

# An Optimal Design Criterion for Within-Individual Covariance Matrices Discrimination and Parameter Estimation in Nonlinear Mixed Effects Models

Un criterio de diseño óptimo para discriminación de matrices de covarianza intra-individual y estimación de parámetros en modelos de efectos mixtos no lineales

MARÍA EUGENIA CASTAÑEDA-LÓPEZ<sup>1,a</sup>, VÍCTOR IGNACIO LÓPEZ-RÍOS<sup>2,b</sup>

<sup>1</sup>INSTITUTO DE MATEMÁTICAS, FACULTAD DE CIENCIAS EXACTAS Y NATURALES, UNIVERSIDAD DE ANTIOQUIA, MEDELLÍN, COLOMBIA

<sup>2</sup>ESCUELA DE ESTADÍSTICA, FACULTAD DE CIENCIAS, UNIVERSIDAD NACIONAL DE COLOMBIA, MEDELLÍN, COLOMBIA

---

## Abstract

In this paper, we consider the problem of finding optimal population designs for within-individual covariance matrices discrimination and parameter estimation in nonlinear mixed effects models. A compound optimality criterion is provided, which combines an estimation criterion and a discrimination criterion. We used the  $D$ -optimality criterion for parameter estimation, which maximizes the determinant of the Fisher information matrix. For discrimination, we propose a generalization of the  $T$ -optimality criterion for fixed-effects models. Equivalence theorems are provided for these criteria. We illustrated the application of compound criteria with an example in a pharmacokinetic experiment.

**Key words:** Compound criteria;  $D$ -optimality; Mixed effects models; Optimal designs;  $T$ -optimality.

## Resumen

En este artículo se considera el problema de encontrar diseños óptimos poblacionales para discriminación entre matrices de covarianza intra-individual y estimación de parámetros en modelos de efectos mixtos no lineales. Se propone un criterio compuesto que combina un criterio para

---

<sup>a</sup>Ph.D. E-mail: [maria.castaneda@udea.edu.co](mailto:maria.castaneda@udea.edu.co)

<sup>b</sup>Ph.D. E-mail: [vilopez@unal.edu.co](mailto:vilopez@unal.edu.co)

estimación y otro para discriminación. Para estimación se usa el criterio de  $D$ -optimalidad el cual maximiza el determinante de la matriz de información de Fisher. Para discriminación se propone una generalización del criterio de  $T$ -optimalidad para modelos de efectos fijos. Para estos criterios se proporcionan los respectivos teoremas de equivalencia. La aplicación del criterio compuesto se ilustra con un ejemplo en un experimento de farmacocinética.

**Palabras clave:** Criterios compuestos; Diseños óptimos;  $D$ -optimalidad; Modelo de efectos mixtos;  $T$ -optimalidad.

## 1. Introduction

In the application of optimal design, one of the basic assumptions is to assume that the model used to describe a given phenomenon or process is the correct model. In practice, however, several candidate models may exist. One way of selecting the most appropriate model from among several candidates is by conducting an experiment designed in such a way that the observations obtained allow us to discriminate between the models in the best possible way. In addition, it is useful and advantageous that this design works well for estimating selected model parameters efficiently. This leads to the problem of finding the optimal experimental conditions by using optimality criteria for both model discrimination and model parameter estimation simultaneously.

In the case of fixed-effects models, a commonly used parameter estimation criterion is  $D$ -optimality, which maximizes the determinant of the variance-covariance matrix of estimated parameters (Atkinson, Donev & Tobias 2007). For models with normal errors, Atkinson & Fedorov (1975) proposed the  $T$ -optimality criterion for discriminating between two competing-response models, which provides the most powerful  $F$ -test for the lack of fit of one model when the other is assumed to be true. The criterion has been generalized for others fixed effects models (e.g., Uciński & Bogacka, 2005). López-Fidalgo, Tommasi & Trandafir (2007) proposed the  $KL$ -optimality criterion based on the Kullback-Leibler distance. This criterion is an extension of the  $T$ -optimality criterion and it can be used when the rival models are nested or not, homoscedastic or heteroscedastic, and with any error distribution. These criteria have been used in compound criteria for model discrimination and efficient parameter estimation (Atkinson, 2008 and Tommasi, 2009).

The nonlinear mixed-effects models are particularly useful in longitudinal studies, such as population pharmacokinetics experiments, assay analyses, and studies of growth in which a limited number of samples can be obtained from each individual. These models distinguish two classes of variation: random variation among observations within a given individual (within-individual) and random variation between-individuals (Davidian & Giltinan 1995). This separation of variability allows the estimation of population characteristics from sparse samples per individual in a set of subjects without requiring individual estimation of the parameters. The pattern of within-individual variability is not necessarily

homogeneous, and depending on the application and nature of the data, it is possible to incorporate heteroscedasticity by specification of a variance function. In the context of mixed effects models, a population experimental design is defined by the number of individuals to study and the individual designs to be performed in the individuals (number of samples and the sampling times) (Mentré, Burtin, Merlé, Bree, Mallet & Steimer 1995). Assuming that the response function and the between-individuals variability model are correctly specified, it may be of interest the problem of designing an experiment in groups of individuals with sparse samples and different sampling scheme, for discrimination between two competing within-individual covariance models and parameter estimation.

The most common approach to optimal population design is D-optimal (e.g., Mentré, Mallet & Baccar, 1997 and Gagnon & Leonov, 2005, among others). To discriminate between two models, some authors proposed extensions of T-optimality criterion (e.g., Waterhouse, Redmann, Duffull & Eccleston, 2005, Vajjah & Duffull, 2012 and Castañeda & López-Ríos, 2016). Kuczewski, Bogacka & Uciński (2008) proposed an extension of the T-optimality criterion for multivariate mixed models and it can be applied directly when all individuals are observed under the same experimental conditions. In contrast to the above problems, the model discrimination and parameter-estimation problem have not been developed in great detail for mixed models. Waterhouse, Redmann, Duffull & Eccleston (2005) proposed the product D-optimality criterion based on the product of the determinants of the Fisher-information matrices. For nonlinear models, however, such designs may be less efficient for discriminating than the T-optimal designs. Therefore, alternative methods for the dual purposes of model discrimination and parameter estimation are required.

In this paper, we consider the problem of finding optimal population designs for discrimination between two within-individual variance models and efficient parameter estimation. A compound optimality criterion is provided, which combines an estimation criterion and a discrimination criterion. The parameter estimation criterion we used is the D-optimality criterion, and for discrimination, we propose a generalization of the T-optimality criterion for fixed-effects models. Our approach can be applied to population studies for groups of individuals with sparse samples and a different sampling scheme, where each sampling scheme is a point in a finite space of admissible sampling sequences.

This paper is organized as follows: In Section 2, we present the nonlinear mixed-effects model, optimal design concepts, and D-optimum designs. In Section 3, we derive a generalization of the T-optimality criterion, and a necessary and sufficient condition for a design to be optimum is given. In Section 4, we propose a compound criterion for discrimination and estimation and provide an equivalence theorem. In Section 5, an illustrative example is presented. Finally, our conclusions and further work are detailed in Section 6.

## 2. Model and Optimal Population Design

### 2.1. Model

We assume that for each individual  $i$  in a sample of  $N$  individuals, from a larger population, the number of different available observations is  $n$ . Let  $\mathbf{y}_i = (y_{i1}, \dots, y_{in})^T$  be the vector of repeated measurements for the  $i$ th individual, and  $\mathbf{t}_i = (t_{i1}, \dots, t_{in})^T$  the  $n \times 1$  vector of sampling times where  $t_{ij}$  belongs to a finite set  $\mathcal{X}$ . It is assumed that measurements made on different subjects are independent. To model the relationship between  $\mathbf{y}_i$  and  $\mathbf{t}_i$ , we consider the nonlinear mixed model (Demidenko 2004):

$$\mathbf{y}_i = \mathbf{f}(\mathbf{t}_i, \boldsymbol{\beta}) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, N \quad (1)$$

where  $\boldsymbol{\beta}$  is a  $p \times 1$  vector of population parameters,  $\mathbf{f}(\mathbf{t}_i, \boldsymbol{\beta})$  is an  $n \times 1$  vector function,  $\mathbf{f}(\mathbf{t}_i, \boldsymbol{\beta}) = (f(t_{i1}, \boldsymbol{\beta}), \dots, f(t_{in}, \boldsymbol{\beta}))^T$  with  $f$  a known nonlinear function of  $\boldsymbol{\beta}$ ,  $\mathbf{Z}_i$  is an  $n \times q$  full-rank matrix of known constants,  $\mathbf{b}_i$  is a  $q \times 1$  vector of random effects associated with individual  $i$ , and  $\boldsymbol{\varepsilon}_i$  is the  $n \times 1$  within-individual errors vector. It is assumed that the  $\mathbf{b}_i$  are independent and normally distributed with mean  $\mathbf{0}$  and variance-covariance matrix  $\mathbf{D}$ , and that  $\mathbf{b}_i$  and  $\boldsymbol{\varepsilon}_i$  are independent. It is assumed that  $\boldsymbol{\varepsilon}_i \sim N_n(\mathbf{0}, \mathbf{G}(\mathbf{t}_i, \boldsymbol{\lambda}))$  where  $\mathbf{G}(\mathbf{t}_i, \boldsymbol{\lambda})$  is an  $(n \times n)$  matrix and is called the within-individual covariance matrix, which depends on the parameter  $\boldsymbol{\lambda} (a \times 1)$ . The matrix  $\mathbf{G}$  takes into account the nature of the variation among observations on a given individual and may be chosen in such a way that reflects the heterogeneity of variance, the correlation structure, or both. In this work, we assume uncorrelated observations and heterogeneous variance for the within-individual variability, so that  $\mathbf{G}(\mathbf{t}_i, \boldsymbol{\lambda})$  is a diagonal matrix with diagonal elements  $\text{Var}(\varepsilon_{ij}) = g(t_{ij}, \boldsymbol{\lambda})$ , where  $g$  is a variance function. We assume that the individuals share a common within-individual covariance pattern that only varies between them through the possibly different values of  $\mathbf{t}_i$ , and that this does not depend on the parameter  $\boldsymbol{\beta}$ . The distinct elements of the parameter vector  $\boldsymbol{\lambda}$  and the covariance matrix  $\mathbf{D}$  can be arranged in a single vector  $\boldsymbol{\gamma} \in \Gamma \subset \mathbb{R}^c$  of variance parameters of dimension  $c = a + q(q+1)/2$ . Under the assumption of independence and normality of the distributions of  $\mathbf{b}_i$  and  $\boldsymbol{\varepsilon}_i$ , the marginal distribution of  $\mathbf{y}_i$  is  $n$ -variate normal with mean vector and variance-covariance matrix given by

$$\begin{aligned} \mathbb{E}(\mathbf{y}_i) &= \mathbf{f}(\mathbf{t}_i, \boldsymbol{\beta}) \\ \text{Cov}(\mathbf{y}_i) &= \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{G}(\mathbf{t}_i, \boldsymbol{\lambda}) \\ &\equiv \boldsymbol{\Sigma}(\mathbf{t}_i, \boldsymbol{\gamma}) \end{aligned} \quad (2)$$

In this model, the full parameters vector to be estimated is  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T \in \Theta \subset \mathbb{R}^d$  of dimension  $d = p + c$ .

### 2.2. Population Designs

Suppose that the sample of  $N$  individuals consists of  $s$  groups each of size  $n_k$  and that the individuals in the same group are observed under the conditions

vector  $\mathbf{t}_k = (t_{k1}, \dots, t_{kn})^T$ . The collection of  $\mathbf{t}_k$  and  $n_k$ , represented by

$$\left\{ \begin{matrix} \mathbf{t}_1, & \dots, & \mathbf{t}_s \\ n_1, & \dots, & n_s \end{matrix} \right\} = \{\mathbf{t}_k, n_k\}_1^s, \sum_{k=1}^s n_k = N \tag{3}$$

is a population design. The set  $\mathcal{X}^n$  is the design region, and points  $\mathbf{t}_k$  are the design points. The collection  $\zeta_N = \{\mathbf{t}_k, \omega_k\}_1^s$  where  $\omega_k = \frac{n_k}{N}$  is the proportion of subjects allocated to each group is the normalized or exact population design with weights vector  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_s)$  (Gagnon & Leonov 2005). For fixed values of the total number of individuals  $N$  and the number of sampling times  $n$ , the population optimal design problem consists of finding the design by a choice of distinct values for the sampling times vector  $\mathbf{t} \in \mathcal{X}^n$  and values for the number subjects assigned to vector  $\mathbf{t}$  in such a way that the resulting design maximizes some optimality criteria. The most commonly used optimality criteria usually depend on the unknown model parameter. One approach is to construct locally optimal designs, which require specifying a prior parameter estimate, then addressing the optimization problem for this specific value (Chernoff 1953). In this article, the approximate design  $\zeta$ , in which the weights  $\omega_k$  may be any real numbers from the interval  $[0, 1]$ , is considered. The set of points  $\mathbf{t}_k$  in the design region  $\mathcal{X}^n$ , for which the design  $\zeta$  has nonzero weights  $\omega_k$ , is the support set of  $\zeta$ .

### 2.3. D-optimality Criterion

D-optimality is the most widely used criterion for efficient parameter estimation, which determines a design  $\zeta$  maximizing

$$\Phi_D(M(\zeta, \boldsymbol{\theta})) = \log \det M(\zeta, \boldsymbol{\theta})$$

where  $M(\zeta, \boldsymbol{\theta})$  is the Fisher information matrix for the population design  $\zeta$  which can be expressed as the sum of the Fisher information matrix  $M(\mathbf{t}_k, \boldsymbol{\theta})$  for each group with design  $\mathbf{t}_k$ ,  $M(\zeta, \boldsymbol{\theta}) = \sum_{k=1}^s \omega_k M(\mathbf{t}_k, \boldsymbol{\theta})$ . The matrix  $M(\mathbf{t}_k, \boldsymbol{\theta})$  is given by  $-E \left[ \frac{\partial^2 \ell(\boldsymbol{\theta}; \mathbf{y}_i)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right]$ , where  $\ell(\boldsymbol{\theta}; \mathbf{y}_i)$  is the log-likelihood of the vector of observation  $\mathbf{y}_i$  for the population parameter  $\boldsymbol{\theta}$ . This matrix can be calculated using the general formula for observations with normal distribution,  $M(\mathbf{t}_k, \boldsymbol{\theta}) = \{M_{uv}(\mathbf{t}_k, \boldsymbol{\theta})\}_{u,v=1}^d$  with

$$\begin{aligned} M_{uv}(\mathbf{t}_k, \boldsymbol{\theta}) &= \frac{\partial \mathbf{f}(\mathbf{t}_k, \boldsymbol{\theta})^T}{\partial \theta_u} \boldsymbol{\Sigma}^{-1}(\mathbf{t}_k, \boldsymbol{\theta}) \frac{\partial \mathbf{f}(\mathbf{t}_k, \boldsymbol{\theta})}{\partial \theta_v} \\ &+ \frac{1}{2} \text{tr} \left[ \boldsymbol{\Sigma}^{-1}(\mathbf{t}_k, \boldsymbol{\theta}) \frac{\partial \boldsymbol{\Sigma}(\mathbf{t}_k, \boldsymbol{\theta})}{\partial \theta_u} \boldsymbol{\Sigma}^{-1}(\mathbf{t}_k, \boldsymbol{\theta}) \frac{\partial \boldsymbol{\Sigma}(\mathbf{t}_k, \boldsymbol{\theta})}{\partial \theta_v} \right] \end{aligned} \tag{4}$$

For the class of approximate population designs  $\Xi$ , a population design  $\zeta^* \in \Xi$  with a nonsingular Fisher information matrix  $M(\zeta, \boldsymbol{\theta})$  is a  $D$ -optimum design for the parameter estimation  $\boldsymbol{\theta}$ , if and only if  $\phi_D(\mathbf{t}, \zeta^*) \leq 0, \forall \mathbf{t} \in \mathcal{X}^n$ , where

$$\phi_D(\mathbf{t}, \zeta) = \text{tr} [M(\mathbf{t}, \boldsymbol{\theta})M^{-1}(\zeta, \boldsymbol{\theta})] - d, \quad d = \dim(\boldsymbol{\theta}) \tag{5}$$

is the directional derivative of the criterion  $\Phi_D(M(\zeta, \boldsymbol{\theta}))$  at  $\zeta$  in the direction of  $\delta_{\zeta_t} = \zeta_t - \zeta$ , and  $\zeta_t$  is the design which puts the whole mass at point  $t$  (Gagnon & Leonov 2005). The  $D$ -efficiency of a design  $\zeta$  relative to the optimum design  $\zeta_D^*$  is  $\text{Eff}_D(\zeta) = \left[ \frac{\det M(\zeta, \boldsymbol{\theta})}{\det M(\zeta^*, \boldsymbol{\theta})} \right]^{1/d}$ . This efficiency is a number in  $(0,1)$  which measures the goodness of a design  $\zeta$  for estimating purposes.

### 3. Discrimination Between Two Within-Individual Variance Models

In this section, we propose a generalization of  $T$ -optimality criterion for discrimination between two nested within-individual variance models. The criterion determines a design that maximize the power of likelihood ratio test when the largest model is assumed to be the true model. Our approach can be applied to population studies for groups of individuals with sparse samples and a different sampling scheme.

Let  $\mathbf{G}_1(t, \boldsymbol{\lambda}_1)$  and  $\mathbf{G}_2(t, \boldsymbol{\lambda}_2)$  be two models for the within-individual variability where one model is nested within the other. For example, if  $\mathbf{G}_2$  is nested within of  $\mathbf{G}_1$ , it would mean that both models involve the same structure  $\mathbf{G}(t, \boldsymbol{\lambda})$ , and the parameter space  $\Gamma_2$  of  $\mathbf{G}_2$  is a subset of the parameter space  $\Gamma_1$  of  $\mathbf{G}_1$ , which is defined by the imposition of  $\kappa$  equality constraints. In order to discriminate between these models, we propose the following criterion: assuming that  $\mathbf{G}_1$  is the true model for a known parameter vector  $\boldsymbol{\gamma}_1^0$ , a locally optimum approximate population design  $\zeta^*$  to discriminate between  $\mathbf{G}_1$  and  $\mathbf{G}_2$  is such that

$$\zeta^* = \arg \max_{\zeta} T_G(\zeta) \quad (6)$$

where

$$T_G(\zeta) = \min_{\boldsymbol{\gamma}_2 \in \Gamma_2} \sum_{k=1}^s \omega_k F(\mathbf{t}_k, \boldsymbol{\gamma}_2) \quad (7)$$

with

$$F(\mathbf{t}, \boldsymbol{\gamma}_2) = \text{tr}[\boldsymbol{\Sigma}_1(\mathbf{t}, \boldsymbol{\gamma}_1^0) \boldsymbol{\Sigma}_2^{-1}(\mathbf{t}, \boldsymbol{\gamma}_2)] - \log \det[\boldsymbol{\Sigma}_1(\mathbf{t}, \boldsymbol{\gamma}_1^0) \boldsymbol{\Sigma}_2^{-1}(\mathbf{t}, \boldsymbol{\gamma}_2)] \quad (8)$$

and  $\boldsymbol{\Sigma}_r(\mathbf{t}, \boldsymbol{\gamma})$ ,  $r = 1, 2$ , is given by (2). The design  $\zeta^*$  be called  $T_G$ -optimal.

The justification of criterion is given below. Without loss of generality, consider the exact population design  $\zeta_N = \{\mathbf{t}_i, \omega_i\}_1^N$  where  $\omega_i = \frac{1}{N}$ . To discriminate between  $\mathbf{G}_1$  and  $\mathbf{G}_2$  consider the problem of testing the hypothesis:

$$H_0 : \mathbf{G}(\mathbf{t}_i, \boldsymbol{\lambda}) = \mathbf{G}_2(\mathbf{t}_i, \boldsymbol{\lambda}_2) \text{ vs } H_1 : \mathbf{G}(\mathbf{t}_i, \boldsymbol{\lambda}) = \mathbf{G}_1(\mathbf{t}_i, \boldsymbol{\lambda}_1) \quad (9)$$

Since  $\mathbf{y}_i$  are  $N_n(\mathbf{f}(\mathbf{t}_i, \boldsymbol{\beta}), \boldsymbol{\Sigma}(\mathbf{t}_i, \boldsymbol{\gamma}))$  random vectors, the likelihood function under the model  $\mathbf{G}_r$  is

$$L_r = \prod_{i=1}^N (2\pi)^{-n/2} (\det \boldsymbol{\Sigma}_r(\mathbf{t}_i, \boldsymbol{\gamma}_r))^{-1/2} \times \exp \left[ \frac{1}{2} (\mathbf{y}_i - \mathbf{f}(\mathbf{t}_i, \boldsymbol{\beta}))^T \boldsymbol{\Sigma}_r^{-1}(\mathbf{t}_i, \boldsymbol{\gamma}_r) (\mathbf{y}_i - \mathbf{f}(\mathbf{t}_i, \boldsymbol{\beta})) \right] \quad (10)$$

Thus, the likelihood-ratio test function usually associated to the hypotheses (9) is given by

$$2 \ln \frac{L_1}{L_2} = \sum_{i=1}^N \{ -\ln \det[\boldsymbol{\Sigma}_2^{-1}(\mathbf{t}_i, \boldsymbol{\gamma}_2) \boldsymbol{\Sigma}_1(\mathbf{t}_i, \boldsymbol{\gamma}_1)] + (\mathbf{y}_i - \mathbf{f}(\mathbf{t}_i, \boldsymbol{\beta}))^T \boldsymbol{\Sigma}_2^{-1}(\mathbf{t}_i, \boldsymbol{\gamma}_2) (\mathbf{y}_i - \mathbf{f}(\mathbf{t}_i, \boldsymbol{\beta})) - (\mathbf{y}_i - \mathbf{f}(\mathbf{t}_i, \boldsymbol{\beta}))^T \boldsymbol{\Sigma}_1^{-1}(\mathbf{t}_i, \boldsymbol{\gamma}_1) (\mathbf{y}_i - \mathbf{f}(\mathbf{t}_i, \boldsymbol{\beta})) \}$$

If the model  $\mathbf{G}_1$  is the true model for a specific alternative value  $\boldsymbol{\gamma}_1^0 \in \Gamma_1$  and  $\boldsymbol{\gamma}_2^*$  is an estimate for  $\boldsymbol{\gamma}_2$ , then

$$\begin{aligned} 2E \left[ \ln \frac{L_1}{L_2} \right] &= \sum_{i=1}^N \{ \text{tr}[\boldsymbol{\Sigma}_1(\mathbf{t}_i, \boldsymbol{\gamma}_1^0) \boldsymbol{\Sigma}_2^{-1}(\mathbf{t}_i, \boldsymbol{\gamma}_2^*)] - \log \det[\boldsymbol{\Sigma}_1(\mathbf{t}_i, \boldsymbol{\gamma}_1^0) \boldsymbol{\Sigma}_2^{-1}(\mathbf{t}_i, \boldsymbol{\gamma}_2^*)] - n \} \\ &= N \left\{ \sum_{i=1}^N \frac{1}{N} F(\mathbf{t}_i, \boldsymbol{\gamma}_2^*) - n \right\} \\ &= N \{ T_G(\zeta_N) - n \} \end{aligned} \quad (11)$$

From this expression, the power of the likelihood ratio test for the hypotheses (9) increases with  $T_G(\zeta_N)$  and hence can be maximized by the choice of design  $\zeta_N$ .

Finally, the exact design  $\zeta_N$  can be replaced by the corresponding approximate design  $\zeta$ , thus we obtain the  $T_G$ -criterion defined in (7).

In order to provide precise conditions for checking whether a particular design is  $T_G$ -optimum, we say that a design  $\zeta$  is a  $T_G$ -regular design if the set

$$\Gamma_2(\zeta) = \left\{ \tilde{\boldsymbol{\gamma}}_2 : \tilde{\boldsymbol{\gamma}}_2(\zeta) = \arg \min_{\boldsymbol{\gamma}_2 \in \Gamma_2} \sum_{k=1}^s \omega_k F(\mathbf{t}_k, \boldsymbol{\gamma}_2) \right\}$$

is singleton; otherwise it is a  $T_G$ -singular design. Hence, if  $\zeta$  is a  $T_G$ -regular design, and  $\tilde{\boldsymbol{\gamma}}_2 \in \Gamma_2(\zeta)$  then  $\tilde{\boldsymbol{\gamma}}_2$  is the unique solution of the equation

$$\sum_{k=1}^s \omega_k F(\mathbf{t}_k, \tilde{\boldsymbol{\gamma}}_2) = \min_{\boldsymbol{\gamma}_2 \in \Gamma_2} \sum_{k=1}^s \omega_k F(\mathbf{t}_k, \boldsymbol{\gamma}_2)$$

Assuming that  $\Gamma_2$  is a compact set and  $\boldsymbol{\Sigma}_2(\mathbf{t}, \boldsymbol{\gamma}_2)$  is a twice continuously differentiable in  $\Gamma_2$ , the equivalence theorem for  $T_G$ -optimality criterion can be formulated as follows. The proof of this theorem is similar to the proof of Theorem 1 in Castañeda & López-Ríos (2016).

**Theorem 1.** Let  $\zeta^*$  be a  $T_G$ -regular design.

(i) A necessary and sufficient condition for the design  $\zeta^*$  to be  $T_G$ -optimal is

$$\phi_{T_G}(\mathbf{t}, \zeta^*) \leq 0, \quad \forall \mathbf{t} \in \mathcal{X}^n \quad (12)$$

where

$$\phi_{T_G}(\mathbf{t}, \zeta) = F(\mathbf{t}, \tilde{\gamma}_2) - T_G(\zeta), \quad \tilde{\gamma}_2 \in \Gamma_2(\zeta) \quad (13)$$

is the directional derivative of  $T_G(\zeta)$  at  $\zeta$  in the direction of  $\delta_{\zeta_t} = \zeta_t - \zeta$ , and  $\zeta_t$  is the design putting all mass at the point  $\mathbf{t} \in \mathcal{X}^n$ .

(ii) The function  $\phi_{T_G}(\mathbf{t}, \zeta^*)$  achieves its maximum value at the support points of the optimal design  $\zeta^*$ .

(iii) The set of  $T_G$ -optimum designs is convex.

A measure of efficiency of a design  $\zeta$  relative to a  $T_G$ -optimum design is  $\text{Eff}_{T_G}(\zeta) = \frac{T_G(\zeta)}{T_G(\zeta^*)}$ .

## 4. $DT_G^\alpha$ -optimality Criterion

In order to provide population designs for both discrimination and efficient parameters estimation, we propose a compound criterion based on the combination of  $D$ -optimality for each model and  $T_G$ -optimality, specifically the criteria used are:

- $\Phi_1(\zeta) = T_G$  for discrimination assuming that  $\mathbf{G}_1$  is the true model
- $\Phi_2(\zeta) = \det M_1(\zeta)^{1/d_1}$  for parameter estimation  $\theta_1$
- $\Phi_3(\zeta) = \det M_2(\zeta)^{1/d_2}$  for parameter estimation  $\theta_2$

where  $M_r(\zeta)$  is the information matrix under the model  $\mathbf{G}_r$  for some known  $\theta_r^0$ .

We adopted the approach proposed by López (2008) which consists of maximizing the geometric mean of the efficiencies for a given design  $\zeta$  with respect to each criterion, weighted by a predefined constant  $\alpha$ , where  $0 \leq \alpha \leq 1$ . We give equal weighting to the estimation criteria and other weighting for discrimination. Thus, assuming that the model  $\mathbf{G}_1$  is the true model, we propose the following criterion:

$$\Phi_\alpha^{DT_G}(\zeta) = \left\{ \frac{\Phi_1(\zeta)}{\Phi_1(\zeta_1^*)} \right\}^\alpha \left\{ \frac{\Phi_2(\zeta)}{\Phi_2(\zeta_2^*)} \frac{\Phi_3(\zeta)}{\Phi_3(\zeta_3^*)} \right\}^{\frac{1-\alpha}{2}} \quad (14)$$

where  $\zeta_j^* = \arg \max_{\zeta \in \Xi} \Phi_j(\zeta)$ . Taking the logarithm of (14), we obtain

$$\ln \Phi_\alpha^{DT_G}(\zeta) = \alpha \ln T_G(\zeta) + \frac{1-\alpha}{2} \left\{ \frac{1}{d_1} \ln \det M_1(\zeta) + \frac{1}{d_2} \ln \det M_2(\zeta) \right\} + C \quad (15)$$

where  $C$  is a constant that does not depend on the design  $\zeta$ . The approximate design that maximizes the criterion (15) over the set  $\Xi$  is  $DT_G^\alpha$ -optimum.

The design criterion (15) is concave because it is a convex combination of concave functions. Therefore, the  $DT_G^\alpha$ -design criterion satisfies the conditions of convex optimum design theory, and the following equivalence theorem may be stated:

**Theorem 2.** *Let  $\zeta^*$  be a  $T_G$ -regular design with a nonsingular Fisher information matrix  $M_r(\zeta)$ ,  $r = 1, 2$ . A necessary and sufficient condition for the design  $\zeta^*$  to be  $DT_G^\alpha$ -optimal is*

$$\phi_{DT_G^\alpha}(\mathbf{t}, \zeta^*) \leq 0, \quad \forall \mathbf{t} \in \mathcal{X}^n, \tag{16}$$

where

$$\begin{aligned} \phi_{DT_G^\alpha}(\mathbf{t}, \zeta) &= \alpha \frac{F(\mathbf{t}, \tilde{\gamma}_2)}{T_G(\zeta)} + \frac{1-\alpha}{2} \left\{ \frac{1}{d_1} \operatorname{tr} [M_1(\mathbf{t})M_1^{-1}(\zeta)] \right. \\ &\quad \left. + \frac{1}{d_2} \operatorname{tr} [M_2(\mathbf{t})M_2^{-1}(\zeta)] \right\} - 1, \quad \tilde{\gamma}_2 \in \Gamma_2(\zeta) \end{aligned} \tag{17}$$

is the directional derivative of criterion function at  $\zeta$  in the direction of  $\delta_{\zeta_t}$ . The function  $\phi_{DT_G^\alpha}(\mathbf{t}, \zeta^*)$  achieves its maximum value at the support points of the optimal design  $\zeta^*$ .

A measure of the efficiency of a design  $\zeta$  relative to a  $DT_G^\alpha$ -optimum design is  $\operatorname{Eff}_{DT_G^\alpha}(\zeta) = \frac{DT_G^\alpha(\zeta)}{DT_G^\alpha(\zeta^*)}$ .

## 5. An Example

To illustrate the proposed criterion, we consider the data set presented in Racine, Grieve, Fluhler & Smith (1986) on the plasma concentrations of cadralazine in 10 German cardiac- failure patients who were observed between 2 and 32 hours after a single 30mg dose of cadralazine. We propose to fit the nonlinear mixed model shown above where only the intercept is random; thus, for simplicity,  $\mathbf{Z}_i = \mathbf{1}$  for all  $i$ . This model is

$$\begin{aligned} y_{il} &= f(t_{il}, \boldsymbol{\beta}_i) + b_i + \varepsilon_{il} \\ f(t_{il}, \boldsymbol{\beta}_i) &= \frac{D}{\beta_1} \exp(-\beta_2 t_{il}) \end{aligned} \tag{18}$$

where  $b_i$  is the random effect associated with individual  $i$ . We assume that  $b_i \sim N(0, \psi)$  and  $\varepsilon_i \sim N_n(\mathbf{0}, \mathbf{G}(\mathbf{t}_i, \boldsymbol{\lambda}))$ . Thus, the full parameters vector to be estimate is  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$  where  $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$  and  $\boldsymbol{\gamma} = (\psi, \boldsymbol{\lambda}^T)^T$ .

We consider two alternative models for the within-individual covariance matrix  $\mathbf{G}(\mathbf{t}_i, \boldsymbol{\lambda})$ :

- (1) Exponential model

$$\mathbf{G}_1(\mathbf{t}_i, \boldsymbol{\lambda}) = \sigma^2 \operatorname{diag}(\exp(\delta t_{i1}), \dots, \exp(\delta t_{in})), \quad \boldsymbol{\lambda} = (\sigma^2, \delta)^T$$

(2) Constant variance:  $\mathbf{G}_2(\mathbf{t}_i, \boldsymbol{\lambda}) = \sigma^2 \mathbf{I}_n$ ,  $\boldsymbol{\lambda} = \sigma^2$ .

We fit the model (18) with each within-individual covariance structure in R Development Core Team (2014) using the `nlme` function for fitting nonlinear mixed-effects models (Pinheiro & Bates 2000). The results of the estimation are:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_1 &= (14.93, 0.21)^T, \hat{\psi}_1 = 0.23^2, \hat{\sigma}_1^2 = 0.08^2, \hat{\delta} = 0.02 \\ \hat{\boldsymbol{\beta}}_2 &= (15.01, 0.19)^T, \hat{\psi}_2 = 0.20^2, \hat{\sigma}_2^2 = 0.10^2\end{aligned}\quad (19)$$

Since the model  $\mathbf{G}_2$  is nested within  $\mathbf{G}_1$ , the true model is  $\mathbf{G}_1$ . We use the estimates of the parameters obtained previously as the local value of  $\boldsymbol{\theta}$  to find the locally optimal designs. Thus,

$$\begin{aligned}\boldsymbol{\theta}_1^0 &= \left( \boldsymbol{\beta}_1^{0T}, \boldsymbol{\lambda}_1^{0T} \right)^T = (14.93, 0.21, 0.23^2, 0.08^2, 0.02)^T \\ \boldsymbol{\theta}_2^0 &= \left( \boldsymbol{\beta}_2^{0T}, \boldsymbol{\lambda}_2^{0T} \right)^T = (15.01, 0.19, 0.20^2, 0.10^2)^T\end{aligned}$$

The times in  $\mathcal{X}$  was taken of the set of measurement times for the experiment:

$$\mathcal{X} = \{2, 4, 6, 8, 10, 24\} \text{ hours}$$

We assume that two observations are available for each patient  $n = 2$ , which is a sparse sampling situation in a pharmacokinetic study. Therefore, the design region is given by the set of combinations of two sampling times from  $\mathcal{X}$ ; that is,  $\mathcal{X}^2 = \{\mathbf{t} = (t_1, t_2); t_j \in \mathcal{X}\}$  contains 15 elements. The  $DT_G^\alpha$ -optimum design was calculated to optimize the respective criterion implemented through an algorithm in R Development Core Team (2014) for different values of  $\alpha$  (0.25, 0.50, 0.75). The function `nlminb` was used for the optimization in the design region  $\mathcal{X}^2$ . Additionally we find the  $D$ -,  $T_G$ , and product  $D$ -optimum designs ( $\zeta_P^*$ ) to obtain an overview of the individual designs. The computing times took approximately 1 hour for the compound designs. The computation complexity depends on the size of the design region and the number of parameters.

The designs obtained were:

$$\begin{aligned}\zeta_{D_1}^* &= \left\{ \begin{array}{cccc} (2, 6) & (2, 8) & (2, 24) & (6, 24) \\ 0.06 & 0.37 & 0.35 & 0.22 \end{array} \right\} \\ \zeta_{D_2}^* &= \left\{ \begin{array}{cc} (2, 6) & (2, 8) \\ 0.42 & 0.58 \end{array} \right\} \\ \zeta_{T_G}^* &= \left\{ \begin{array}{c} (2, 24) \\ 1 \end{array} \right\}\end{aligned}$$

$$\zeta_P^* = \begin{Bmatrix} (2, 8) & (2, 24) & (6, 24) \\ 0.68 & 0.19 & 0.13 \end{Bmatrix}$$

$$\zeta_{DT_G^{0.25}}^* = \begin{Bmatrix} (2, 6) & (2, 24) & (6, 24) \\ 0.29 & 0.47 & 0.24 \end{Bmatrix}$$

$$\zeta_{DT_G^{0.50}}^* = \begin{Bmatrix} (2, 6) & (2, 24) & (6, 24) \\ 0.11 & 0.66 & 0.23 \end{Bmatrix}$$

$$\zeta_{DT_G^{0.75}}^* = \begin{Bmatrix} (2, 24) & (6, 24) \\ 0.84 & 0.16 \end{Bmatrix}$$

To check the optimality of the calculated compound designs, we use the Equivalence Theorem. First, we enumerate all candidate sampling sequences (i.e., the elements of  $\mathcal{X}^2$ ) and calculate the sensitivity function for each sampling sequence. Then, we plot the sensitivity function as a function of index  $i$ . The resulting plot for each design is shown in Figure 1.

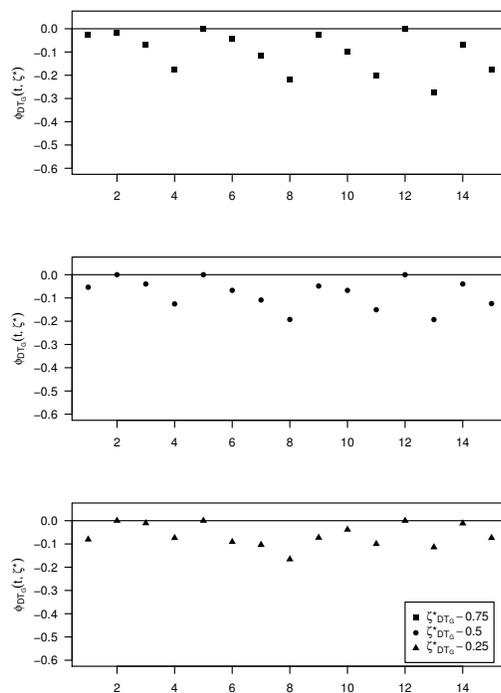


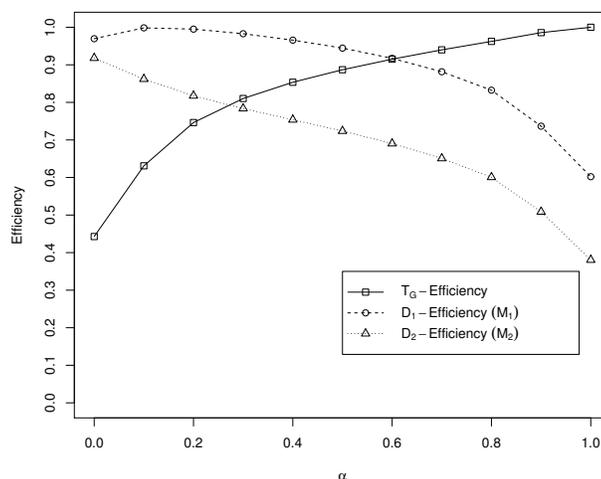
FIGURE 1: Sensitivity function for compound designs.

Table 1 presents the efficiencies of different optimum designs with respect to the  $D$ - and  $T_G$ -optimum designs. The first and second column list the efficiencies of the optimum designs with respect to the  $D$ -optimum designs and corresponds to  $G_1$  and  $G_2$  models, respectively. These efficiencies measure the goodness of

optimum designs for parameter estimation. The  $T_G$ -optimum design does not seem good for estimation purposes; for the  $G_2$  especially, the efficiency is approximately 38%. On the other hand, the efficiencies of  $\zeta_P$ - and  $DT_G^\alpha$ -optimal designs are always high; the worst efficiency ( $\alpha = 0.75$ ) is approximately 62%. The third column presents the efficiencies of the optimum designs with respect to the  $T_G$ -optimum design, which measure the goodness of a design for model discrimination. For the  $DT_G^\alpha$ -optimal designs, these efficiencies seem good; the lowest efficiency ( $\alpha = 0.25$ ) is approximately 78%. The product-optimum design does not seem to be good for discrimination purposes; the efficiency is approximately 44%. Thus, the best strategy for both parameter estimation and model discrimination purposes seems to be the  $DT_G^\alpha$ -optimality criterion. In order to choose an adequate value  $\alpha$  to successfully discrimination and estimation, we compute the efficiencies for  $D$ - and  $T_G$ -optimality for different values of  $\alpha$ . Figure 2 graphically displays these efficiencies. For  $\alpha$  approximately 0.3, the efficiencies are high; therefore, the compound design  $DT_G^{0.3}$  is good for discriminating and estimating.

TABLE 1: D-efficiency and  $T_G$ -efficiency.

Design	$\text{Eff}_{D_1}(\zeta)$	$\text{Eff}_{D_2}(\zeta)$	$\text{Eff}_{T_G}(\zeta)$
$\zeta_{D_1}^*$	1	0.8478	0.6644
$\zeta_{D_2}^*$	0.6715	1	0.0605
$\zeta_{T_G}^*$	0.6021	0.3804	1
$\zeta_P^*$	0.9695	0.9180	0.4428
$\zeta_{DT_G^{0.25}}^*$	0.9899	0.8009	0.7798
$\zeta_{DT_G^{0.50}}^*$	0.8869	0.7237	0.9734
$\zeta_{DT_G^{0.75}}^*$	0.8549	0.6230	0.9534

FIGURE 2: Plot of the efficiencies for different values of  $\alpha$ .

## 6. Conclusions

In this article, a compound optimality criterion for model discrimination and parameter estimation was proposed for two nonlinear mixed effects nested models with respect to the within-individual variance pattern. We used the  $D$ -optimality criterion for estimation, and we proposed a generalization of  $T$ -optimality criterion (the  $T_G$ -optimality criterion) for the problem of model discrimination. An equivalence theorem for the criterion was provided. To illustrate the use of proposed methodology, a pharmacokinetic example was presented. To show the good properties of the optimum compound design, the  $D$ -,  $T_G$ - and product  $D$ -optimum designs were computed; for discrimination, the efficiencies of the compound optimum designs were higher than the  $D$ - and product  $D$ -optimal designs. The design efficiencies for different values of the weight in the compound design  $\alpha$  show that the design for  $\alpha = 0.3$  is good for estimating and discriminating. Further development of the method for estimating the model parameters, some nonlinear functions of parameters, and simultaneously discrimination between models would also be interesting. In this work, we considered nonlinear mixed model with normality of both within-subject random error and random-effects, however, there is often interest in models with nonnormal distribution. Thus, the  $KL$ -optimality criterion could be an interesting methodology to explore in this class of models. Finally, it should be noted that we have found local designs, but the other option is the use of the Bayesian approach.

[Recibido: septiembre de 2019 — Aceptado: abril de 2020]

## References

- Atkinson, A. C. (2008), ‘DT-Optimum Designs for Model Discrimination and Parameter Estimation’, *Journal of Statistical Planning and Inference* **138**, 56–64.
- Atkinson, A. C., Donev, A. N. & Tobias, R. D. (2007), *Optimum Experimental Designs, with SAS*, first edn, Oxford, New York.
- Atkinson, A. C. & Fedorov, V. V. (1975), ‘The Design of Experiments for Discriminating Between Two Rival Models’, *Biometrika* **62**(1), 57–70.
- Castañeda, M. E. & López-Ríos, V. I. (2016), ‘Optimal Population Designs for Discrimination Between Two Nested Nonlinear Mixed Effects Models’, *Ciencia en Desarrollo* **7**(1), 71–81.
- Chernoff, H. (1953), ‘Locally Optimal Designs for Estimating Parameters’, *The Annals of Mathematical of Statistics* **24**(4), 586–602.
- Davidian, M. & Giltinan, D. M. (1995), *Nonlinear Models for Repeated Measurement Data*, first edn, Chapman & Hall, London.

- Demidenko, E. (2004), *Mixed Models: Theory and Applications*, first edn, John Wiley & Sons, Inc, New York.
- Gagnon, R. & Leonov, S. (2005), ‘Optimal Population Designs for PK Models with Serial Sampling’, *Journal of Biopharmaceutical Statistics* **15**(1), 143–163.
- Kuczewski, B., Bogacka, B. & Uciński, D. (2008), ‘Optimum Designs for Discrimination Between Two Nonlinear Multivariate Dynamic Mixed-Effects Models’, *Biometrical Letters* **45**(1), 1–28.
- López-Fidalgo, J., Tommasi, C. & Trandafir, P. C. (2007), ‘An Optimal Experimental Design Criterion for Discriminating Between Non-Normal Models’, *Journal of the Royal Statistical Society: Series B* **69**, 231–242.
- López, V. I. (2008), Diseños óptimos para estimación y discriminación en modelos no lineales, Tesis de Doctorado, Centro de Investigación en Matemáticas, CIMAT, México.
- Mentré, F., Burtin, P., Merlé, Y., Bree, J., Mallet, A. & Steimer, J. (1995), ‘Sparse-Sampling Optimal Designs in Pharmacokinetics and Toxicokinetics’, *Drug Information Journal* **29**(3), 997–1019.
- Mentré, F., Mallet, A. & Baccar, D. (1997), ‘Optimal Designs in Random-Effects Regression Models’, *Biometrika* **84**(2), 429–442.
- Pinheiro, J. C. & Bates, D. M. (2000), *Mixed-Effects Models in S and S-PLUS*, first edn, Springer-Verlag, New York.
- R Development Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.  
\*<http://www.R-project.org>
- Racine, A., Grieve, A., Fluhler, H. & Smith, A. (1986), ‘Bayesian Methods in Practice Experiences in the Pharmaceutical Industry’, *Journal of the Royal Statistical Society* **35**(2), 93–150.
- Tommasi, C. (2009), ‘Optimal Designs for Both Model Discrimination and Parameter Estimation’, *Journal of Statistical Planning and Inference* **139**, 4123–4132.
- Uciński, D. & Bogacka, B. (2005), ‘T-Optimum Designs for Discrimination Between Two Multiresponse Dynamic Models’, *Journal of the Royal Statistical Society: Series B* **67**, 3–18.
- Vajjah, P. & Duffull, S. (2012), ‘A Generalisation of T-optimality for Discriminating Between Competing Models With an Application to Pharmacokinetics Studies’, *Pharmaceuticals Statistics* **11**(6), 503–510.

- Waterhouse, T., Redmann, S., Duffull, S. & Eccleston, J. (2005), 'Optimal Design for Model Discrimination and Parameter Estimation for Itraconazole Population Pharmacokinetics in Cystic Fibrosis Patients', *Journal of Pharmacokinetics and Pharmacodynamics* **32**(3), 521–545.