of the next section.

7 Numerical algorithm

In addition to the minimax properties of optimal designs, the theorems of previous section give a powerful tool for checking the optimality of intuitively sensible designs or of those obtained numerically. We need efficient numerical procedures to construct Ψ-optimum design measures. Substantial difficulty in determining the population designs arises from the fact that they are not unique. Indeed, the criterion Ψ is most often strictly convex on $\mathfrak{M}(\Xi)$, and this guarantees that the optimal FIM is unique, but this does not necessarily mean that $(\zeta, N) \mapsto \Psi[M(\zeta, N)]$ is strictly convex in (ζ, N).

Hence, there is no guarantee that the optimum population design is unique. Multiple global solutions $(\zeta^\star, N^\star)$ may yield the same minimizing value of $M(\zeta, N)$.

Furthermore, there may be multiple local minima to $\Psi(\,\cdot\,)$ which highly interferes with the optimization process. In fact there is lack of efficient numerical tools for determining population designs.

Here we propose a new method of finding optimum population design based on the reformulation of the optimization problem presented in the previous section.

Instead of solving the original problem of minimizing $\Psi(M(\zeta, N))$ subject to (12) we first solve the equivalent problem of minimization of $\Psi(M(\omega))$ subject to $\sum_{k=1}^{s} q_k = 1$, which is far more simple. But, the solution to the latter requires a method to transform the optimal $\omega^\star$ into the original population design pair $(\zeta^\star, N^\star)$.

The determination of final solution can be achieved in three steps:

Step 1. Solve the optimization problem:

$$\omega^\star = \arg\min_{\omega\in\Xi} \Psi(M(\omega)). \tag{30}$$

Step 2. Transform $\omega^\star$ into an equivalent design $\upsilon^\star \in \Upsilon$, which satisfies

$$\upsilon^\star = \arg\min_{\upsilon\in\Upsilon} \Psi(M(\upsilon)). \tag{31}$$

Step 3. Transform $\upsilon^\star$ into an equivalent design pair $(\zeta^\star, N^\star)$.

After solving the minimization problem in Step 1 we have

$$\omega^\star = \left\{ \begin{array}{ccc} t_1^\star & \dots & t_s^\star \\ q_1^\star & \dots & q_s^\star \end{array} ; \quad \sum_{k=1}^{s} q_k^\star = 1 \right\}, \tag{32}$$

where $q_k^\star > 0,\ k = 1, \dots, s$. Then, we have to retrieve the components of $\upsilon^\star$, i.e $\{(\xi_j^\star, \beta_j^\star)\}_{j=1,\dots,G}$ or $\{(t_i^{j\star}, w_i^{j\star}, \beta_j^\star)\}_{j=1,\dots,G;\, i=1,\dots,s_j}$. Here we allow the weights w_i^j's take the zero values. Then each individual design has the same form

$$\xi_j^\star = \left\{ \begin{array}{ccc} t_1^\star & \dots & t_s^\star \\ w_1^{j\star} & \dots & w_s^{j\star} \end{array} ; \quad \sum_{k=1}^{s} w_k^{j\star} = 1 \right\},$$

and, technically, the problem of determining the design $\upsilon^\star$ simplifies to the determination of the weights $w_i^{j\star}$ and $\beta_j^\star$. This can be achieved solving the following system of equations:

$$\left\{ \begin{array}{lll} \beta_j w_i^j - q_i^j = 0, & i = 1, \dots, s, \ \ j = 1, \dots, G & (sG \text{ nonlinear equations, see (23)} \\ \sum_{i=1}^{s} w_i^j = 1, & j = 1, \dots, G & (G \text{ linear equations}) \\ \sum_{j=1}^{G} q_i^j = q_i^\star, & j = 1, \dots, s & (s \text{ linear equations}) \end{array} \right. \tag{33}$$

It is clear, that this system of $sG + s + G$ equations with $G + 2sG$ variables is unspecified if the number of groups $G > 1$. There are $s(G-1)$ more variables than equations. Although the first sG equations are nonlinear, it is easy to show that the solution with nonnegative values of all variables always exists. Treating $s(G-1)$ variables q_i^j as nonnegative parameters (satisfying the condition that for any j, $j = 1, \dots, G$, there exists at least one positive value q_i^j) the solution becomes simple:

$$\left\{ \begin{array}{ll} \beta_j^\star = \sum_{i=1}^{s} q_i^j, & j = 1, \dots, G, \\ w_i^{j\star} = q_i^j / \beta_j^\star, & i = 1, \dots, s, \ \ j = 1, \dots, G, \end{array} \right. \tag{34}$$

These values are further used in Step 3.

The optimal values of the population parameters $\alpha_j^\star, n_j^\star,\ j=1,\ldots,G$ and $N^\star$ can be retrieved solving the system of equations:

$$\begin{cases} \dfrac{N}{C_0}\alpha_j c(\xi_j^\star, n_j) = \beta_j^\star, \quad j=1,\ldots,G & (G \text{ nonlinear equations}) \\ \sum_{i=1}^G \alpha_j = 1, & (1 \text{ linear equation}) \end{cases} \tag{35}$$

Here again, we have unspecified system (since we have $G+1$ equations and $2G+1$ variables) of nonlinear equations which can be easily solved numerically, with accuracy to the G parameters. It is clear that such formulation of the optimal population design problem leads to not unique solutions. This, however, does not worry us. It gives room for tailoring optimum designs to practical requirements and makes it sensible to use some additional information an experimenter may have. It is a great advantage of the method as it gives the experimenter some freedom to impose some additional constraints on the design variables.

Furthermore, the solution depends on the cost function $c(\,\cdot\,,\,\cdot\,)$. For example, if the numbers of observations per individual in each group, $n_j,\ j=1,\ldots,G$, are known or can be chosen arbitrarily, then the optimal solution exists and takes the following form

$$N^\star = C_0 \sum_{j=1}^G \frac{\beta_j^\star}{c(\xi_j^\star, n_j)}, \quad \alpha_j^\star = \frac{C_0}{N^\star}\frac{\beta_j^\star}{c(\xi_j^\star, n_j)},\ \ j=1,\ldots,G. \tag{36}$$

Two special forms of population designs are given in the following result:

Theorem 8. *Assume, that the cost function $c(\xi_j, n_j)$ takes positive values for any positive n_j and nonempty design ξ_j. If $\omega^\star$ is a solution to* (30) *given by* (32)*, then:*

(i) the design $\zeta^\star = \left\{ \begin{smallmatrix} \omega_1^\star \\ 1 \end{smallmatrix} \right\}$ and $N^\star = \frac{C_0}{c(\omega^\star, n_1)}$ for any $n_1 > 0$ minimizes $\Psi[M(\zeta^\star, N^\star)]$,

(ii) the design $\zeta^\star = \left\{ \begin{smallmatrix} \omega_1^\star & \ldots & \omega_s^\star \\ q_1^\star & \ldots & q_s^\star \end{smallmatrix} \right\}$ and $N^\star = C_0 \sum_{j=1}^s \frac{q_j^\star}{c(\omega_j^\star, n_j)}$, where $\omega_j^\star = \left\{ \begin{smallmatrix} t_j^\star \\ 1 \end{smallmatrix} \right\}$ for any $n_j > 0,\ \ j=1,\ldots,s$ minimizes $\Psi[M(\zeta^\star, N^\star)]$.

Proof. To prove the assertions above it is sufficient to show that the information matrices $M(\zeta^\star, N^\star)$ and $M(\omega^\star)$ are the same, and this is a matter of simple calculations. Indeed:

(i) In this case from (14) we have

$$M(\zeta^\star, N^\star) = \frac{\frac{C_0}{c(\omega^\star, n_1)}}{C_0} \sum_{i=1}^s c(\omega^\star, n_1) q_i^\star M(t_i^\star) = \sum_{i=1}^s q_i^\star M(t_i^\star) = M(\omega^\star)$$

what proves the first part of the claim.

(*ii*) Here, taking from (34) that $\beta_j^\star = q_j^\star,\ j = 1, \ldots, s$ and applying (36) we obtain

$$\alpha_j^\star = \frac{C_0}{N^\star} \frac{\beta_j^\star}{c(\xi_j^\star, n_j)},\ \ j = 1, \ldots, s$$

and

$$\begin{aligned} M(\zeta^\star, N^\star) &= \frac{C_0 \sum_{j=1}^s \frac{q_j^\star}{c(\omega_j^\star, n_j)}}{C_0} \sum_{j=1}^s \frac{C_0}{C_0 \sum_{j=1}^s \frac{q_j^\star}{c(\omega_j^\star, n_j)}} \frac{q_j^\star}{c(\omega_j^\star, n_j)} c(\omega_j^\star, n_j) M(t_j^\star) \\ &= \sum_{j=1}^s q_j^\star M(t_j^\star) = M(\omega^\star) \end{aligned}$$

what finishes the proof. ■

The result above indicates that when the observations are independent there exist population designs of very simple forms: the identical population design (in the sense of the same observation schedule for every individual) and the one-point population designs (with only one observation time instant for every individual).

Despite the complex formulation of the problem, the approximations of the optimal population designs are simple. This fact, seemed to be unrevealed in the related literature.

With such a general definition of the population design as in this paper, it is clear that the very difficult optimization problem related to direct determination of the population design can be solved efficiently through the proper reformulation. Then, the most cumbersome subtask of the whole procedure becomes the optimization problem in Step 1. But, this can be considered as a classical experimental design problem for dynamic systems, which has been thoroughly studied for many years and there are many efficient algorithms for this purpose. For reviews, reader can be referred to (Fedorov and Hackl, 1997; Walter and Pronzato, 1997; Patan, 2004; Uciński, 2005; Atkinson and Donev, 1992; Pázman, 1986; Rafajłowicz, 1996). In this work we analyze various optimality criteria and it is convenient to use the Semi-Definite Programming algorithms based on the convex optimization theory, (Boyd and Vandenberghe, 2004).

8 Examples

8.1 Toxicokinetic study

As the first example we consider the toxicokinetic studies performed on the rodents (Mentré et al., 1997). After a single dose of a drug, the model is:

$$y = \frac{D}{V} e^{-\frac{CL}{V} t} e^{\varepsilon}, \tag{37}$$

where CL is the oral clearence and V is the apparent volume for the distribution, D is a dose and ε is a zero-mean uncorrelated Gaussian measurement noise with a constant variance. Under the log-transformation the model becomes:

$$z = \log y = \log D - \theta_2 - e^{\theta_1 - \theta_2} t + \varepsilon$$

where $\theta = (\theta_1, \theta_2)^{\mathrm{T}} = (\log CL, \log V)^{\mathrm{T}}$ are the parameters which are assumed to be independent and normally distributed. The prior values of the population parameters are:

$$\begin{aligned}\psi_0 &= \big(E(\log CL), E(\log V), \mathrm{var}(\log CL), \mathrm{var}(\log V), \mathrm{cov}(\log CL, \log V)\big)^{\mathrm{T}} \\ &= (-0.476, 1.937, 0.073, 0.187, 0.000)^{\mathrm{T}} \quad \text{and} \quad \mathrm{var}(\varepsilon) = 0.108.\end{aligned}$$

We seek a population design to estimate the population parameters as precisely as possible. In our example the D-optimality criterion was applied.

We consider the design space $T = [0.5, 24]$ scaled in hours after administration of the drug. Three cost functions were studied, imposing the total cost $C_0 = 60$, namely:

- $c_1(\xi, n) = n$, restriction of the number of measurements n in the design ξ,
- $c_2(\xi, n) = n + 2$, penalization of the additional individual in the group,
- $c_3(\xi, n) = n(1 + \max \mathrm{supp}\, \xi - \min \mathrm{supp}\, \xi)$, additional term for the duration of the experiment.

In Step 1 of the procedure described in the previous section we obtain the optimal D-optimal design

$$\omega^\star = \begin{Bmatrix} 0.5000 & 9.0300 & 24.0000 \\ 0.3340 & 0.3319 & 0.3341 \end{Bmatrix}$$

with $\det(M(\omega^\star)) = 1.6093 \cdot 10^5$. The variance of the model prediction function is shown in Fig. **??**. It entirely determines the locations of observations.

Then, according to the required structure of the experiment it is possible to find suitable population design. For example, if we assume that the number of measurements for each individual design is fixed *a priori*, e.g. $n = 6$ then the identical population design is $\zeta^\star = \{(\omega^\star, 1)\}$ with the population numbers $N_1^\star = 10$, $N_2^\star = 7.5$, $N_3^\star = 0.4082$ for the cost functions c_1, c_2 and c_3, respectively, is D-optimal:

$$\zeta = \left\{ \begin{array}{c} \left(\begin{Bmatrix} 0.5000 & 9.0300 & 24.0000 \\ 0.3340 & 0.3319 & 0.3341 \end{Bmatrix}, 6 \right) \\ 1 \end{array} \right\}; \quad N^\star.$$

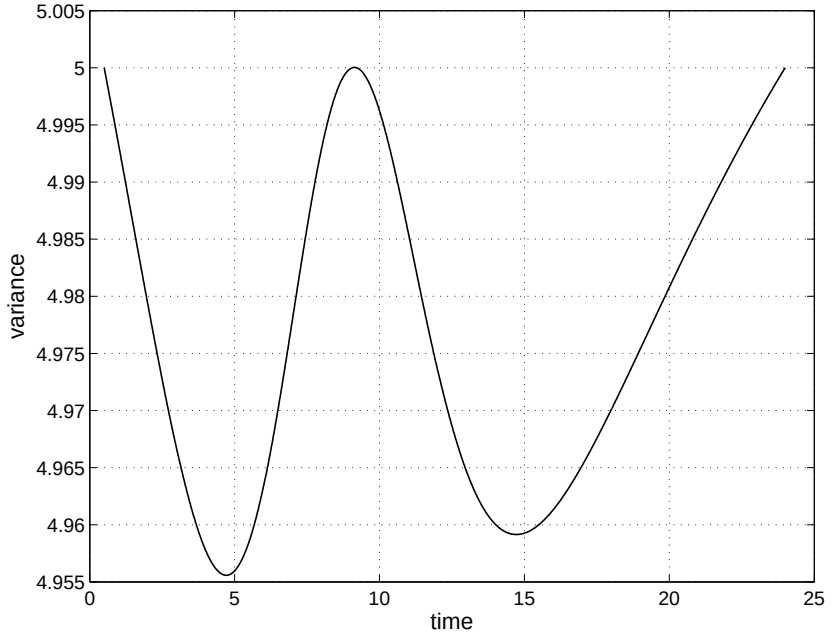

Figure 1: Variance of the model response prediction for the example 8.1.

On the other hand one can be interested in possibly simplest in application one-point individual designs (sparse sampling). Then, the population sizes are $N_1^\star = 10$,$N_2^\star = 7.5$, $N_3^\star = 10$, respectively:

$$\zeta = \left\{ \begin{array}{ccc} \left(\left\{ {}^{0.5000}_{1}, \right\}, 6\right) & \left(\left\{ {}^{9.0300}_{1}, \right\}, 6\right) & \left(\left\{ {}^{24.0000}_{1}, \right\}, 6\right) \\ 0.3340 & 0.3319 & 0.3341 \end{array} \right\}; \quad N^\star.$$

It becomes clear that for some cost functions we may obtain unreasonable values of population parameters, as the case of $N_3^\star$ for identical design. But then, due to the freedom in choosing the parameters when solving Steps 2 and 3 in our procedure it is possible to appropriately adjust their values.

These are two "extreme" population designs, but there is a class of equivalent solutions. For instance, arbitrarily assigning values to the weights q_i^j in three groups in step 2 of the procedure we obtain the following population design

$$\zeta^\star = \left\{ \begin{array}{ccc} \left(\left\{ {}^{0.5}_{0.5226}\ {}^{9.03}_{0.4774} \right\}, 3\right) & \left(\left\{ {}^{0.5}_{0.1700}\ {}^{9.03}_{0.4440}\ {}^{24.0}_{0.3861} \right\}, 3\right) & \left(\left\{ {}^{0.5}_{0.2842}\ {}^{24.0}_{0.7158} \right\}, 3\right) \\ 0.3743 & 0.3451 & 0.2806 \end{array} \right\}$$

with population size $N^\star = 20$ which yield to be optimal.

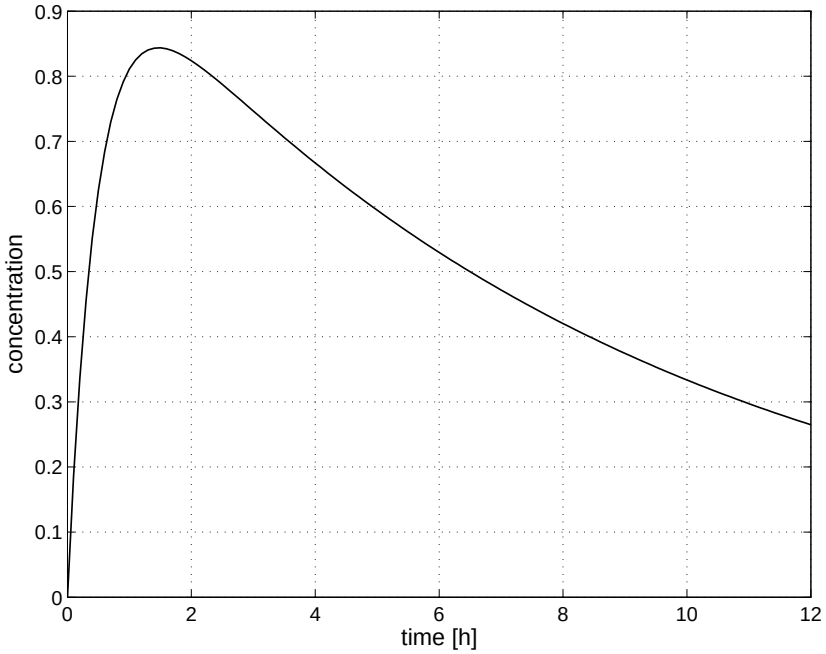

Figure 2: Absorption model for the example 8.2.

8.2 Pharmacokinetic study

The second example is based on (Jonsson et al., 1996) where using population analysis, sparsely sampled Phase 3 clinical data are utilized to determine the pharmacokinetic characteristics of the target population. The typical open one-compartment model with first-order drug absorption is used:

$$y = \frac{Dk_a}{V(k_e - k_e)}\left(e^{-k_e t} - e^{-k_a t}\right) + \varepsilon, \tag{38}$$

where k_a and k_e are the first-order absorption and elimination rates, respectively, V is the apparent volume of distribution, D is a dose and ε is an additive zero-mean uncorrelated Gaussian measurement noise with a constant variance. In this example we assume that the regression parameters $\theta = (\theta_1, \theta_2, \theta_3)^{\mathrm{T}} = (V, k_a, k_e)^{\mathrm{T}}$ are independent and normally distributed. The prior values of the population parameters are:

$$\begin{aligned}\psi_0 &= \big(E(\theta), \mathrm{var}(\theta)\big)^{\mathrm{T}} \\ &= (100, 2.08, 0.1155, 0.3, 0.3, 0.03)^{\mathrm{T}} \quad \text{and} \quad \mathrm{var}(\varepsilon) = 0.15.\end{aligned}$$

The response shown in Fig. 2 represents the fast absorption of the drug and slower decay of its concentration. As in the previous example we are looking for an D-optimum population design to estimate the population parameters as precisely as possible.

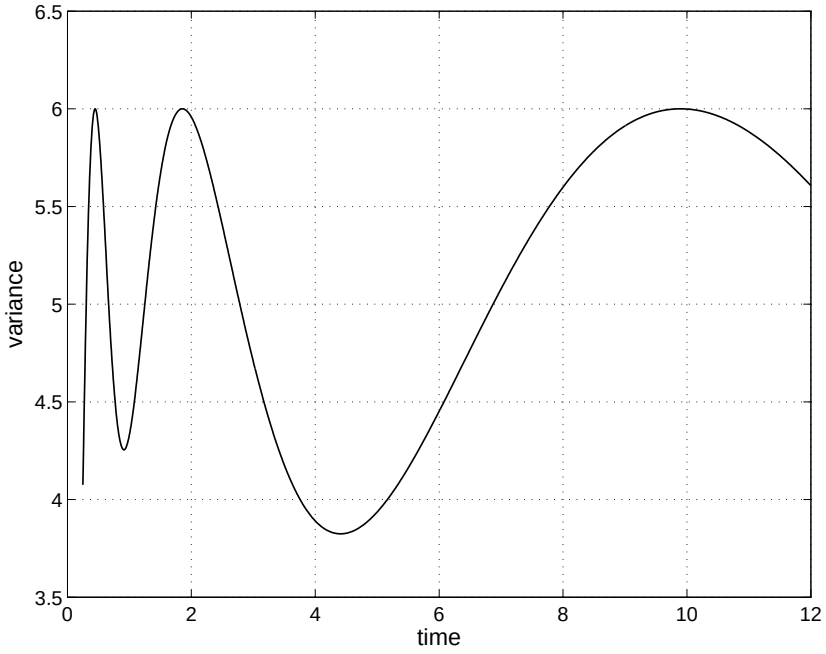

Figure 3: Variance of the model response prediction for the example 8.2.

We consider that the concentration of the drug can be measured within the design space $T = [0.25, 12]$ scaled in hours after administration. In this example only cost function $c_1(\xi, n) = n$ is taken into account and the total number of measurements is assumed to be $C_0 = 900$.

In this case the global design from the Step 1 of algorithm is:

$$\omega^\star = \left\{ \begin{array}{ccc} 0.45 & 1.86 & 9.90 \\ 0.3334 & 0.3334 & 0.3333 \end{array} \right\}$$

and the corresponding variance of the model prediction function is shown in Fig. 3, where the support points are indicated by its maxima.

Within this example let us explore further the properties of the set of population designs and freedom of constructing the suitable structures of the experiment:

(1) Identical design (one group design), $G = 1$, $n_1 = 9$,

(a) forcing in Step 2 the matrix $Q = [q_i^j]$ to be the column vector we have:
$Q^\star = [q_i^{j\star}] = \left[0.3333 \quad 0.3334 \quad 0.3333\right]^{\mathrm{T}} \Longrightarrow \beta^\star = \left[1\right], \quad W^\star = [w_i^{j\star}] = Q^\star$

(b) then, from Step 3: $\alpha^\star = 1,\ N^\star = 100$

with final population design:

$$\zeta = \left\{ \begin{array}{c} \left(\left\{ \begin{array}{ccc} 0.45 & 1.86 & 9.90 \\ 0.3333 & 0.3334 & 0.3333 \end{array} \right\}, 9 \right) \\ 1 \end{array} \right\}; \quad N^\star = 100.$$

In this case final population design is also exact design, so for each patient we have to conduct exactly three measurements at each time instant.

(2) one-point population design, $G = 3$, $n_1 = n_2 = n_3 = 10$

(a) this time in Step 2 matrix Q should be diagonal:

$$Q^\star = [q_i^{j\star}] = \begin{bmatrix} 0.3333 & 0.0 & 0.0 \\ 0.0 & 0.3334 & 0.0 \\ 0.0 & 0.0 & 0.3333 \end{bmatrix} \Longrightarrow \beta^\star = \begin{bmatrix} 0.3333 & 0.3334 & 0.3333 \end{bmatrix}, W^\star = I$$

(b) from Step 3:

$$\alpha^\star = \beta^\star = \begin{bmatrix} 0.3333 & 0.3334 & 0.3333 \end{bmatrix}, \quad N^\star = 90$$

and final population design is:

$$\zeta^\star = \begin{Bmatrix} \left(\left\{ \begin{smallmatrix} 0.45 \\ 1 \end{smallmatrix} \right\}, 10 \right) & \left(\left\{ \begin{smallmatrix} 1.86 \\ 1 \end{smallmatrix} \right\}, 10 \right) & \left(\left\{ \begin{smallmatrix} 9.90 \\ 1 \end{smallmatrix} \right\}, 10 \right) \\ 0.3333 & 0.3334 & 0.3333 \end{Bmatrix}; \quad N^\star = 90$$

which again appears to be exact one (since each group consist of 30 patients) and rounding is unnecessary.

(3) unstructured design, $G = 3$, $n_1 = n_2 = n_3 = 10$

(a) in Step 2 inside matrix Q the weights of global design are randomly split into required numer of groups, e.g.

$$Q^\star = [q_i^{j\star}] = \begin{bmatrix} 0.2298 & 0 & 0.1710 \\ 0.1036 & 0.2089 & 0.0855 \\ 0 & 0.1245 & 0.0768 \end{bmatrix},$$

then

$$\beta^\star = \begin{bmatrix} 0.4008 & 0.3979 & 0.2013 \end{bmatrix}, \quad W^\star = \begin{bmatrix} 0.5733 & 0.2603 & 0 \\ 0 & 0.5248 & 0.6184 \\ 0.4267 & 0.2149 & 0.3816 \end{bmatrix}$$

(b) from Step 3:

$$\alpha^\star = \beta^\star = \begin{bmatrix} 0.4008 & 0.3979 & 0.2013 \end{bmatrix}, \quad N^\star = 90.$$

and final population design:

$$\zeta^\star = \begin{Bmatrix} \left(\left\{ \begin{smallmatrix} 0.45 & 9.90 \\ 0.5733 & 0.4267 \end{smallmatrix} \right\}, 10 \right) & \left(\left\{ \begin{smallmatrix} 0.45 & 1.86 & 9.90 \\ 0.2603 & 0.5248 & 0.2149 \end{smallmatrix} \right\}, 10 \right) & \left(\left\{ \begin{smallmatrix} 1.86 & 9.90 \\ 0.6184 & 0.3816 \end{smallmatrix} \right\}, 10 \right) \\ 0.4008 & 0.3979 & 0.2013 \end{Bmatrix}.$$

For this setting, the population design cannot be realized in practice, therefore it has to be rounded in order to have integer numbers of patients in each group and integer numbers of observations allocated to each time instant. As an effect of such procedure we obtain the following exact design:

$$\zeta^\star = \begin{Bmatrix} \left(\left\{ \begin{smallmatrix} 0.45 & 9.90 \\ 0.6 & 0.4 \end{smallmatrix} \right\}, 10 \right) & \left(\left\{ \begin{smallmatrix} 0.45 & 1.86 & 9.90 \\ 0.3 & 0.5 & 0.2 \end{smallmatrix} \right\}, 10 \right) & \left(\left\{ \begin{smallmatrix} 1.86 & 9.90 \\ 0.6 & 0.4 \end{smallmatrix} \right\}, 10 \right) \\ 0.4 & 0.4 & 0.2 \end{Bmatrix}$$

which efficiency is:

$$\left(\frac{\det M(\zeta, N)}{\det M(\zeta^\star, N^\star)} \right)^{1/6} = 0.9984$$

As we can see the relatively large number of total measurements ensures small decrease in the efficiency caused by the rounding of design. However, the freedom of constructing the structure of required optimal design allows us to reduce the influence of rounding on efficiency of final design.

(4) unstructured design, $G = 3$, $n_1 = n_2 = n_3 = 10$

(a) this time in Step 2 we can force the elements of matrix W to be appropriate rationals, e.g.

$$W^\star = [w_i^{j\star}] = \begin{bmatrix} 0 & 0 & \frac{3}{10} \\ 0 & \frac{2}{5} & \frac{7}{10} \\ 1 & \frac{3}{5} & 0 \end{bmatrix},$$

then

$$Q^\star = \begin{bmatrix} 0 & 0 & 0.3333 \\ 0 & 0.1333 & 0.2000 \\ 0.1000 & 0.2333 & 0 \end{bmatrix},$$

$$\beta^\star = \begin{bmatrix} 0.3333 & 0.3334 & 0.3333 \end{bmatrix}$$

(b) from Step 3:

$$\alpha^\star = \beta^\star = \begin{bmatrix} 0.3333 & 0.3334 & 0.3333 \end{bmatrix}, \quad N^\star = 90.$$

and final population design

$$\zeta^\star = \begin{Bmatrix} (\{\begin{smallmatrix}9.90\\1\end{smallmatrix}\},10) & (\left\{\begin{smallmatrix}1.86 & 9.90\\ \frac{2}{5} & \frac{3}{5}\end{smallmatrix}\right\},10) & (\left\{\begin{smallmatrix}0.45 & 1.86\\ \frac{3}{10} & \frac{7}{10}\end{smallmatrix}\right\},10) \\ 0.3333 & 0.3334 & 0.3333 \end{Bmatrix}$$

is almost exact design with efficiency after rounding all its global weights to $\frac{1}{3}$

$$\left(\frac{\det M(\zeta, N)}{\det M(\zeta^\star, N^\star)}\right)^{1/6} \approx 1.0000$$

In such a way, the dependence of the efficiency of final population design on the rounding can be minimized.

A Proofs of some properties of optimal population designs

A.1 Proof of Lemma 3

First, note that the set Ξ is compact, what is a direct consequence of assumption (A1), since all the elements of any design $\xi \in \Xi$ are bounded and the set Ξ is closed. Moreover, from the (A2) comes that the $M(\xi)$ as a sum of continuous mappings is also continuous in Ξ.

Then, the set $S(\Xi) = \{M(\xi) : \xi \in \Xi)$ is compact, being the image of the compact set Ξ under the continuous mapping $M(\xi)$, from Ξ into the space of all $q \times q$ matrices. $\mathfrak{M}(\Xi)$ form the convex hull of $S(\Xi)$. Because in the Euclidean space the convex hull of a compact set is compact (Rockafellar, 1970) and the spaces $\mathbb{R}^{q\times q}$ and $\mathbb{R}^{q^2}$ are isomorphic, then the $\mathfrak{M}(\Xi)$ is compact.

To prove the convexity, let us introduce the design

$$\upsilon = (1-\lambda)\upsilon_1 + \lambda\upsilon_2, \tag{39}$$

where $\upsilon_1, \upsilon_2 \in \Upsilon$ and $\lambda \in [0,1]$. Now, constructing the convex combination of the matrices corresponding to υ_1, υ_2 we have

$$\begin{aligned}(1-\lambda)M(\upsilon_1) + \lambda M(\upsilon_2) &= (1-\lambda)\int_\Xi M(\xi)\upsilon_1(\mathrm{d}\xi) + (1-\lambda)\int_\Xi M(\xi)\upsilon_2(\mathrm{d}\xi) \\ &= \int_\Xi M(\xi)[(1-\lambda)\upsilon_1(\mathrm{d}\xi) + \lambda\upsilon_2(\mathrm{d}\xi)] = M(\upsilon)\end{aligned} \tag{40}$$

According to (18), $M(\upsilon) \in \mathfrak{M}(\Xi)$ which proves the second part of the lemma.

A.2 Proof of Theorem 4

The existence of an optimal design $\upsilon^\star$ follows from the compactness of $\mathfrak{M}(\Xi)$ (see Lem. 3) and the existence of designs with finite measure Ψ, cf. (A5).

The set $\mathfrak{M}(\Xi)$ is the convex hull of the set $S(\Xi) = \{M(\xi) : \xi \in \Xi\}$. The dimension of this set is $d = q(q+1)/2$ due to the symmetry of the matrix $M(\xi)$ (it is sufficient to use only elements lying over and on the main diagonal). Applying the Carathedéodory theorem (Fedorov, 1972; Pukelsheim, 1993), we may represent any $M_0 \in \mathfrak{M}(\Xi)$ as a convex combination of no more than d_0 points from $S(\Xi)$:

$$M_0 = \sum_{i=1}^{d_0} w_i M(\xi_i), \quad \sum_{i=1}^{d_0} w_i = 1 \tag{41}$$

where $d_0 \leq d+1$ in a general case and $d_0 \leq d$ for boundary points. Choosing

$$\xi = \begin{Bmatrix} \xi_1 & \cdots & \xi_{d_0} \\ w_1 & \cdots & w_{d_0} \end{Bmatrix},$$

we have $M_0 = M(\xi)$. From the monotonicity of the criterion Ψ in (A4) it follows that $M(\upsilon^\star)$ has to be a boundary point of $\mathfrak{M}(\Xi)$. Indeed, if we assume that $M(\upsilon^\star)$ is an interior point of $\mathfrak{M}(\Xi)$ then there exists $\lambda > 1$ such that $\lambda M(\upsilon^\star) \in \mathfrak{M}(\Xi)$. Consequently, there exist some design υ, whose information matrix is given by $\lambda M(\upsilon^\star)$, but then we have

$$\Psi[M(\upsilon^\star)] > \Psi[M(\lambda\upsilon^\star)] = \Psi[M(\upsilon)]$$

and this contradicts the optimality of the design $\upsilon^\star$. Thus, if $M(\upsilon^\star)$ is a boundary point of $\mathfrak{M}(\Xi)$, then it have no more than d support points.

For the last part of the assertion, assume that $\upsilon_1^\star$ and $\upsilon_2^\star$ are optimal and $\upsilon^\star = \lambda\upsilon_1^\star + (1-\lambda)\upsilon_2^\star$. From the convexity of $\Psi(\,\cdot\,)$ (A3) and the set $\mathfrak{M}(\Xi)$, we have

$$\begin{aligned}\Psi[M(\upsilon^\star)] &= \Psi[\lambda M(\upsilon_1^\star) + (1-\lambda)M(\upsilon_2^\star)] \leq \lambda\Psi[M(\upsilon_1^\star)] + (1-\lambda)\Psi[M(\upsilon_2^\star)] \\ &= \lambda \min_{\upsilon\in\Upsilon} \Psi[M(\upsilon)] + (1-\lambda) \min_{\upsilon\in\Upsilon} \Psi[M(\upsilon)] = \min_{\upsilon\in\Upsilon} \Psi[M(\upsilon)],\end{aligned}$$

hence $\upsilon^\star$ is an optimal design.

A.3 Proof of Theorem 5

In order to prove *equivalence theorem*, we have to derive some auxiliary results.

Lemma 9. *For any design $\upsilon \in \Upsilon$, we have*

(i) $\int_\Xi \phi_P(\xi, \upsilon)\upsilon(\mathrm{d}\xi) = \varsigma_P(\upsilon)$, *and*

(ii) $\max_{\xi\in\Xi} \phi_P(\xi, \upsilon) \geq \varsigma_P(\upsilon)$.

Proof. Taking into account (21), we obtain

$$\begin{aligned}\int_{\Xi} \phi_P(x,\upsilon)\upsilon(\mathrm{d}\xi) &= -\int_{\Xi} \operatorname{trace}\left[\overset{\circ}{\Psi}[M(\upsilon)]M(\xi)\right]\upsilon(\mathrm{d}\xi)\\ &= -\operatorname{trace}\left[\overset{\circ}{\Psi}[M(\upsilon)]\int_{\Xi} M(\xi)\upsilon(\mathrm{d}\xi)\right]\\ &= -\operatorname{trace}\left[\overset{\circ}{\Psi}[M(\upsilon)]M(\upsilon)\right] = \varsigma_P(\upsilon)\end{aligned} \tag{42}$$

This establishes (i). Then (ii) is a direct consequence of (42). ■

Lemma 10. *If* $\upsilon \in \Upsilon_q$, $\bar{\upsilon} \in \Upsilon$ *and* $\upsilon_\alpha = (1-\alpha)\upsilon + \alpha\bar{\upsilon}$, *then*

$$\left.\frac{\partial \Psi[M(\upsilon_\alpha)]}{\partial \alpha}\right|_{\alpha=0^+} = \varsigma_P(\upsilon) - \int_{\Xi} \phi_P(\xi,\upsilon)\,\bar{\upsilon}(\mathrm{d}\xi). \tag{43}$$

Proof. We have

$$\begin{aligned}\left.\frac{\partial \Psi[M(\upsilon_\alpha)]}{\partial \alpha}\right|_{\alpha=0^+} &= \operatorname{trace}\left[\overset{\circ}{\Psi}[M(\upsilon)]\int_{\Xi} M(\xi)\,\bar{\upsilon}(\mathrm{d}\xi)\right] - \operatorname{trace}\left[\overset{\circ}{\Psi}[M(\upsilon)]M(\upsilon))\right]\\ &= \int_{\Xi}\left\{\operatorname{trace}\left[\overset{\circ}{\Psi}[M(\upsilon)]M(\xi)\right]\right\}\bar{\upsilon}(\mathrm{d}x) + \varsigma_P(\upsilon)\\ &= \varsigma_P(\upsilon) - \int_{\Xi}\phi_P(\xi,\upsilon)\,\bar{\upsilon}(\mathrm{d}x).\end{aligned} \tag{44}$$

■

Now, we are capable of deriving our main result:

First, define $\upsilon_\alpha = (1-\alpha)\upsilon^\star + \alpha\upsilon_1$, where $\upsilon^\star \in \upsilon_q$, and $\upsilon_1 \in \Upsilon$.

$(i) \Rightarrow (ii)$ If the optimal design $\upsilon^\star$ minimizes $\Psi(M(\upsilon))$, then $\Psi(M(\upsilon^\star)) \leq \Psi[M(\upsilon_\alpha)]$ for any $\upsilon_1 \in \Upsilon$, therefore

$$\left.\frac{\partial \Psi[M(\upsilon_\alpha)]}{\partial \alpha}\right|_{\alpha=0^+} \geq 0, \quad \forall \upsilon_1 \in \Upsilon. \tag{45}$$

In particular substituting, $\upsilon = \upsilon^\star$ and $\bar{\upsilon} = \upsilon_\xi = \left\{\begin{smallmatrix}\xi\\1\end{smallmatrix}\right\}$ into (43), we get

$$\left.\frac{\partial \Psi[M(\upsilon_\alpha)]}{\partial \alpha}\right|_{\alpha=0^+} = \varsigma_P(\upsilon^\star) - \phi_P(\xi,\upsilon^\star) \geq 0, \quad \forall \xi \in \Xi. \tag{46}$$

In connection with the second part of Lemma 9 this establishes (ii).

$(ii) \Rightarrow (iii)$ Lemma 9 implies that $\max_{\xi\in\Xi}\phi_P(\xi,\upsilon) - \varsigma_P(\upsilon)$ is bounded from below by zero. From (46) it follows that this zero bound is achieved at any design minimizing $\Psi[M(\upsilon)]$ (the existence of such a design is guaranteed by Theorem 4). This means that if $\upsilon^\star$ is a design characterized in (ii), then necessarily $\max_{x\in X}\phi_P(\xi,\upsilon^\star) - \varsigma_P(\upsilon^\star) = 0$, which is exactly (iii).

$(iii) \Rightarrow (i)$ Let $\upsilon^\star \in \Upsilon$ satisfy $\max_{\xi\in\Xi} \phi_P(\xi, \upsilon^\star) = \varsigma_P(\upsilon^\star)$. Setting $\upsilon_\alpha = (1-\alpha)\upsilon^\star + \alpha\bar{\upsilon}$ for $\bar{\upsilon} \in \Upsilon$, from Lemma 10 we obtain

$$\left.\frac{\partial\Psi[M(\upsilon_\alpha)]}{\partial\alpha}\right|_{\alpha=0^+} = \varsigma_P(\upsilon^\star) - \int_\Xi \phi_P(\xi, \upsilon^\star)\,\bar{\upsilon}(\mathrm{d}\xi) \geq \varsigma_P(\upsilon^\star) - \max_{\xi\in\Xi} \phi_P(\xi, \upsilon^\star) = 0, \qquad (47)$$

which implies the optimality of $\upsilon^\star$.

The unicity of the information matrix for each optimal design follows from the convexity of the set $\mathfrak{M}(\Xi)$ and the strict convexity of the function $\Psi : M \mapsto \Psi[M]$ (from classical optimization theory it is known that there exists at most one global minimum of a strictly convex function over a convex set).

References

Atkinson, A. C. and A. N. Donev (1992). *Optimum Experimental Design*. Oxford: Oxford University Press.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge: Cambridge University Press.

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

Fedorov, V. V. and P. Hackl (1997). *Model-Oriented Design of Experiments*. Lecture Notes in Statistics. New York: Springer-Verlag.

Jones, B. and J. Wang (1999). Constructing optimal designs for fitting pharmacokinetic models. *Statistics and Computing 9*, 209–218.

Jonsson, E. N., J. R. Wade, and M. O. Karlsson (1996). Comparison of some practical sampling strategies for population pharmacokinetic studies. *Journal of Pharmacokinetics and Pharmacodynamics 24*(2), 245–263.

Kiefer, J. and J. Wolfowitz (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics 12*, 363–366.

Mentré, F., A. Mallet, and D. Baccar (1997). Optimal design in random-effects regression models. *Biometrica 84*(2), 429–442.

Patan, M. (2004). *Optimal Observation Strategies for Parameter Estimation of Distributed Systems*. Zielona Góra: Technical University Press.

Pázman, A. (1986). *Foundations of Optimum Experimental Design*. Mathematics and Its Applications. Dordrecht: D. Reidel Publishing Company.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. Probability and Mathematical Statistics. New York: John Wiley & Sons.

Rafajłowicz, E. (1996). *Algorithms of Experimental Design with Implementations in MATHEMATICA*. Warsaw: Academic Publishers PLJ. (in Polish).

Retout, S., S. Duffull, and F. Mentré (2001). Development and implementation of the population Fisher information matrix for the evaluation of population pharmacokinetic designs. *Computer methods and Programs in Biomedicine 65*, 141–151.

Retout, S. and F. Mentré (2003). Further developments of the Fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics. *Journal of Biopharmaceutical Statistics 13*(2), 209–227.

Retout, S., F. Mentré, and R. Bruno (2002). Fisher information matrix for non-linear mixed-effect models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. *Statistics in Medicine 21*, 2623–2639.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton: Princeton University Press.

Uciński, D. (2005). *Optimal Measurement Methods for Distributed Parameter System Identification*. Boca Raton: CRC Press.

Walter, É. and L. Pronzato (1997). *Identification of Parametric Models from Experimental Data*. Communications and Control Engineering. Berlin: Springer-Verlag.