Statistica Sinica Preprint No: SS-2023-0284	
Title	Locally Optimal Designs for Estimating One or More Functions of Shared Parameters Between Two Groups in Biomedical Studies
Manuscript ID	SS-2023-0284
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0284
Complete List of Authors	Xin Liu, Rong-Xian Yue and Weng Kee Wong
Corresponding Authors	Weng Kee Wong
E-mails	wkwong@ucla.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0284

Locally Optimal Designs for Estimating One or More Functions of Shared Parameters Between Two Groups in Biomedical Studies

Xin Liu[1] Rong-Xian Yue[2,3] and Weng Kee Wong[4]

[1] *Donghua University,* [2] *Shanghai Normal University* ,

[3] *Fuyao University of Science and Technology and* [4] *University of California, Los Angeles*

Abstract: Models with shared parameters arise quite naturally in the biological sciences and we use optimal design theory to construct c-optimal approximate designs for estimating one or more functions of the model parameters in two regression models with shared parameters. We assume sample sizes for the two groups are fixed and establish equivalence theorems to confirm the optimality of the design. As applications, we consider the parallel dose response model, the EMAX model and the Exponential model, each with shared parameters. The methodology is general and can be applied to other models or design problems. For example, we show the theoretical framework can be directly extended to the case when we are interested to find a c-optimal design to estimate the mean difference between the expected responses at an extrapolated dose for a nonlinear model, or when the total sample size for the whole study is fixed, and we wish to determine the optimal proportions of observations to allocate to the two groups, or we have multivariate responses.

Key words and phrases: Approximate design, Equivalence theorem, Group comparison, L-optimal design.

1. Introduction

Group comparison is a basic and important problem across disciplines. A statistical model is usually used to model the outcome from the two or more groups and assess group differences. Statistical models with shared parameters are increasingly common and they arise quite naturally in various research areas. For instance, Chakraborty et al. (2016) employed dynamic regimens trials for precision medicine in a depression clinical trial using models with shared parameters. Models with shared parameters are also increasingly used in pharmaceutical studies, where experts believe that some parameters in the drug profiles have the same values when the drugs are in the same class.

There is some research work on constructing D-optimal designs analytically for a two-group experiments with shared parameters, see for example, Feller et al. (2017). D-optimal designs minimize the generalized variance of all the estimated parameters in the mean function and so they are best for estimating all the parameters in the mean response. For nonlinear models, the generalized variance depends on the unknown parameters which we want to estimate! To circumvent this issue, we assume some best guesses of the unknown parameters (nominal values) of the model parameters are available. When the unknown parameters are replaced by the nominal values, the generalized variance can be directly minimized. Such designs are locally D-optimal designs because they depend on the nominal values.

Sometimes, the goal of the study is to estimate a given function of the model parameters. In this case, c-optimal designs are more appropriate because they minimize the asymptotic variance of the estimated function of interest. Locally c-optimal designs for models with shared parameters have not been discussed before and their analytic construction is also generally more difficult to find than locally D-optimal designs. Such design problems arise frequently in practice. For example, in dose-response studies, there is often interest in estimating the ED_{50}, which is the dose for which one-half of its maximum response is attained. When the drug response from a drug is modeled along with a comparable drug, we may want to estimate one or more functions of parameters in the two models with shared parameters. The method developed here can be directly generalized for finding Bayesian c-optimal designs if a prior density for the shared parameters is available or a multiple-objective optimal design when the objective functions are convex functions of the information matrix (Cook and Wong, 1994).

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0284

Section 2 is background material and reviews the general construction and confirmation of c-optimal designs via an equivalence theorem and how they are extended to make inference on models with shared parameters. Section 3 demonstrates applications using common models in the biological sciences, including how to optimally allocate resources to each of the two groups. Section 4 provides results when there are several functions to estimate for models with shared parameters. Section 5 considers multi-response models and Section 6 concludes with a summary.

2. Background

This section provides the statistical setup, including notation, terminology, model specification and the concept of two types of optimal designs. We also review optimal design theory before we specialize to finding a design that can optimally estimate a given function of the shared parameters in the two regression models.

2.1 Models and optimal designs

Let y_{ijk} be the kth continuous outcome from the ith group at experimental condition $x_{ij} \in \mathcal{X}$ and $\mathcal{X}$ is a user-selected compact design space. Let the mean response from the ith group at the jth condition be $f(x_{ij}, \theta_i)$ and let

$$y_{ijk} = f(x_{ij}, \theta_i) + \varepsilon_{ijk}, \quad i = 1, 2; j = 1, \ldots, l_i; k = 1, \ldots, n_{ij}. \tag{2.1}$$

Here, $f(x_{ij}, \theta_i)$ is known apart from the unknown parameters θ_i and the ε_{ijk}'s are independent and normally distributed random variables each with mean 0 and variance σ^2. This means that for the ith group, observations are taken at l_i different experimental conditions $x_{i1}, \ldots, x_{il_i}$ and n_{ij} observations are taken at each x_{ij}. The x_{ij}'s are the support point of the design for the ith group and we denote them by $\mathrm{Supp}(\xi_i)$. The two vectors of parameters θ_1 and θ_2 are p-dimensional and describe the different dependence between the mean response and the explanatory variable or variables in the two groups. Let $n_i = \sum_{j=1}^{l_i} n_{ij}$ be the pre-determined number of observations in the ith group, $i = 1, 2$, and let $n = n_1 + n_2$ be the total sample size. Given an optimality criterion, our task is to find an exact optimal design with optimal values for $l_i, x_{ij}, n_{ij}, j = 1, \ldots, l_i, i = 1, 2$.

Under the common assumptions of regularity, the maximum likelihood estimates, say $\hat{\theta}_1, \hat{\theta}_2$ in both samples are asymptotically normally distributed with covariance matrix

$$cov(\hat{\theta}_i) = \frac{\sigma_i^2}{n_i} M^{-1}(\xi_i, \theta_i), \quad i = 1, 2,$$

where

$$M(\xi_i, \theta_i) = \int_{\mathcal{X}} g(x, \theta_i) g^T(x, \theta_i) d\xi_i(x) \quad \text{and} \quad g(x, \theta_i) = \frac{\partial}{\partial \theta_i} f(x, \theta_i).$$

We assume we have a large sample size with $\lim_{n_i \to \infty} \frac{n_{ij}}{n_i} = w_{ij}$ and focus on finding optimal approximate designs. These designs are probability measures on the dose interval and we denote a generic approximate design by $\xi_i = \{x_{i1}, \ldots, x_{il_i}; w_{i1}, \ldots, w_{il_i}\}, i = 1, 2$, where w_{ij} is the proportion of observations to be taken at x_{ij} subject to $\sum_{j=1}^{l_i} w_{ij} = 1$. Given an optimality criterion and a fixed sample size n_1 for group 1 and n_2 for group 2, we optimize the variables l_i, x_{ij}, w_{ij} and implement the optimal approximate design by taking roughly $n_i w_{ij}$ observations at $x_{ij}, j = 1, \ldots, l_i, i = 1, 2$. We do not directly optimize the number of replicates n_{ij} at each design point to find an optimal exact design because for large samples, the differences between the two types of optimal designs do not have much practical consequences. We revisit this issue in the summary.

2.2 General strategies for finding c-optimal designs

The bulk of optimal design work in the literature focuses on finding D-optimal designs. D-optimal designs minimize the generalized variance of the estimates of all the model parameters in the mean function and so they are frequently used for estimating all the model parameters. However, the end goal in a real study is rarely on estimating parameters. Frequently, the goal is in prediction or estimating a function of the model parameters.

A c-optimal design is used for estimating a function of the model parameters as accurately as possible. The function of interest can be nonlinear even if the model is linear. For example, if y is a continuous outcome and its mean is $E(y) = \theta_0 + \theta_1 x + \theta_2 x^2$, interest may be in estimating the turning point x_0 of the mean response and not just simply estimating the three parameters. In this case, $x_0 = -\theta_1/(2\theta_2)$ is a nonlinear function of the model parameters θ_0, θ_1 and θ_2 in a linear model. Thus our goal is to find an optimal allocation scheme that minimizes the (asymptotic) variance of the estimated x_0. The design criterion is a function of the model parameters and so a c-optimal design is appropriate.

This paper constructs optimal designs to estimate a function of the differences between the parameters θ_1 and θ_2. Let this function be $h(\theta_1, \theta_2)$ and let its estimator be $h(\hat{\theta}_1, \hat{\theta}_2)$. Its asymptotic variance is

$$
\begin{aligned}
var(h(\hat{\theta}_1, \hat{\theta}_2)) \approx & \left(\frac{1}{n_1} c_1^T(\theta_1, \theta_2) M^{-1}(\xi_1, \theta_1) c_1(\theta_1, \theta_2) \right. \\
& \left. + \frac{1}{n_2} c_2^T(\theta_1, \theta_2) M^{-1}(\xi_2, \theta_2) c_2(\theta_1, \theta_2) \right) \sigma^2,
\end{aligned}
$$

where $c_i(\theta_1, \theta_2) = \frac{\partial}{\partial \theta_i} h(\theta_1, \theta_2)$ is the gradient of the function h with respect to $\theta_i, i = 1, 2$.

An appropriate choice of an optimality criterion for a precise estimation of the function $h(\theta_1, \theta_2)$ is given by

$$\psi_c(\xi_1, \xi_2) = \sum_{i=1}^{2} \frac{1}{n_i} c_i^T(\theta_1, \theta_2) M^{-1}(\xi_i, \theta_i) c_i(\theta_1, \theta_2). \tag{2.2}$$

As usual, for n_1, n_2 fixed, a pair of designs $\xi^* = (\xi_1^*, \xi_2^*)$ is called c-optimal if it minimize the function $\psi_c(\xi_1, \xi_2)$ over the space of all approximate pairs of designs (ξ_1, ξ_2) on $\mathcal{X}^2$.

One observes that if the design ξ_i is locally c-optimal for the single model $f(x, \theta_i), i = 1, 2$, with respect to $c_i(\theta_1, \theta_2)$, then the pair $\xi = (\xi_1, \xi_2)$ is locally c-optimal design for model (2.1). The following equivalence theorem provides necessary and sufficient conditions for a design to be optimal under (2.2). The proof follows by standard arguments of optimal design theory and is therefore omitted.

Theorem 1. *For the model (2.1), let*

$$\phi_c(x_1, x_2, \xi_1, \xi_2) = \sum_{i=1}^{2} \frac{1}{n_i} (c_i^T(\theta_1, \theta_2) M^{-1}(\xi_i, \theta_i) g(x_i, \theta_i))^2. \tag{2.3}$$

A design $\xi^* = (\xi_1^*, \xi_2^*)$ is c-optimal if and only if the inequality

$$\phi_c(x_1, x_2, \xi^*) \leq \psi_c(\xi^*) \tag{2.4}$$

holds for all $x_1, x_2 \in \mathcal{X}$. Moreover, equality is achieved in (2.4) for any $(x_1^*, x_2^*) \in supp(\xi_1^*) \times\ supp(\xi_2^*)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0284

2.3 c-optimal design for models with shared parameters

Suppose there are common parameters in the two models in (2.1) and they are denoted by $\theta_c \in \mathbb{R}^q$ in the models

$$y_{ijk} = f(x_{ij}, \theta_c, \theta_{di}) + \varepsilon_{ijk}, \quad i = 1, 2; j = 1, \ldots, l_i; k = 1, \ldots, n_{ij}. \tag{2.5}$$

They are assumed to have the same value in the two groups and the rest of the parameters $\theta_{di} \in \mathbb{R}^{p-q}$ have different values in the two groups. Assume that the designs for the two models are ξ_1 and ξ_2 and $\theta = (\theta_c^T, \theta_{d1}^T, \theta_{d2}^T)^T$ is the vector of unknown parameters. Under the assumption $\lim_{n\to\infty} \frac{n_i}{n} = \lambda_i \in (0,1), i = 1, 2$, the maximum likelihood estimate $\hat{\theta} = (\hat{\theta}_c^T, \hat{\theta}_{d1}^T, \hat{\theta}_{d2}^T)^T$ is asymptotically normally distributed with covariance matrix $cov(\hat{\theta}) = \frac{\sigma^2}{n}\mathcal{M}^{-1}(\xi, \theta)$, where

$$\mathcal{M}(\xi, \theta) = \sum_{i=1}^{2} \lambda_i M^{(i)}(\xi_i, \theta) \tag{2.6}$$

is the information matrix of ξ and the matrices $M^{(i)}(\xi_i, \theta)$ are defined by

$$M^{(i)}(\xi_i, \theta) = \int_{\mathcal{X}} g_i(x, \theta) g_i^T(x, \theta) d\xi_i(x) \tag{2.7}$$

with

$$\begin{aligned} g_1^T(x, \theta) &= (\frac{\partial}{\partial \theta_c} f^T(x, \theta_c, \theta_{d1}), \frac{\partial}{\partial \theta_{d1}} f^T(x, \theta_c, \theta_{d1}), 0_{p-q}^T) && (2.8) \\ g_2^T(x, \theta) &= (\frac{\partial}{\partial \theta_c} f^T(x, \theta_c, \theta_{d2}), 0_{p-q}^T, \frac{\partial}{\partial \theta_{d2}} f^T(x, \theta_c, \theta_{d2})). && (2.9) \end{aligned}$$

Consequently, the asymptotic variance of the estimated function $h(\hat{\theta})$ is

$$var(h(\hat{\theta})) \approx \frac{\sigma^2}{n} c^T(\theta)\mathcal{M}^{-1}(\xi,\theta)c(\theta)$$

where $c(\theta) = \frac{\partial}{\partial\theta}h(\theta)$ denotes the gradient of the function h with respect to θ. It follows that the sought design is a c-optimal design ξ^* that minimizes

$$\psi_c(\xi) = c^T(\theta)\mathcal{M}^{-1}(\xi,\theta)c(\theta). \tag{2.10}$$

The following theorem gives a characterization of a c-optimal design for models with shared parameters.

Theorem 2. *For the model (2.5), let*

$$\phi_c(x_1,x_2,\xi) = \sum_{i=1}^{2} \lambda_i (g_i^T(x_i,\theta)\mathcal{M}^{-1}(\xi,\theta)c(\theta))^2 \tag{2.11}$$

A design $\xi^* = (\xi_1^*, \xi_2^*)$ is c-optimal if and only if the inequality

$$\phi_c(x_1,x_2,\xi^*) \leq \psi_c(\xi^*) \tag{2.12}$$

holds for all $x_1, x_2 \in \mathcal{X}$. In addition, the above inequality (2.12) becomes an equality for any $(x_1^*, x_2^*) \in supp(\xi_1^*) \times supp(\xi_2^*)$.

The function on the left hand side of the above inequality (or similarly, the one in (2.4) is frequently called the sensitivity function of the design ξ^* and if the design interval is one or two-dimensions, a visual inspection of the graph of the sensitivity function can confirm optimality of the design. To compare the worth of a design ξ with the c-optimal design ξ^*, we calculate its c-efficiency given by

$$\mathrm{Eff}_c(\xi) = \frac{\psi_c(\xi^*)}{\psi_c(\xi)}. \tag{2.13}$$

If the above ratio is $1/2$, this means that the design ξ needs to be replicated twice to do as well as the c-optimal design ξ^*. In general, designs with high c-efficiencies are preferred.

The above ratio requires that the optimal design be known. Otherwise, if the objective is still a convex function over the space of approximate designs, one can derive a lower bound for $\mathrm{Eff}_c(\xi)$ via an equivalence theorem. This efficiency lower bound provides the minimum efficiency of a design without knowledge of the optimum design. In particular, a direct application of Theorem 2 and [Pilz [9], p.137, Lemma 11.5] shows that for any approximate design ξ, we have

$$2c^T\mathcal{M}^{-1}(\xi,\theta)c - \sup_{x_1,x_2\in\mathcal{X}}\sum_{i=1}^{2}\lambda_i(g_i^T(x_i,\theta)\mathcal{M}^{-1}(\xi,\theta)c)^2 \le c^T\mathcal{M}^{-1}(\xi^*,\theta)c,$$

which leads to the following corollary.

Corollary 1.

$$\mathit{Eff}_c(\xi) \ge 2 - \frac{\sup_{x_1,x_2\in\mathcal{X}}\phi_c(x_1,x_2,\xi)}{\psi_c(\xi)}. \tag{2.14}$$

Such an efficiency lower bound is helpful in practice. For example, if an algorithm stops prematurely or is trapped at a local optimum, the efficiency of the design found at that point can be evaluated using such an efficiency lower bound. If the value of the bound is sufficiently high, the design may be deemed adequate for practical purposes and there may not be the need to find the optimal design.

3. Applications

In this section, we apply our results to find optimal designs for models with shared parameters in four scenarios. The first concerns parallel line assay models where we want to estimate the location-shift parameter μ in parallel models commonly used in a bioassay. An early work is Puri and Gupta (1986) with many follow-up work. Some recent ones are Fleetwood et al. (2015) and Faya et al. (2020), just to name a few. The second assumes two patient groups reacts to a drug intake via an EMAX model with 3 parameters, and one of them is shared and assumed to have the same value when no treatment is involved. The third concerns an exponential model where the interest is to find an optimal design to estimate the difference in the AUCs from two treated groups in a shared parameter model. For this application, analytical solution is not available and we use Particle Swarm Optimization, which is a general purpose optimization algorithm, to find a numerical solution. Throughout, we assumed the proportion of subjects assigned to the various groups is pre-determined. The fourth application shows that for a total fixed sample size, our method can also finds the optimal proportion of subjects to assign to the two groups and a c-optimal design for a model with shared parameters at the same time.

In all cases, our optimal designs are model-based, which is increasingly embraced among dose-finding oncology trialists (Love et al., 2017). Applications of model-based optimal designs can also be found in toxicology studies (Dette et al., 2011).

Application 1 (Parallel model)

The parallel model defined on a scaled compact interval $\mathcal{X} = [-1, 1]$ has a mean response given by

$$f(x, \theta_c, \theta_{di}) = \theta_{di} + \theta_c x, \quad i = 1, 2. \tag{3.1}$$

This model is commonly used in bioassay experiments to analyze responses from different doses of the standard and test preparations; see Huang et al. (2006), for example. The parallelism assumption is stringent and users usually test the validity of the assumption. One such recent test was proposed by Novick and Yang (2019), which is also applicable to test parallelism for more complicated models. The location-shift parameter depends on the model parameters $\theta = (\theta_{d1}, \theta_{d2}, \theta_c)^\top$ and is given by

$$\mu(\theta) = \frac{\theta_{d1} - \theta_{d2}}{\theta_c}.$$

This parameter is widely used to measure the location-shift between the standard and test preparations in parallel line assays. Let ξ_1 and ξ_2 be the two designs for the two models and let $\xi = (\xi_1, \xi_2)$. If $c = (1, -1, -\mu(\theta))^T$, the variance of $\mu(\hat{\theta})$ from design ξ is approximately

$$var(\mu(\hat{\theta})) \approx \frac{\sigma^2}{n\theta_c^2} c^T \mathcal{M}^{-1}(\xi) c.$$

As an example, we suppose the parameters obtained from initial information are $\theta_{d1} = 3, \theta_{d2} = 0.5$ and $\theta_c = 1$, implying that $\mu = 2.5$. We also assume that $\lambda_1 = 1/4$ and $\lambda_2 = 3/4$. Let ξ_1^* be the one point design which put mass on -1 and ξ_2^* be the design which put mass $1/6$ on -1 and $5/6$ on 1. We now verify that the pair ξ^* is c-optimal for model (3.1). For the design ξ^*, it is straightforward to show that $\psi_c(\xi^*) = 7$ and $\phi_c(x_1, x_2, \xi^*) = (x_1 - 1)^2 + 3x_2^2$. Obviously, the function $\phi_c(x_1, x_2, \xi^*)$ attains its maximum value of 7 at $(-1, -1)$ and $(-1, 1)$. It follows from Theorem 2 that the pair is c-optimal.

To compare the c-optimal design with the design $\xi_E = (\xi_{E1}, \xi_{E2})$, where both ξ_{E1} and ξ_{E2} are equally supported at -1 and 1, we calculate the criterion value of ξ_E which is $\psi_c(\xi_E) = 124/7$. Consequently, its c-efficiency

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0284

is $\text{Eff}_c(\xi_E) = 49/124$.

Application 2 (EMAX model)

The Michaelis-Menten model has two parameters and takes the form

$$f(x, \theta) = \frac{\theta_{d_0} x}{x + \theta_{d_1}}, \tag{3.2}$$

with $\theta^T = (\theta_{d_0}, \theta_{d_1})$. It is widely used to study enzyme-substrate dose-relationship in kinetic biological systems but it is also consistently used in various ways across disciplines, see for example, in biology (Butler and Wolkowicz, 1985) and in agriculture (Yu and Gu, 2007), to name a few. Optimal experimental designs to estimate any one or two of the parameters have been investigated, including the case when there is unequal interest in estimating each of the parameters (Lopez-Fidalgo and Wong, 2002).

The EMAX model generalizes the Michaelis-Menten model by having an extra parameter to model outcomes more flexibly. Dette et al. (2005) found optimal designs that addressed various lack of fit issues for the model

$$f(x, \theta_c, \theta_{di}) = \theta_c + \frac{\theta_{di1} x}{x + \theta_{di2}}, x \in [0, b], \quad i = 1, 2. \tag{3.3}$$

Here b is pre-specified, x is the dose, θ_c is the placebo effect at dose $x = 0$, θ_{di1} the asymptotic maximum treatment benefit over placebo for the ith group and θ_{di2} the dose that gives half of the asymptotic maximum effect for the ith group. The common parameter θ_c means that the placebo effect is the same between two groups.

The $ED_p, 0 < p < 1$, is the smallest dose that achieves $100p\%$ of the maximum effect in the given dose range. For model (3.3), the ED_p of the ith group in terms of the underlying model parameter is

$$ED_i p = \frac{bp\theta_{di2}}{\theta_{di2} + b(1-p)}.$$

Our interest is to find a design that best estimates the difference of ED_p's between the two groups. If ξ_i is the design used for the i^{th} group, the implemented design for model (3.3) is $\xi = (\xi_1, \xi_2)$, and the approximate variance of $\hat{ED}_{1p} - \hat{ED}_{2p}$ is

$$var(\hat{ED}_{1p} - \hat{ED}_{2p}) \approx \frac{\sigma^2}{n} c^T \mathcal{M}^{-1}(\xi) c.$$

Here $c = \frac{\partial}{\partial \theta}(ED_{1p} - ED_{2p}) = (0, 0, \gamma_1, 0, -\gamma_2)$ and

$$\gamma_i = \frac{b^2(1-p)p}{(\theta_{di2} + b(1-p))^2}, i = 1, 2.$$

Using the argument in Dette et al. (2010), it is sufficient to restrict the search on designs of the form $\xi = (\xi_1, \xi_2)$ with $\xi_1 = (x_1, b; w_1, 1-w_1)$ and $\xi_2 = (0, x_2, b; 1 - w_2 - w_3, w_2, w_3)$. Hence, the only values to determine in this case are the interior support points x_1, x_2 and the weights w_1, w_2, w_3. As an example of a locally c-optimal design, suppose that $b = 150, \lambda_1 =$

$1/5, \lambda_2 = 4/5$ $\theta_{d11} = 1/2, \theta_{d12} = 20$, $\theta_{d21} = 1/4$ and $\theta_{d22} = 25$, where we recall that λ_1 and λ_2 are the user-selected relative size of the two groups when the total sample size is fixed. Table 1 presents c-optimal designs for selected values of p from Theorem 2. For example, when $p = 0.5$, the locally c-optimal design for estimating the difference of the ED_{50}'s in the two groups is $\xi^* = (\xi_1^*, \xi_2^*)$, where $\xi_1^* = (7.1638, 150; 0.7699, 0.2301)$ and $\xi_2^* = (0, 18.7500, 150; 0.0910, 0.6060, 0.3030)$.

It is instructive to compare consequences of using a model with shared parameters or not. When there is no shared parameter, there is an extra parameter to estimate for the placebo effect in each group. For the i^{th} group, the EMAX model is

$$f(x, \theta) = \theta_{di0} + \frac{\theta_{di1} x}{x + \theta_{di2}}, x \in [0, b], \quad i = 1, 2,$$

and a direct application of Theorem 5.2 in Dette et al. (2010) for the EMAX model shows the locally ED_{50}-optimal design for the $i = 1$ group is $\xi_{S1} = (0, 15.7895, 150; 0.25, 0.5, 0.25)$, and that for the $i = 2$ group is $\xi_{S2} = (0, 18.7500, 150; 0.25, 0.5, 0.25)$. The implemented design is $\xi_{NS} = (\xi_{S1}, \xi_{S2})$ and for various values of p, its efficiency relative to the locally c-optimal design when the shared model is adopted is shown in the last column of Table 1. For example, if $p = 0.6$, the loss in efficiency of the design ξ_{NS} relative to the c-optimal design for the model with the shared parameter ξ^*

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0284

Table 1: The interior support points and weights of the c-optimal design for model (3.3) with shared parameter defined on [0, 150], and the efficiency of the optimal design ξ_{NS} for the model without the shared parameter.

p	x_1	w_1	x_2	w_2	w_3	$Eff_c(\xi_{NS})$
0.1	6.9496	0.7738	18.7500	0.6039	0.3019	0.8268
0.2	6.9866	0.7731	18.7500	0.6042	0.3021	0.8254
0.3	7.0321	0.7723	18.7500	0.6047	0.3023	0.8241
0.4	7.0895	0.7712	18.7500	0.6052	0.3026	0.8223
0.5	7.1638	0.7699	18.7500	0.6060	0.3030	0.8200
0.6	7.2637	0.7681	18.7500	0.6071	0.3035	0.8168
0.7	7.4053	0.7656	18.7500	0.6088	0.3044	0.8121
0.8	7.6208	0.7618	18.7500	0.6117	0.3058	0.8047
0.9	7.9870	0.7556	18.7500	0.6177	0.3088	0.7911

is about 19%.

Other c-optimal designs for model (3.3) for similar problems can be found similarly. For instance, if there is interest for predicting the mean difference in responses from the two groups at a given dose $x_0 > b$, i.e. we first set $c(\theta)$ equal to the derivative of $h(\theta) = \frac{\theta_{d11}x_0}{x_0+\theta_{d12}} - \frac{\theta_{d21}x_0}{x_0+\theta_{d22}}$ and find a design that minimizes its asymptotic variance. The vector of interest to use in the c-optimality criterion is the derivative of $h(\theta$ with respect to θ, i.e.

$$c = \frac{\partial}{\partial\theta}h(\theta) = \left(\frac{x_0}{x_0+\theta_{d12}}, -\frac{\theta_{d11}x_0}{(x_0+\theta_{d12})^2}, -\frac{x_0}{x_0+\theta_{d22}}, \frac{\theta_{d21}x_0}{(x_0+\theta_{d22})^2}\right)^T.$$

For example, a direct calculation shows that if $x_0 = 160$, the locally c-optimal design for the model with shared parameter is $\xi^* = (\xi_1^*, \xi_2^*)$, where $\xi_1^* = (11.0301, 150; 0.0295, 0.9705)$, $\quad \xi_2^* = (15.7831, 150; 0.0346, 0.9654)$. We note that 0 is not a support point of ξ_2^* either and most of its mass is at the right end of the design space at $b = 150$.

Application 3 (Exponential model)

Consider the Exponential model with mean function given by

$$f(x,\theta) = \theta_0 + \theta_1 \exp(x/\theta_2), \tag{3.4}$$

where θ_0 is the placebo effect, θ_1 is the slope of the curve, θ_2 determines the rate of effect increase and x is the dose level. Pinheiro et al. (2006) used this model in dose finding clinical studies and Dette et al. (2010) found optimal designs for the model. In a two-group trial, we assume the placebo effect is the same for each group and the models for the two groups become

$$f(x,\theta_c,\theta_{di}) = \theta_c + \theta_{di1}\exp(x/\theta_{di2}), x \in [0,b], \quad i = 1,2, \tag{3.5}$$

Our interest is to find a design that best estimates the difference of AUC between the two groups. Tthe area under the curve for the i^{th} group is AUC_i and is given by

$$AUC_i = \theta_c b + \theta_{di1}\theta_{di2}(\exp(b/\theta_{di2}) - 1), \quad i = 1,2.$$

For the design $\xi = (\xi_1, \xi_2)$, the approximate variance of $A\hat{U}C_1 - A\hat{U}C_2$ is

$$var(A\hat{U}C_1 - A\hat{U}C_2) \approx \frac{\sigma^2}{n} c^T \mathcal{M}^{-1}(\xi) c,$$

where $c = \frac{\partial}{\partial\theta}(AUC_1 - AUC_2) = (0, \nu_{11}, \nu_{12}, \nu_{21}, \nu_{22})$ with

$$\nu_{i1} = \theta_{di2}(\exp(b/\theta_{di2}) - 1), \quad \nu_{i2} = \theta_{di1}(\exp(b/\theta_{di2})(1 - 1/\theta_{di2}) - 1), i = 1,2.$$

To find a locally optimal design, we use thee nominal values for the setup: $b = 150, \lambda_1 = 1/5, \lambda_2 = 4/5$ and $\theta_{d11} = 1/2, \theta_{d12} = 20, \theta_{d21} = 1/4, \theta_{d22} = 25$. A direct computation similar to Applications 1 and 2 using Theorem 2 produces the sought optimal design but this will not be possible for more complicated models. For the latter, numerical methods are required. We recommend Particle Swarm solver available in Matlab, which implements the particle swarm optimizer (PSO) to search for a c-optimal design. PSO is a general-purpose optimization tool inspired by animals or nature and hence is called a nature-inspired metaheuristic algorithm. There are usually fast, assumptions free and easy to use. It is also now increasingly used to find all types of efficient designs in the optimal design literature; Chen, et al. (2022) gives a short overview of its use to find different types of optimal designs in various scenarios. In this example, a direct application of PSO with its default settings produces a design $\xi^* = (\xi_1^*, \xi_2^*)$, where $\xi_1^* = (130.0826; 1)$ and $\xi_2^* = (0, 125.4117, 150; 0.16130.83610.0026)$, which can be verified to be c-optimal using an equivalence theorem.

Application 4 (Determining optimal proportion of observations for two groups)

We have assumed that the sample sizes n_1 and n_2 in the two groups are fixed in advance. Sometimes, there is flexibility and one may find an optimal design that optimizes the triplet $\xi = (\xi_1, \xi_2, \lambda)$, where $\lambda = n_1/(n_1 + n_2)$. Accordingly, we find a design to minimize

$$\psi_c(\xi) = c^T(\theta)\mathcal{M}^{-1}(\xi, \lambda, \theta)c(\theta), \tag{3.6}$$

where

$$\mathcal{M}(\xi, \lambda, \theta) = \lambda M^{(1)}(\xi_1, \theta) + (1 - \lambda)M^{(2)}(\xi_2, \theta). \tag{3.7}$$

Similar to Theorem 2, we have the following equivalence theorem.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0284

Theorem 3. *For the model (2.5), let*

$$\phi_c(x_1, x_2, t, \xi, \lambda) = \sum_{i=1}^{2} I\{t = i\}(g_i^T(x_i, \theta)\mathcal{M}^{-1}(\xi, \lambda, \theta)c(\theta))^2. \tag{3.8}$$

A design $\xi^* = (\xi_1^*, \xi_2^*, \lambda^*)$ is c-optimal if and only if the inequality

$$\phi_c(x_1, x_2, t, \xi^*, \lambda^*) \leq \psi_c(\xi^*) \tag{3.9}$$

holds for all $x_1, x_2 \in \mathcal{X}$ and $t \in \{1, 2\}$. Moreover, equality is achieved in (3.9) for any $(x_1^*, x_2^*, t) \in supp(\xi_1^*)\times$ $supp(\xi_2^*) \times \{1, 2\}$.

Proof. Let $\lambda \in [0, 1]$, let $\xi = (\xi_1, \xi_2) \in \Xi \times \Xi$ and let δ_x be the Dirac measure at the point $x \in \mathcal{X}$. After noting that the set $\mathfrak{M}_\lambda^{(2)} = \{\mathcal{M}(\xi, \lambda, \theta) : (\xi_1, \xi_2) \in \Xi \times \Xi, \lambda \in [0, 1]\}$ is the convex hull of the set $\mathfrak{D}_\lambda^{(2)} = \{\sum_{i=1}^2 I\{t = i\}\mathcal{M}^{(i)}(\delta_{x_i}, \theta) : x_1, x_2 \in \mathcal{X}, t = 1, 2\}$, the rest of the proof parallels to the proof of Theorem 2 and for space consideration, it is omitted.

We revisit Application 1 and find a locally c-optimal design for estimating the location-shift parameter μ. We consider the case when $|\mu| > 2$ or not, separately. In the former case, let $\xi_1^* = (-1, 1; w_1^*, 1 - w_1^*)$ and let $\xi_2^* = (-1, 1; w_2^*, 1 - w_2^*)$ with

$$w_1^* = \frac{\mu\lambda^* + 1}{2\mu\lambda^*} \quad \text{and} \quad w_2^* = \frac{\mu(1 - \lambda^*) - 1}{2\mu(1 - \lambda^*)}.$$

If λ^* satisfies the condition that $1/\mu < \lambda^* < 1 - 1/\mu$, a straightforward

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0284

calculation shows the information matrix of the design $\xi^* = (\xi_1^*, \xi_2^*, \lambda^*)$ is

$$\mathcal{M}(\xi^*, \lambda^*) = \begin{pmatrix} \lambda^* & 0 & -1/\mu \\ 0 & 1-\lambda^* & 1/\mu \\ -1/\mu & 1/\mu & 1 \end{pmatrix}.$$

It follows that

$$\phi_c(x_1, x_2, t, \xi^*, \lambda^*) = (I\{t=1\}x_1^2 + I\{t=2\}x_2^2)\mu^2$$

and $\psi_c(\xi^*) = \mu^2$. Obviously, the function $\phi_c(x_1, x_2, t, \xi^*, \lambda^*)$ attains its maximum μ^2 at $(\pm 1, \pm 1, 1)$ and $(\pm 1, \pm 1, 2)$ and the design ξ^* satisfies Theorem 3, so it is c-optimal. For example, the design ξ^* with $w_1^* = 9/10, w_2^* = 1/10$ and $\lambda^* = 1/2$ is a c-optimal design for $\mu = 2.5$. The corresponding value of the criterion is $\psi_c(\xi^*) = 6.25 < 7$, which is the value of the c-optimal design given in Application 1.

When $|\mu| \leq 2$, it can be similarly shown that the design with $w_1^* = \mu/2, w_2^* = 0$ and $\lambda^* = 1/2$ is c-optimal because $\phi_c(x_1, x_2, t, \xi^*, \lambda^*) = \psi_c(\xi^*) = 4$. Further, one can verify directly that the c-efficiency of $\xi_E = (\xi_{E1}, \xi_{E2}, 1/2)$ is $\text{Eff}_c(\xi_E) = \frac{\mu^2}{\mu^2+4}$ if $|\mu| > 2$ and $\text{Eff}_c(\xi_E) = \frac{4}{\mu^2+4}$ if $|\mu| \leq 2$.

4. Multiple-objective c-optimal designs and L-optimal designs

The previous sections focus on constructing optimal designs for estimating a single function of the model parameters. Sometimes, there are several

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0284

interesting functions to estimate, and some are more interesting than others. For instance, in addition to estimating ED_{50}, other percentiles may also be of interest. The goal then is to find a design that provides higher efficiencies for the more important objectives. Accordingly, we first rank each objective by its importance and want a design that delivers a user-specified efficiency of e_i under the i^{th} objective. Assuming objective 1 is the most important, we have $e_1 \geq e_2 \geq \ldots \geq e_m$. This formulation is intuitive but finding such a constrained optimal design can be problematic.

An alternative is to find a design that best estimates a convex combination of the m convex functions of interest. Since the resulting criterion is still convex, we can treat the optimization problem as a single-objective c-optimal design problem. After expressing each of the convex functionals as a vector $c_i(\theta), i = 1, \ldots, m$, we find an approximate design that minimizes

$$\psi_C(\xi) = \sum_{i=1}^{m} \alpha_i c_i^T(\theta)\mathcal{M}^{-1}(\xi, \lambda, \theta)c_i(\theta) \tag{4.1}$$

among all approximate designs on $\mathcal{X}$. Here each $\alpha_i \in [0, 1]$ is a preselected weight in the compound criterion and they are normalized so that $\sum_{i=1}^{m} \alpha_i = 1$. The interpretation is that a larger weight implies that there is greater interest in estimating the particular function of model parameters.

The following theorem is a direct generalization of the previous theorem and can be used to verify whether a design is a compound optimal

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0284

design.The proof is similar to the proof of Theorem 2 and is omitted.

Theorem 4. *For the model (2.5), let*

$$\phi_C(x_1, x_2, \xi) = \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{2} \lambda_j (g_j^T(x_j, \theta)\mathcal{M}^{-1}(\xi, \theta)c_i(\theta))^2 \tag{4.2}$$

A design $\xi^* = (\xi_1^*, \xi_2^*)$ *is compound c-optimal if and only if the inequality*

$$\phi_C(x_1, x_2, \xi^*) \leq \psi_C(\xi^*) \tag{4.3}$$

holds for all $x_1, x_2 \in \mathcal{X}$. *Moreover, equality is achieved in (4.3) for any* $(x_1^*, x_2^*) \in supp(\xi_1^*)\times\ supp(\xi_2^*)$.

Cook and Wong (1994) proposed a graphical method for finding a dual-objective optimal design with $m = 2$. First the two single-objective optimal designs are determined and the two types of efficiencies of each compound optimal design are computed. These efficiencies are then plotted against values of $\lambda \in [0, 1]$. To find the sought constrained optimal design indirectly from the easier-to-find compound optimal designs, one then draw a horizontal line at the sought efficiency level for the primary objective and note the corresponding λ where the horizontal line meets the efficiency plot for the primary objective. The compound optimal design corresponding to that λ is the sought constrained optimal design. The method applies to any two convex functionals but when m is large, the plots are high-dimensional

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0284

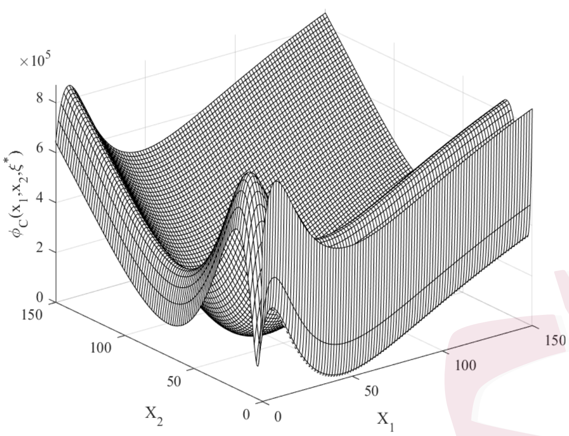

Figure 1: Plot of the sensitivity function $\phi_C(x_1, x_2, \xi^*)$ of the design x_i^* for Application 2 (Continued) confirms its compound c-optimality.

and it becomes hard to appreciate visually the features in the plot and infer whether the conditions of the equivalence theorem are met.

Application 2 (Continued).

Suppose we want to find an optimal design to ascertain differences of E_{25}, E_{50} and E_{75} with equal interest. This means that the weights in (4.1) are $\alpha_i = 1/3, i = 1, 2, 3$. A direct calculation shows that the compound c-optimal design is $\xi^* = (\xi_1^*, \xi_2^*)$, with $\xi_1^* = (7.4418, 150; 0.7649, 0.2351)$ and $\xi_2^* = (0, 18.7500, 150; 0.0872, 0.6085, 0.3043)$. Figure 1 plots the sensitivity function of this design and confirms the c-optimality ξ^*.

Some studies aim to estimate several functions of the model parameters

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0284

Table 2: The interior support points and weights of the compound c-optimal design for model (3.3).

α_1	α_2	α_3	x_1	w_1	x_2	w_2	w_3
0.1	0.8	0.1	7.2954	0.7675	18.7500	0.6067	0.3034
0.2	0.6	0.2	7.3756	0.7661	18.7500	0.6075	0.3038
0.25	0.5	0.25	7.4045	0.7656	18.7500	0.6079	0.3040
1/3	1/3	1/3	7.4418	0.7649	18.7500	0.6085	0.3043
0.4	0.2	0.4	7.4646	0.7645	18.7500	0.6090	0.3045

simultaneously. In this case, L-optimal designs are more appropriate. They generalize c-optimal designs and an approximate design ξ^* is defined to be L-optimal if it minimizes

$$\psi_L(\xi) = \mathrm{tr}(L\mathcal{M}^{-1}(\xi, \theta)). \tag{4.4}$$

Here L is a user-selected constant matrix and it can be shown that the criterion is a convex function of the information matrix. The following equivalence theorem characterizes a L-optimal design.

Theorem 5. *For the model (2.5), let*

$$\phi_L(x_1, x_2, \xi) = \sum_{i=1}^{2} \lambda_i (g_i^T(x_i, \theta)\mathcal{M}^{-1}(\xi, \theta) L \mathcal{M}^{-1}(\xi, \theta) g_i(x_i, \theta)) \tag{4.5}$$

A design $\xi^* = (\xi_1^*, \xi_2^*)$ is L-optimal if and only if the inequality

$$\phi_L(x_1, x_2, \xi^*) \leq \psi_L(\xi^*) \tag{4.6}$$

holds for all $x_1, x_2 \in \mathcal{X}$. In addition, the above inequality (4.6) becomes an equality for any $(x_1^*, x_2^*) \in$ supp$(\xi_1^*)\times$ supp(ξ_2^*).

Proof. Let NND(k) be the set of all $k \times k$ non-negative definite matrices, let Ξ be the set of approximate designs on $\mathcal{X}$ and let $\xi = (\xi_1, \xi_2) \in \Xi \times \Xi$. Then

$$\mathcal{M}(\xi, \theta) = \sum_{i=1}^{2} \lambda_i M^{(i)}(\xi_i, \theta).$$

The set $\mathfrak{M}^{(2)} = \{\mathcal{M}(\xi, \theta) : (\xi_1, \xi_2) \in \Xi \times \Xi\}$ is a convex subset of the set NND$(2p - q)$. Let δ_x be the Dirac measure at the point $x \in \mathcal{X}$. It follows that $\mathfrak{M}^{(2)}$ is the convex hull of the set $\mathfrak{D}^{(2)} = \{\mathcal{M}(\delta_{x_1}, \delta_{x_2}, \theta) : x_1, x_2 \in \mathcal{X}\}$, and the function $\psi_L(\xi)$ defined in (4.4) is convex on the set $\Xi_1 \times \Xi_2$. Since $\psi_L(\xi)$ depends on ξ only through $\mathcal{M}(\xi, \theta)$, it can be viewed as a function on $\mathfrak{M}^{(2)}$ and denoted by $\psi_L(\mathcal{M})$.

Because of the convexity of ψ_L the design $\xi^* = (\xi^*_{(1)}, \xi^*_{(2)})$ is c-optimal if and only if its Fréchet derivative satisfies

$$\partial\psi_L(\mathcal{M}^*, E_0^*) = \lim_{\alpha\to 0^+} \frac{\psi_L(\mathcal{M}^* + \alpha E_0) - \psi_L(\mathcal{M}^*)}{\alpha} \geq 0,$$

for all directions $E_0^* = E - \mathcal{M}^*$. Since $\mathfrak{M}^{(2)}$ =conv($\mathfrak{D}^{(2)}$), it is sufficient to verify the above inequality for all $E \in \mathfrak{D}^{(2)}$. Let $\alpha \in (0,1)$ and let $\xi_\alpha = (\xi_{\alpha 1}, \xi_{\alpha 2}) = (1-\alpha)(\xi_1, \xi_2) + \alpha(\delta_{x_1}, \delta_{x_2})$. Noting that $\mathcal{M}_\alpha := \mathcal{M}(\xi_\alpha) = (1-\alpha)\mathcal{M} + \alpha\mathcal{M}(\delta_{x_1}, \delta_{x_2})$, we have

$$\begin{aligned} \frac{d\psi_L(\mathcal{M}_\alpha)}{d\alpha} &= \frac{d}{d\alpha}\mathrm{tr}(L\mathcal{M}_\alpha^{-1}) \\ &= \mathrm{tr}\left(\frac{d}{d\alpha}L\mathcal{M}_\alpha^{-1}\right) \\ &= \mathrm{tr}\left(L(-\mathcal{M}_\alpha^{-1}(\mathcal{M}(\delta_{x_1}, \delta_{x_2}) - \mathcal{M})\mathcal{M}_\alpha^{-1})\right). \end{aligned}$$

It follows that the directional derivative of $\psi_L(\mathcal{M}_\alpha)$ at $\mathcal{M}$ in the direction of $E_0 = \mathcal{M}(\delta_{x_1}, \delta_{x_2}) - \mathcal{M}$ is

$$\begin{aligned} \partial\psi_L(\mathcal{M}, E_0) &= \lim_{\alpha\to^+} \frac{\psi_L(\mathcal{M} + \alpha E_0) - \psi_L(\mathcal{M})}{\alpha} \\ &= \frac{d}{d\alpha}\psi_L(\mathcal{M}_\alpha)\bigg|_{\alpha=0^+} \\ &= \mathrm{tr}\left(L(-\mathcal{M}^{-1}(\mathcal{M}(\delta_{x_1}, \delta_{x_2}) - \mathcal{M})\mathcal{M}^{-1})\right) \\ &= \mathrm{tr}(L\mathcal{M}^{-1}) - \mathrm{tr}\left(L\mathcal{M}^{-1}\mathcal{M}(\delta_{x_1}, \delta_{x_2})\mathcal{M}^{-1}\right) \\ &= \mathrm{tr}(L\mathcal{M}^{-1}) - \sum_{i=1}^{2} \lambda_i(g_i^T(x_i,\theta)\mathcal{M}^{-1}L\mathcal{M}^{-1}g_i(x_i,\theta)). \end{aligned}$$

Consequently, the design $\xi^* = (\xi_1^*, \xi_2^*)$ is L-optimal if and only if

$$\sum_{i=1}^{2} \lambda_i (g_i^T(x_i, \theta)\mathcal{M}^{-1} L \mathcal{M}^{-1} g_i(x_i, \theta)) \leq \mathrm{tr}(L\mathcal{M}^{-1}) \tag{4.7}$$

for all $x_1, x_2 \in \mathcal{X}$ and the proof Theorem 5 is complete. $\square$

Application 2 (continued)

Suppose we want to find a L-optimal design to ascertain differences of the two responses at $x_{01} = 155$ and $x_{02} = 160$. We let A be a matrix with rows

$$A_i = \left(0, \frac{x_{0i}}{x_0 i + \theta_{d12}}, -\frac{\theta_{d11} x_{0i}}{(x_{0i} + \theta_{d12})^2}, -\frac{x_{0i}}{x_{0i} + \theta_{d22}}, \frac{\theta_{d21} x_{0i}}{(x_{0i} + \theta_{d22})^2}\right),$$

where $x_{01} = 155, x_{02} = 160$, and let $L = AA^T$. A direct calculation shows the L-optimal design for model (3.3) is given by $\xi^* = (\xi_1^*, \xi_2^*)$, where $\xi_1^* = (11.1625, 150; 0.0237, 0.9763)$ and $\xi_2^* = (15.9591, 150; 0.0278, 0.9722)$.

5. Extension to multi-response models

Multi-response models are common in real applications. For example, the Berman model is frequently used to analyze data for calibrating apparatus in microwave engineering (Berman, 1983). There are also applications of the model for modeling concentric circles and ellipses in biometrics and medical diagnostics. The shapes of the eyeballs are increasingly used in biometrics to securely identify individuals and this involves fitting data to images of concentric circles. In medical imaging, a person infected with

Malaria is identified by the concentric elliptical shapes of the parasite cells in the body; see Al-Sharadqah and Nguyen (2022) for details.

Let $Y_i = (y_{i1}(x), \ldots, y_{is}(x))^T$ be our s-dimensional response vector, let $F(x) = (f_1(x), \ldots, f_s(x))^T$ be a $s \times p$ matrix of regression functions, let θ_i, be a vector of p unknown parameters of the model for the i^{th} group and let ϵ be a s-dimensional vector of random errors, each with mean zero and non-singular covariance matrix $\Sigma = (\sigma_{ij})_{s\times s}$. Our multi-response linear models for the two groups are

$$Y_{ijk} = F(x_{ij})\theta_i + \epsilon_{ijk}, \quad i = 1, 2; j = 1, \ldots, l_i; k = 1, \ldots, n_{ij}, \tag{5.1}$$

The two groups may represent two machines providing readings of the shapes of eyeballs or cell shapes. Suppose further that θ_c is the vector of shared parameters for the two groups and the linear multi-response model (5.1) becomes

$$Y_{ijk} = F_1(x_{ij})\theta_c + F_2(x_{ij})\theta_{di} + \epsilon_{ijk}. \tag{5.2}$$

Assume that n_i points are observed around the ith concentric circle with radius r_i centered at the common center (θ_1, θ_2). Denote the two coordinates of measured data point on the ith circle by y_{i1} and $y_{i2}, i = 1, 2$.

The Berman model is usually expressed as

$$\begin{cases} y_{i1}(t) = \theta_1 + \theta_{i3}\cos t - \theta_{i4}\sin t + \epsilon_{i1}, \\ y_{i2}(t) = \theta_2 + \theta_{i3}\sin t + \theta_{i4}\cos t + \epsilon_{i2}. \end{cases} \quad t \in [0, 2\pi], \tag{5.3}$$

Letting I_2 be the identity matrix of dimension 2, we have from (5.2),

$$F_1(t) = I_2, F_2(t) := A(t) = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix},$$

$\theta_c = (\theta_1, \theta_2)^T, \theta_{di} = (\theta_{i3}, \theta_{i4})^T$ and $r_i = \sqrt{\theta_{i3}^2 + \theta_{i4}^2}$, for $i = 1, 2$. The covariance matrix of $\epsilon_i = (\epsilon_{i1}, \epsilon_{i2})^T$ is assumed to be $\Sigma = \sigma^2 I_2$. Under model (5.2) the information matrix of a design pair $\xi = (\xi_1, \xi_2)$ is

$$\mathcal{M}(\xi) = \sum_{i=1}^{2} \lambda_i M^{(i)}(\xi_i), \tag{5.4}$$

where the matrices $M^{(i)}(\xi_i), i = 1, 2$ are defined by

$$M^{(i)}(\xi_i) = \int_{\mathcal{X}} G_i^T(x)\Sigma^{-1}G_i(x)d\xi_i(x) \tag{5.5}$$

with $G_1(x) = (F_1(x), F_2(x), 0_{s\times(p-q)})$ and $G_2(x) = (F_1(x), 0_{s\times(p-q)}, F_2(x))$. Similar to the single response case, we have the following theorem that provides a characterization of a c-optimal design:

Theorem 6. *For the model (2.5), let*

$$\phi_c(x_1, x_2, \xi) = \sum_{i=1}^{2} \lambda_i c^T(\theta)\mathcal{M}^{-1}(\xi)G_i^T(x_i)\Sigma^{-1}G_i(x_i)\mathcal{M}^{-1}(\xi)c(\theta). \tag{5.6}$$

A design $\xi^* = (\xi_1^*, \xi_2^*)$ is c-optimal if and only if the inequality

$$\phi_c(x_1, x_2, \xi^*) \leq \psi_c(\xi^*) \tag{5.7}$$

holds for all $x_1, x_2 \in \mathcal{X}$, with equality at any $(x_1^*, x_2^*) \in supp(\xi_1^*) \times supp(\xi_2^*)$.

Application 5 Let η_{E_i} be the equidistant sampling design with sample size m_i on $[0, 2\pi]$ and let its m_i support points be at $t_j = (j-1)2\pi/m_i$ for $j \in \{1, \ldots, m_i\}, i = 1, 2$. The design problem is to find a c-optimal design for estimating the difference of the radii $r_1 - r_2$ under model (5.3), which is $\sqrt{\theta_{13}^2 + \theta_{14}^2} - \sqrt{\theta_{23}^2 + \theta_{24}^2}$. Here $c = \frac{\partial}{\partial\theta}(r_1 - r_2)$ and a direct calculation shows $c = (0, 0, \theta_{13}/r_1, \theta_{14}/r_1, \theta_{23}/r_2, \theta_{24}/r_2)^T$.

The information matrix of a design pair $\xi = (\xi_1, \xi_2)$ for model (5.3) is

$$M^{(1)}(\xi_1) = \begin{pmatrix} \Sigma^{-1} & \Sigma^{-1}A(\xi_1) & 0 \\ A^\top(\xi_1)\Sigma^{-1} & \frac{\sigma^2}{\det(\Sigma)}I_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad M^{(2)}(\xi_2) = \begin{pmatrix} \Sigma^{-1} & 0 & \Sigma^{-1}A(\xi_2) \\ 0 & 0 & 0 \\ A^\top(\xi_2)\Sigma^{-1} & 0 & \frac{\sigma^2}{\det(\Sigma)}I_2 \end{pmatrix},$$

where

$$A(\xi_i) = \int_{\mathcal{X}} A(t)d\xi_i = \begin{pmatrix} c(\xi_i) & -s(\xi_i) \\ s(\xi_i) & c(\xi_i) \end{pmatrix}, \quad c(\xi_i) = \int_{\mathcal{X}} \cos(t)d\xi_i, \quad \text{and} \quad s(\xi_i) = \int_{\mathcal{X}} \sin(t)d\xi_i.$$

Since $c(\eta_{E_i}) = \int_{\mathcal{X}} \cos(t)d\eta_{E_i} = 0, \quad s(\eta_{E_i}) = \int_{\mathcal{X}} \sin(t)d\eta_{E_i} = 0,$ we have

$A(\eta_{E_i}) = 0_{2\times 2}$. Thus, if $\eta_E = \lambda_1\eta_{E_1} + \lambda_2\eta_{E_2}$, then

$$\mathcal{M}(\eta_E) = \lambda_1 M^{(1)}(\eta_{E_1}) + \lambda_2 M^{(2)}(\eta_{E_2}) = \begin{pmatrix} \Sigma^{-1} & 0 & 0 \\ 0 & \frac{\sigma^2}{\det(\Sigma)} I_2 & 0 \\ 0 & 0 & \frac{\sigma^2}{\det(\Sigma)} I_2 \end{pmatrix}.$$

This implies that

$$\begin{aligned} \phi_c(t_1, t_2, \eta_E) &= \sum_{i=1}^{2} \frac{\sigma^2}{\lambda_i r_i^2} (\theta_{i3}, \theta_{i4}) A^T(t) A(t) (\theta_{i3}, \theta_{i4})^T \qquad (5.8) \\ &= \sum_{i=1}^{2} \frac{\sigma^2}{\lambda_i r_i^2} ((\theta_{i3}\cos t - \theta_{i4}\sin t)^2 + (\theta_{i3}\sin t + \theta_{i4}\cos t)^2) \\ &= \sum_{i=1}^{2} \frac{\sigma^2}{\lambda_i} \end{aligned}$$

and

$$\psi_c(\eta_E) = c^T(\theta)\mathcal{M}^{-1}(\eta_E)c(\theta) = \sum_{i=1}^{2} \frac{\sigma^2}{\lambda_i}. \qquad (5.9)$$

By Theorem 6, η_E is c-optimal for estimating $r_1 - r_2$ under model (5.3).

6. Summary

This paper constructs c and L-optimal designs for estimating one or more functions of the parameters in regression models with shared parameters. Some work on finding locally D-optimal designs to estimate all parameters in the model are available for such a setting, but not for estimating one or more functions of the parameters in the models with shared parameters.

The optimal designs also allow different levels of interests among the objectives, including how to choose the optimal proportions of patients to assign to the two groups given a fixed sample size. The methodology also extends to finding locally c-optimal designs for multi-response models with shared parameters.

Our focus was in finding locally optimal designs. It is possible to extend the method to find Bayesian optimal designs or minimax types of optimal designs, where both strategies do not require a single set of nominal values. For more complicated models or criteria, the optimal designs will have to be found numerically, such as using a nature-inspired metaheuristic algorithm like PSO used in Application 3 to find a c-optimal design. Masoudi et al. (2019), Chen et al. (2020) and Liu et al. (2021) showed how PSO can be used to find different types of optimal designs for more complicated models and optimality criteria.

In practice, we implement exact designs, which requires the sample size n to be known in advance. This implies that if the optimal proportion of observations at the design point x_i of an approximate design is w_i, we would have to round $n * w_i$ to a positive integer $[n * w_i]$ such that $\sum_i^n [n * w_i] = n$. This is typically done using some intuitive rounding-off procedures or rules in Pukelsheim and Rieder (1992). However, our take is that optimal exact

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0284

designs are useful only when we have a small sample and it becomes problematic to round an approximate design to an exact design with replicates that sum to n.

Other reasons for not working with exact designs even for large sample problems, are : (a) using theory to find an optimal exact design is nearly theoretically impossible for almost all non-linear models with multiple parameters; (b) the solution depends sensitively on each assumption of the model and a slight mis-specification will quickly invalidate the proof; (c) there is no general algorithm for finding an optimal exact design and no unified method for confirming an exact design is an optimal exact design, and (d) they depend on n, the statistical model, the design criterion and one would need to provide an endless list of tables of different types of optimal exact designs for practitioners. More importantly, Kiefer (1983) argued that when the sample size is large, the difference in efficiencies between the true and unknown optimal exact design and the one obtained by rounding the optimal approximate design into an optimal exact design directly is no larger than $O(1/n^2)$ for any intuitive rounding methods.

Optimal designs are meant to serve as a rough guide for the practitioners to select an efficient design. They are not meant to be implemented strictly since they may not meet requirements of the user. For instance, the optimal exact designs may have too few design points than the user wants in the experiment, or the user wants to always include a 0 dose that is not a support point of the optimal design.

There are recent work on constructing optimal exact designs, and as expected, they are more difficult and laborious to find than optimal approximate designs. Some recent work include Chen, et al. (2023), who found optimal exact designs for longitudinal models, and both Duarte et al. (2020) and Vaquez, et al. (2023), who used mixed integer programming to find different types of optimal exact designs for different scenarios. Since these tools require specialized software and good solvers, practitioners may prefer to use optimal approximate designs, which likely should suffice in practice especially for large sample problems.

Acknowledgments

Dr. Liu and Dr. Yue were partially supported by the National Natural Science Foundation of China (No.11971318) and Natural Science Foundation of Shanghai (No.24ZR401500). Wong was partially supported by the Yushan Fellowship Award from the Ministry of Education (MOE), Taiwan and he is grateful for the additional support and hospitality from The National Cheng Kung University in Tainan, Taiwan.

References

Al-Sharadqah, A. and Nguyen, P. (2022). Statistical analysis of concentric objects estimation problem under the heteroscedastic Berman model.*Journal of Statistical Planning and Inference* **218**, 9–24.

Berman, M. (1983). Estimating the parameters of a circle when angular differences are known. *J. Roy. Statist. Soc. Ser. C* **32**, 1–6.

Butler, G.J. and Wolkowicz, G.S.K. (1985). A mathematical model of the chemostat with a general class of functions describing nutrient uptake. *SIAM Journal on Applied Mathematics* **45**, 138–151.

Chakraborty, Bibhas, Ghosh, Palash, Moode, Erica E. M. and Rush, J. A. (2016). Estimating optimal shared-parameter dynamic regimens with application to a multistage depression clinical trial. *Biometrics* **73**, 865–876.

Chen, R. B., Chen, P. Y., Hsu, C. L. and Wong, W. K. (2020). Hybrid algorithms for generating optimal designs for discriminating multiple nonlinear models under various error distributional assumptions. *PlosOne* **15(10)**:e0239864.

Chen, P. Y., Chen, R. B. and Wong, W. K. (2022). Particle swarm optimization for searching efficient experimental designs – a review. *Wires Journal: Computational Statistics* **14**, e1578. doi:abs/10.1002/wics.1578.

Chen, P. Y., Chen, R. B. and Wong, W. K. (2023). Particle swarm optimization for finding efficient longitudinal exact designs for nonlinear models. *The New England Journal of Statistics in Data Science* **1**, 299–313.

Cook, R. D. and Wong, W. K. (1994). On the equivalence of constrained and compound optimal designs. *Journal of the American Statistician Association* **89**, 687–692.

Dette, H., Melas, V. B. and Wong, W. K. (2005). Optimal designs for goodness of fit of the Michaelis-Menten enzyme kinetic function. *Journal of American Statistical Association* **100**, 1370–1381.

Dette, H., Kiss, C., Bevanda, M. and Bretz, F. (2010). Optimal designs for the EMAX, log-linear and exponential models. *Biometrika* **97**, 513–518.

Dette, H., Pepelyshev, A. and Wong, W. K. (2011). Experimental design strategies for detecting hormesis. *Risk Analysis* **31**, 1949–1960.

Duarte, B. P. M., José F. O. Granjo abd Wong, W. K. (2020). Optimal exact designs of experiments via Mixed Integer Nonlinear Programming. *statistics & Computing* **30**, 93–112.

Faya P., Rauk A. P., Griffiths K. L. and Parekh B. (2020). A curve similarity approach to parallelism testing in bioassay. *J.Biopharm. Stat.* **30**, 721–733.

Feller, C., Schorning, K., Dette, H., Bermann, G. and Bornkamp, B. (2017). Optimal designs for dose response curves with common parameters. *Annals of Statistics* **45**, 2102–2132.

Fleetwood K., Bursa F. and Yellowlees A. (2015). Parallelism in practice: approaches to parallelism in bioassays. *PDA. J. Pharm. Sci. Technol.* **69**, 248–263.

Huang, M. N. L., Chen, R. B., Lin, C. S. and Wong, W. K. (2006). Optimal designs for parallel models with correlated responses. *Statist. Sinica* **16**, 121–133.

Liu, X., Yue, R. X., Zhang, Z. and Wong, W. K. (2021). G-optimal designs for hierarchical linear models: an equivalence theorem and a nature-inspired meta-heuristic algorithm. *Soft Computing* **5**, 13549–13565.

Lopez-Fidalgo, J. and Wong, W. K. (2002). Optimal designs for the Michaelis-Menten model. *Journal of Theoretical Biology* **215**, 1–11.

Love, S. B., Brown, S., Weir, C. J., Harbron, C., Yap, C., Gascher-Markefski, B., Matcham, J., Caffrey, L., McKevitt, C., Clive, S., Craddock, C., Spicer, J. and Cornelius, V. (2017). Embracing model-based designs for dosefinding trials. *British Journal of Cancer* **117**, 332–339.

Masoudi, E., Holling, H., Duarte, B. P. M. and Wong, W. K. (2019). Metaheuristic adaptive cubature based algorithm to find Bayesian optimal designs for nonlinear models. *Journal of Computational and Graphical Statistics* **28**, 861–876.

Novick, S. and Harry Yang, H. (2019). A fast and reliable test for parallelism in bioassay. *J. Biopharm. Stat.* **29**, 1011–1023.

Pinheiro, J. Bretz, F. and Branson, M. (2006). Analysis of dose response studies: Modeling approaches. Pp. 146-171 in Ting, N. (ed.) *Dose finding in drug development.* Springer, New York.

Pukelsheim, F. and Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika* **79**, 763–770.

Vaquez, A., Wong, W. K. and Goos, P. (2022). Constructing two-level-optimal screening designs using mixed-integer programming and heuristic algorithms. *Statistics & Computing* **33**, DOI https://doi.org/10.1007/s11222-022-10168-.

Yu, X. Z. and Gu, J. D. (2007). Differences in Michaelis-Menten kinetics for different cultivars of maize during cyanide removal. *Ecotoxicology and Environmental Safety* **67**, 254–259.

School of Mathematics and Statistics, Donghua University, Shanghai 201620, P. R. China.

E-mail: (liuxin@dhu.edu.cn)

School of Arts and Sciences, Fuyao University Sciencee and Technology, Fuzhou 350109, P. R. China; College of Mathematics and Science, Shanghai Normal University, Shanghai 200234, P. R. China.

E-mail: (yue2@shnu.edu.cn)

Department of Biostatistics, University of California, Los Angeles, CA, USA.

E-mail: (wkwong@ucla.edu)