Statistica Sinica **29** (2019), 283-302
doi:https://doi.org/10.5705/ss.202017.0020

OPTIMAL DESIGNS FOR NONLINEAR MODELS WITH RANDOM BLOCK EFFECTS

Xin Wang[1], Min Yang[1] and Wei Zheng[2]

[1] *University of Illinois at Chicago, and* [2] *University of Tennessee*

Abstract: Optimal designs for nonlinear model with random block effects are systematically studied. For a large class of nonlinear models, we prove that any optimal design can be based on some simple structure. We further derive the corresponding general equivalence theorem. This allows us to propose an efficient algorithm for deriving specific optimal designs. The application of the algorithm is demonstrated through deriving a variety of locally optimal designs and accessing their robustness under different nonlinear models.

Key words and phrases: Algorithm, A-optimality, D-optimality, Loewner ordering.

1. Introduction

Nonlinear models have found broad applicability during the last decades. They have been applied in such fields as drug discovery, clinical trials, social sciences, marketing, etc. Methods of analysis and inference for these models are well established (see for example McCullagh and Nelder (1989); McCulloch and Searle (2001)). While using nonlinear models to analyze such data has become common with advances in computational tools, the study of optimal design for such problems is far behind the current use of nonlinear models, especially when observations are correlated.

An optimal/efficient design can reduce the sample size needed for achieving pre-specified precision of estimation or improving the precision of estimation for the fixed sample size. While the importance of optimal design cannot be overstated, there are many scientific problems for which tools that can help to identify optimal or efficient designs are simply inadequate, not infrequently leading to the use of inferior designs. This inadequacy is partly due to the fact that identifying optimal designs is a very challenging problem, especially for nonlinear models. As a result, solutions have often been developed on a case-by-case basis, requiring a separate proof for each combination of model, objective, and optimality criterion.

Recently, a series papers by Yang and Stufken (2009); Yang (2010); Dette and Melas (2011); Stufken and Yang (2012); Yang and Stufken (2012); and Dette and Schorning (2013) discovered if the functions that are elements of the information matrix generate the so called Chebyshev system, the number of support points in locally optimal designs is small and often is equal to the number of parameters to be estimated (saturated designs). The new tools simplify the process of deriving optimal designs, and most of the available optimality results for nonlinear models can be derived as special cases with the new tools.

These results focus on the situation where observations are independent, in which the information matrix has the "additive" property: the information matrix of a design can be written as the summation of the information matrices at each point. When the observations are correlated, the additive property does not hold anymore. This new framework does not apply. Even the celebrated general equivalence theorem, which allows us to verify a design is indeed optimal, is no longer available. Relatively little is known of how to produce optimal designs for nonlinear models when the observations are correlated. Müller and Pázman (1999) presented an iterative algorithm for regression models with correlated error. Pázman (2010) studied contribution of information from subset of finite design points when correlated observations are indicated. Dette et al. (2010) derived asymptotic optimal design for population pharmacokinetics model with random effects. Kiefer and Wynn (1981) discussed optimal balanced block and Latin square designs for linear model with various correlation structures. Kunert, Martin and Eccleston (2010) and Cutler (1993) considered optimal design for comparing treatment and control effects under autoregressive correlation structure. Atkinson (2008) gave some examples applying an equivalence theorem for D-optimal in constructing optimal design for nonlinear model with correlated observations. Ucinski and Atkinson (2004) studied design for nonlinear time-dependent models. Dette and Kunert (2014) studied optimal design for Michaelis-Menten model and Holland-Letz, Dette and Renard (2012) proposed an algorithm approach of deriving optimal design based on linear approximation.

With random block effects, Cheng (1995) and Atkins and Cheng (1999) studied optimal design under linear models. Recently, Huang and Cheng (2016) extended their results to quadratic regression with block size two. In this manuscript, we consider a class of nonlinear models with arbitrary block size. We prove that any optimal design can be based on a simple structure. We further derive the corresponding general equivalence theorem under the correlated errors structure. This result allows us to propose an efficient algorithm of deriving specific optimal designs. Our approach works for all general non-linear models and provides a strategy of searching specific optimal designs.

For the layout of the remainder of this paper, in Section 2, we introduce the model and the information matrix. In Section 3, we show that searching for optimal designs can be restricted to those with identical groups and demonstrate a "complete class" result for several specific nonlinear models. This result allows us to focus on a specific structure when we derive any optimal design. In Section 4, we derive the corresponding general equivalence theorem and propose an efficient algorithm for deriving D-optimal and A-optimal designs. It is understood that the algorithm can be extended to other optimalities readily. Some numerical examples are given to demonstrate the results in Section 5. Saturated D-optimal design and robustness issue are also discussed in this section. A short discussion is given in Section 6. Some lengthy proofs are postponed to the appendix, which could be found in the on-line supplement material.

2. Model Setup and Information Matrix

Suppose there are b groups, each having k observations. Consider a nonlinear model

$$y_{ij} = f_\theta(x_{ij}) + \epsilon_{ij}, 1 \le i \le b, 1 \le j \le k,$$

where $f_\theta(\cdot)$ is a smooth function with its form only depending on the parameter θ to be estimated, y_{ij} is the response of the jth unit of the ith group, x_{ij} is the corresponding design point in a given design region, say $\mathcal{X}$. Here we assume ϵ_{ij} to be normally distributed with a constant variance σ^2. Observations in same group are assumed to have equal correlation coefficient ρ and those in different groups are uncorrelated. For the sake of finding optimal designs, we set $\sigma^2 = 1$ without loss of generality. Then, for group i we have

$$\begin{aligned} E(\boldsymbol{Y_i}) &= f_\theta(\boldsymbol{X_i}), \\ Cov(\boldsymbol{Y_i}) &= (1-\rho)\boldsymbol{I_k} + \rho \boldsymbol{J_k} := \boldsymbol{V}, \end{aligned} \tag{2.1}$$

where $\boldsymbol{Y_i} = (y_{i1}, \ldots, y_{ik})^T$, $\boldsymbol{X_i} = (x_{i1}, \ldots, x_{ik})^T \in \mathcal{X}^k$, $f_\theta(\boldsymbol{X_i}) = \{f_\theta(x_{i1}), \ldots, f_\theta(x_{ik})\}^T$, $\boldsymbol{I_k}$ is the $k \times k$ identify matrix, and $\boldsymbol{J_k}$ is the $k \times k$ matrix with all elements 1. Since the covariance matrix is completely symmetric, the order of the components x_{ij} in $\boldsymbol{X_i}$ is irrelevant from a design perspective. Suppose the components of $\boldsymbol{X_i}$ consist of m_i distinct points, say $\{x_{i1}, \ldots, x_{im_i}\}$ with corresponding number of replications as $\{k_{i1}, \ldots, k_{im_i}\}$. Automatically we have $\sum_{j=1}^{m_i} k_{ij} = k$. By direct calculations, the information matrix of group i regarding θ is

$$\boldsymbol{M_i} = \boldsymbol{F_i^T} diag(\boldsymbol{1^T_{k_{i1}}}, \boldsymbol{1^T_{k_{i2}}}, \ldots, \boldsymbol{1^T_{k i_{m_i}}}) \boldsymbol{V^{-1}} diag(\boldsymbol{1_{k_{i1}}}, \boldsymbol{1_{k_{i2}}}, \ldots, \boldsymbol{1_{k_{i_{m_i}}}}) \boldsymbol{F_i}.$$
$$= c_1(\rho) \boldsymbol{F_i^T} diag(k_{i1}, \ldots, k_{im_i}) \boldsymbol{F_i} - k^{-1} c_2(\rho, k) \boldsymbol{F_i^T} (k_{i1}, \ldots, k_{im_i})^T (k_{i1}, \ldots, k_{im_i}) \boldsymbol{F_i}. \quad (2.2)$$

where $\boldsymbol{F_i} = (g(x_{i1}), \ldots, g(x_{im_i}))^T$ with $g(x_{ij}) = \partial f_\theta(x_{ij})/\partial \boldsymbol{\theta}$, $c_1(\rho) = (1-\rho)^{-1}$, $c_2(\rho, k) = k\rho \{1 + (k-1)\rho\}^{-1} c_1(\rho)$, and $\boldsymbol{1_k}$ is the $k \times 1$ vector with all elements 1. Here we utilized the fact $\boldsymbol{V}^{-1} = c_1(\rho) \boldsymbol{I}_k - k^{-1} c_2(\rho, k) \boldsymbol{J}_k$. In the sequel we abbreviate $c_1(\rho)$ and $c_2(\rho, k)$ by c_1 and c_2, respectively, unless there is a necessity to emphasise their dependence on ρ and k. If $w_{ij} = k_{ij}/k$ and $\boldsymbol{W_i} = diag(w_{i1}, \ldots, w_{im_i})$, then $\boldsymbol{M_i}$ can be written as

$$\boldsymbol{M_i} \;\; = \;\; \boldsymbol{F_i^T} \left(c_1 k \boldsymbol{W_i} - c_2 k \boldsymbol{W_i} \boldsymbol{J_{m_i}} \boldsymbol{W_i} \right) \boldsymbol{F_i}. \quad (2.3)$$

It turns out the information matrix $\boldsymbol{M_i}$ depends on the group size k and the design measure of group i, $\boldsymbol{\xi_i} = \{(x_{ij}, w_{ij}), j = 1, \ldots, m_i\}$ with $w_{ij} = k_{ij}/k$. In the classical approximate design theory, we denote the information matrix of $\boldsymbol{\xi_i}$ by

$$\begin{aligned} \boldsymbol{M}(\boldsymbol{\xi_i}) &= \frac{\boldsymbol{M_i}}{k} = \boldsymbol{F_i^T} \left(c_1 \boldsymbol{W_i} - c_2 \boldsymbol{W_i} \boldsymbol{J_{m_i}} \boldsymbol{W_i} \right) \boldsymbol{F_i}, \\ &= c_1 \int g(x) g(x)^T \boldsymbol{\xi_i}(dx) - c_2 \left\{ \int g(x) \boldsymbol{\xi_i}(dx) \right\} \left\{ \int g(x) \boldsymbol{\xi_i}(dx) \right\}^T. \end{aligned} \quad (2.4)$$

Since there is no between-group correlations, we have the model for the full data as

$$\begin{aligned} E(\boldsymbol{Y}) &= f_\theta(\boldsymbol{X}), \\ Cov(\boldsymbol{Y}) &= \boldsymbol{I_b} \otimes \boldsymbol{V}, \end{aligned} \quad (2.5)$$

where $\boldsymbol{Y} = (\boldsymbol{Y_1^T}, \ldots, \boldsymbol{Y_b^T})^T$, $\boldsymbol{X} = (\boldsymbol{X_1^T}, \ldots, \boldsymbol{X_b^T})^T$ and $f_\theta(\boldsymbol{X}) = \{ f_\theta(\boldsymbol{X_1})^T, \ldots, f_\theta(\boldsymbol{X_b})^T \}^T$. Suppose there are b^* distinct groups designs $\{\boldsymbol{\xi_1}, \ldots, \boldsymbol{\xi_{b^*}}\}$, which appear $\{n_1, \ldots, n_{b^*}\}$ times with the restriction $\sum_{i=1}^{b^*} n_i = b$. Denote the whole design measure by $\boldsymbol{\delta} = \{(\boldsymbol{\xi_i}, \zeta(\boldsymbol{\xi_i})), i = 1, \ldots, b^*\}$, where $\zeta(\boldsymbol{\xi_i}) = n_i/b$. Then the information matrix of $\boldsymbol{\delta}$ is

$$\boldsymbol{M}(\boldsymbol{\delta}) = \sum_{i=1}^{b^*} \zeta(\boldsymbol{\xi_i}) \boldsymbol{M}(\boldsymbol{\xi_i}). \quad (2.6)$$

In approximate design theory, we relax the integer constraints on k_{ij} and n_i and work on the space $\{\xi_i : \sum_{j=1}^{m_i} w_{ij} = 1, w_{ij} \geq 0\}$ for ξ_i and $\{\zeta : \sum_{i=1}^{b^*} \zeta(\boldsymbol{\xi_i}) = 1, \zeta(\boldsymbol{\xi_i}) \geq 0\}$ for ζ.

3. Complete Class of Designs

In this section, we find the complete class, a subclass of designs containing the optimal designs under various design criteria simultaneously. The designs in the derived complete class have very small (mostly minimum) number of supporting points, which facilitates the numerical search of specific optimal designs. As compared to existing results on complete class, (2.5) imposes additional challenges. There are two layers of approximate designs as represented by (2.4) and (2.6), and the information matrix in (2.4) does not possess the desirable additivity property as in most studies. We establish complete classes separately for the two layers.

3.1. Complete class of between-group designs

By (2.4), the within-group information matrix under a design, say $\boldsymbol{\xi}$, can be represented by

$$M(\boldsymbol{\xi}) = c_1 L(\boldsymbol{\xi}) - c_2 G(\boldsymbol{\xi}) G(\boldsymbol{\xi})^T, \tag{3.1}$$

$$L(\boldsymbol{\xi}) = \int g(x) g(x)^T \boldsymbol{\xi}(dx), \tag{3.2}$$

$$G(\boldsymbol{\xi}) = \int g(x) \boldsymbol{\xi}(dx). \tag{3.3}$$

Lemma 1. *$M(\boldsymbol{\xi})$ is concave in $\boldsymbol{\xi}$ by Lowner's ordering.*

Proof. Since $L(\boldsymbol{\xi})$ is linear in $\boldsymbol{\xi}$, it is sufficient to show that $G(\boldsymbol{\xi})G(\boldsymbol{\xi})^T$ is convex in $\boldsymbol{\xi}$, in view of $c_2 > 0$. For a constant $0 < \alpha < 1$ and two measures $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$, we have

$$\begin{aligned}&\alpha G(\boldsymbol{\xi}_1)G(\boldsymbol{\xi}_1)^T + (1-\alpha)G(\boldsymbol{\xi}_2)G(\boldsymbol{\xi}_2)^T - G(\alpha\boldsymbol{\xi}_1 + (1-\alpha)\boldsymbol{\xi}_2)G(\alpha\boldsymbol{\xi}_1 + (1-\alpha)\boldsymbol{\xi}_2)^T \\ &= \alpha(1-\alpha)\left\{G(\boldsymbol{\xi}_1) - G(\boldsymbol{\xi}_2)\right\}\left\{G(\boldsymbol{\xi}_1) - G(\boldsymbol{\xi}_2)\right\}^T \geq 0.\end{aligned}$$

Hence, the proof is completed.

Theorem 1. *Consider approximate designs under* (2.5)*. If for any design $\boldsymbol{\delta} = \{(\boldsymbol{\xi_i}, \zeta(\boldsymbol{\xi_i})) | i = 1, \ldots, b^*\}$, $\boldsymbol{\delta}^* = (\bar{\boldsymbol{\xi}}, 1)$ with $\bar{\boldsymbol{\xi}} = \sum_{i=1}^{b^*} \zeta(\boldsymbol{\xi_i})\boldsymbol{\xi_i}$. Then*

$$\boldsymbol{M}(\boldsymbol{\delta}^*) \geq \boldsymbol{M}(\boldsymbol{\delta}). \tag{3.4}$$

by Loewner's ordering.

This theorem is a direct result of Lemma 1 through Jensen's Inequality. This result is similar to that of Schmelter (2007), where the mixed effects model with uncorrelated error terms was studied. Theorem 1 indicates that we can focus on the class of designs which have identical design in each group. This greatly simplifies the procedure of deriving approximate optimal design.

3.2. Complete class of within-group designs

Even though the within-group information matrix does not share the desirable additive property of traditional design problems, it is still possible to identify the complete class as in Theorem 1 in Yang (2010). We can show that only a small number of support points are necessary to achieve optimal design under Model (2.5). First, there exists a $p \times p$ nonsingular transformation matrix $P(\boldsymbol{\theta})$, such that the (3.1) can be written as

$$\boldsymbol{M(\xi)} = P(\boldsymbol{\theta}) \left\{ c_1 \sum_{j=1}^{N} w_j \Phi_1(C_j) - c_2 \sum_{j=1}^{N} w_j \Phi_2(C_j) \sum_{j=1}^{N} w_j \Phi_2(C_j)^T \right\} P(\boldsymbol{\theta})^T, \tag{3.5}$$

where $\Phi_2(C_j) = (\phi_{01}(C_j), \ldots, \phi_{0p}(C_j))^T$,

$$\Phi_1(C_j) = \begin{pmatrix} \phi_{11}(C_j) & \phi_{12}(C_j) & \ldots & \phi_{1p}(C_j) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{1p}(C_j) & \phi_{2p}(C_j) & \ldots & \phi_{pp}(C_j) \end{pmatrix},$$

$P(\boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}$ only and does not depend on x_j or C_j. C_j may depend on $\boldsymbol{\theta}$, and is a one-to-one map from $x_j \in \mathcal{X}$ to $[A, B]$. Thus, it is equivalent to write $\boldsymbol{\xi}$ as $\boldsymbol{\xi} = \{(C_i, w_i), i = 1, \ldots, m\}$.

locally optimal design context, for any two given designs $\boldsymbol{\xi} = \{(C_i, w_i), i = 1, \ldots, m\}$ and $\boldsymbol{\xi^*} = \{(\tilde{C}_i, \tilde{w}_i), i = 1, \ldots, m^*\}$, to show that $\boldsymbol{M_\xi} \leq \boldsymbol{M_{\xi^*}}$, it suffices to show that

$$\sum_{i=1}^{m} w_i \phi_{lt}(C_i) = \sum_{i=1}^{m^*} \tilde{w}_i \phi_{lt}(\tilde{C}_i), \tag{3.6}$$

and that for all $l = 0, 1, \ldots, p, t = 1, \ldots, p$ except for one $l = t$,

$$\sum_{i=1}^{m} w_i \phi_{ll}(C_i) \leq \sum_{i=1}^{m^*} \tilde{w}_i \phi_{ll}(\tilde{C}_i). \tag{3.7}$$

Theorem 2. *Suppose for* (2.5) *there exists a matrix $P(\boldsymbol{\theta})$ s.t. its information matrix can be written in the form of* (3.5)*. Let $\{\phi_1, \ldots, \phi_n\}$ be the set of distinct functions from $\{\phi_{01}, \ldots, \phi_{pp}\}$ in* (3.5)*, which are defined on $[A, B]$, and $\Gamma(C) = \prod_{l=1}^{n} \gamma_{ll}(C), \forall C \in [A, B]$, where*

$$\gamma_{lt} = \begin{cases} \phi_l'(C) & t = 1, l = 1, \ldots, n, \\ \left\{ \dfrac{\gamma_{l,t-1}(C)}{\gamma_{t-1,t-1}(C)} \right\}' & 2 \leq t \leq n, t \leq l \leq n. \end{cases} \tag{3.8}$$

For any given design $\xi = \{(C_j, w_j), j = 1, \ldots, N\}$, there always exists a design

$\tilde{\xi} = \{(\tilde{C}_j, \tilde{w}_j)\}$ *such that*

$$\boldsymbol{M}_{\boldsymbol{\xi}} \leq \boldsymbol{M}_{\tilde{\boldsymbol{\xi}}} \tag{3.9}$$

with respect to Loewner ordering.

(a) When $n = 2m-1, N \geq m$ *and* $\Gamma(C) < 0$ *for* $C \in [A, B]$, $\tilde{\boldsymbol{\xi}}$ *has* m *support points and one of them is* A*;*

(b) When $n = 2m-1, N \geq m$ *and* $\Gamma(C) > 0$ *for* $C \in [A, B]$, $\tilde{\boldsymbol{\xi}}$ *has* m *support points and one of them is* B*;*

(c) When $n = 2m, N \geq m$ *and* $\Gamma(C) > 0$ *for* $C \in [A, B]$, $\tilde{\boldsymbol{\xi}}$ *has* $m+1$ *support points and two of them are* A *and* B*;*

(d) When $n = 2m, N \geq m+1$ *and* $\Gamma(C) < 0$ *for* $C \in [A, B]$, $\tilde{\boldsymbol{\xi}}$ *has* m *support points.*

Remark 1. In Theorems 1 and 2, we consider optimality with respect to Loewner's ordering, which is stronger than most commonly used optimal criteria, like A-, D- and E- optimality.

The proof is skipped because it is a direct application of Theorem 1 in Yang (2010). Theorem 2 allows us to restrict the search of optimal within-group designs to a small subclass, where designs typically have minimum number of distinct support points. This greatly reduces the computational burden.

4. Numerical Search of Optimal Design

There remains a challenge in finding a specific optimal design for a given model and optimality criterion. For the example of the exponential model in Section 5, by Theorems 1 and 2, we can focus on the class of designs with at most three points, one of which is the upper bound. We need therefore to find the remaining two design points and their weights. The classical general equivalence theorem (GET) is a powerful device for verifying the optimality of a candidate design, but, existing results on GET are based on the assumption that the observations are independent, which is not true here. In this section, a new version of GET under Model (2.5) is derived and an efficient algorithm is proposed. We focus on the A- and D-optimal criteria for the algorithm with the understanding that the algorithm can be readily extended to other optimality criterion.

4.1. The general equivalence theorem

A-optimal design minimizes the average (or sum) of variances of the parameter estimators, $\min_{\xi} Tr(\boldsymbol{M}(\boldsymbol{\xi})^{-1})$. A D-optimal design minimizes the volume of the confidence region of the estimators, $\min_\xi |\boldsymbol{M}(\boldsymbol{\xi})^{-1}|$. Kiefer (1974) unified these criteria by the function $\Phi_p(\boldsymbol{M}(\boldsymbol{\xi})) = \{(1/v)Tr(\boldsymbol{M}(\boldsymbol{\xi})^{-p})\}^{1/p}$, where the D- and A- criteria are special cases of $p=0$ and $p=1$, respectively. Here $p=0$ is understood as $\lim_{p\to 0}\Phi_p(\boldsymbol{M}(\boldsymbol{\xi})^{-1}) = |\boldsymbol{M}(\boldsymbol{\xi})^{-1}|^{1/v}$. As $\Phi_p(\boldsymbol{M}(\boldsymbol{\xi}))$ is convex in $\boldsymbol{M}(\boldsymbol{\xi})$ (Fedorov and Hackl (1997, Sec. 2.2)), Lemma 1 allows us to establish the following.

Lemma 2. *$\Phi_p(\boldsymbol{M}(\boldsymbol{\xi}))$ is convex in ξ.*

Theorem 3. *A within-group design, $\boldsymbol{\xi}$, minimizes $\Phi_p(\boldsymbol{M}(\boldsymbol{\xi}))$ if and only if*

$$\min_{x\in\chi}\eta(x,\boldsymbol{\xi}) = Tr(D(\boldsymbol{\xi})\psi(\boldsymbol{\xi},\boldsymbol{\xi})), \tag{4.1}$$

where

$$\begin{aligned} D(\boldsymbol{\xi}) &= \left.\frac{\partial\Phi_p(\boldsymbol{M})}{\partial\boldsymbol{M}}\right|_{\boldsymbol{M}=\boldsymbol{M}(\boldsymbol{\xi})}, \\ \psi(\boldsymbol{\nu},\boldsymbol{\xi}) &= c_1L(\boldsymbol{\nu}) - c_2\left\{G(\boldsymbol{\nu})G(\boldsymbol{\xi})^T + G(\boldsymbol{\xi})G(\boldsymbol{\nu})^T\right\}, \\ \eta(x,\boldsymbol{\xi}) &= c_1g(x)^TD(\boldsymbol{\xi})g(x) - 2c_2G(\boldsymbol{\xi})^TD(\boldsymbol{\xi})g(x), \end{aligned}$$

and $L(\boldsymbol{\xi})$ and $G(\boldsymbol{\xi})$ are defined as in (3.2) *and* (3.3). *Moreover, all supporting points of $\boldsymbol{\xi}$ satisfy the equality in* (4.1).

Proof. By direct calculation we have

$$\left.\frac{\partial\boldsymbol{M}((1-\alpha)\boldsymbol{\xi}+\boldsymbol{\alpha\nu})}{\partial\alpha}\right|_{\alpha=0} = \psi(\boldsymbol{\nu},\boldsymbol{\xi}) - \psi(\boldsymbol{\xi},\boldsymbol{\xi}). \tag{4.2}$$

By Lemma 2, $\boldsymbol{\xi}$ is Φ_p-optimal if and only if

$$\begin{aligned} 0 \le& \left.\frac{\partial\Phi_p(\boldsymbol{M}((1-\alpha)\boldsymbol{\xi}+\alpha\boldsymbol{\nu}))}{\partial\alpha}\right|_{\alpha=0} \\ &= Tr(D(\boldsymbol{\xi})\{\psi(\boldsymbol{\nu},\boldsymbol{\xi}) - \psi(\boldsymbol{\xi},\boldsymbol{\xi})\}), \end{aligned} \tag{4.3}$$

for any design $\boldsymbol{\nu}$. If ν_x is a degenerated design supported on only one point x, then

$$Tr(D(\boldsymbol{\xi})\psi(\nu_x,\boldsymbol{\xi})) = \eta(x,\boldsymbol{\xi}). \tag{4.4}$$

By (4.3), we have

$$\min_{\xi\in\chi}\eta(x,\boldsymbol{\xi}) \ge Tr(D(\boldsymbol{\xi})\psi(\boldsymbol{\xi},\boldsymbol{\xi})). \tag{4.5}$$

Due to (4.4) and $\int\psi(\boldsymbol{\nu},\boldsymbol{\xi})\boldsymbol{\xi}(dx) = \psi(\boldsymbol{\xi},\boldsymbol{\xi})$, we have

$$\int\eta(x,\boldsymbol{\xi})\boldsymbol{\xi}(dx) = Tr(D(\boldsymbol{\xi})\psi(\boldsymbol{\xi},\boldsymbol{\xi})). \tag{4.6}$$

which implies

$$\min_{\xi\in\chi}\eta(x,\boldsymbol{\xi}) \le Tr(D(\boldsymbol{\xi})\psi(\boldsymbol{\xi},\boldsymbol{\xi})). \tag{4.7}$$

The proof is complete in view of (4.5)-(4.7).

Remark 2. We find $D(\boldsymbol{\xi}) = -(1/v)\boldsymbol{M(\xi)}^{-2}$ and $-(1/v)|\boldsymbol{M(\xi)}|^{1/v}\boldsymbol{M(\xi)}^{-1}$, respectively. It can be shown that condition (4.1) is equivalent to $\max_{x\in\mathcal{X}} d(\boldsymbol{\xi}) \leq 0$, where

$$d(\boldsymbol{\xi}, x) = \begin{cases} c_1 g(x)^T \boldsymbol{M(\xi)}^{-2} g(x) - 2c_2 G(\boldsymbol{\xi})^T \boldsymbol{M(\xi)}^{-2} g(x) & \text{A-optimal,} \\ -Tr(\boldsymbol{M(\xi)}^{-2}\psi(\boldsymbol{\xi},\boldsymbol{\xi})), & \\ c_1 g(x)^T \boldsymbol{M(\xi)}^{-1} g(x) - 2c_2 G(\boldsymbol{\xi})^T \boldsymbol{M(\xi)}^{-1} g(x) & \text{D-optimal.} \\ -Tr(\boldsymbol{M(\xi)}^{-1}\psi(\boldsymbol{\xi},\boldsymbol{\xi})), & \end{cases} \tag{4.8}$$

4.2. Optimal weights for given support points

In this section, we propose an algorithm suggested by the optimal weights exchange algorithm (OWEA) of Yang, Biedermann and Tang (2013). The OWEA can be viewed as an extension of the Fedorov-Wynn algorithm (Wynn (1970), Fedorov (1972)) that adds an optimization step for the weights, but this step is for the model with independent observation at each design point. Theorems 4 and 5 show that such technique can be extended to the correlated errors case under D- and A- optimality criteria. Although the two theorems can be proved through the convexity of $\Phi_p(\boldsymbol{M(\xi)})$ (Lemma 2), we give different proofs in the appendix by showing the corresponding Hessian matrix is a nonnegative definite matrix. The proofs provide the needed expressions of the Gradient vector and Hessian matrix in the deriving of optimal weights.

Notice that the D- and A-optimality criteria are equivalent to minimizing

$$\tilde{\Phi}_p(\boldsymbol{\xi}) = \begin{cases} \log|\Sigma_{\boldsymbol{\xi}}(\theta)|, & \text{if } p = 0, \\ Tr(\Sigma_{\boldsymbol{\xi}}(\theta)), & \text{if } p = 1. \end{cases} \tag{4.9}$$

Let $\boldsymbol{\xi} = \{(x_i, w_i), i = 1, \ldots, n\}$ be the within-group design. Take $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)^T$ and $\Omega = \{\omega_i \geq 0, i = 1, \ldots, n-1, \sum_{i=1}^{n-1} \omega_i \leq 1\}$. For a given set of support points, Theorems 4 and 5 provde direct support that the A- and D-optimal criteria functions are convex with respect to the weight vector, as well as expressions (first and second derivative in their proof) which helps derive the algorithm in Section 4.3.

Theorem 4. *The minimum value of* $\log|\Sigma_\xi(\boldsymbol{\theta})|$*, as a function of* $\boldsymbol{w}$*, is achieved at any critical point in* Ω *or at the boundary of* Ω*.*

Theorem 5. *The minimum value of* $Tr(\Sigma_\xi(\boldsymbol{\theta}))$*, as a function of* $\boldsymbol{w}$*, is achieved at any critical point in* Ω *or at the boundary of* Ω*.*

4.3. Implementation of the algorithm

When the design space is continuous, we shall the design space to $\mathcal{X}_n$, which is the collection of n evenly spaced points in $\mathcal{X}$. If $\mathcal{X}$ is discrete, let $\mathcal{X}_n = \mathcal{X}$. The algorithm proceeds as follows.

1. Initialization. Set $S^{(0)}$ to be the set of $m+1$ design points uniformly distributed in χ_n, where m is the parameter in Theorem 2. Derive the optimal design $\boldsymbol{\xi}_0$ for the given initial support points with uniform initial weights.

2. Update. At iteration $t \geq 1$, derive the set of supporting points
$$S^{(t)} = S^{(t-1)} \cup \{x_t^*\}, \text{where } x_t^* = \arg\max_{x \in \chi_n} d(\boldsymbol{\xi}_{t-1}, x), \tag{4.10}$$
and $d(\boldsymbol{\xi}, x)$ is defined as in (4.8). Derive $\boldsymbol{\xi}_t$, the optimal design on the supporting set $S^{(t)}$. The weight in $\boldsymbol{\xi}_{t-1}$ is the initial solution in deriving the weights in $\boldsymbol{\xi}_t$. Points with zero weight in $\boldsymbol{\xi}_t$ shall be removed from $S^{(t)}$.

3. Stopping rule. If $\max_{x \in \chi_n} d(\boldsymbol{\xi}_t, x) \leq \epsilon_0$, for some pre-specified value of ϵ_0, stop and output $\boldsymbol{\xi}_t$ as the optimal design. Otherwise, go back to the updating step.

We give more details for deriving the optimal weight in the update step of the algorithm. This is a modification of the classical Newton-Raphson method. Let $\boldsymbol{w}_0^{(t)}$ be the initial candidate value of the weight for $\boldsymbol{\xi}_t$, $\boldsymbol{w}_j^{(t)}$ is value at the jth iteration Algorithm proceeds as follows.

(a) $\boldsymbol{w}_{j+1}^{(t)} = \boldsymbol{w}_j^{(t)} - a\left\{(\partial^2\tilde{\Phi}_p^{(t)})/(\partial\boldsymbol{w}\partial\boldsymbol{w}^T)\right\}^{-1}(\partial\tilde{\Phi}_p^{(t)})/(\partial\boldsymbol{w})$. Expressions of $(\partial^2\tilde{\Phi}_p^{(t)})/(\partial\boldsymbol{w}\partial\boldsymbol{w}^T)$, $(\partial\tilde{\Phi}_p^{(t)})/(\partial\boldsymbol{w})$ can be found in proof of Theorem 4 and 5 in the appendix in the on-line supplement material.

(b) If there are non-positive components in $\boldsymbol{w}_{j+1}^{(t)}$, go to (d), otherwise go to (c).

(c) If $||\boldsymbol{w}_{j+1}^{(t)} - \boldsymbol{w}_j^{(t)}|| < \epsilon$, where $\epsilon > 0$ is a pre-specified small positive value as threshold for convergence, output $\boldsymbol{w}^{(t)}$ as the optimal weight. Otherwise, go back to (a).

(d) Reduce a to $a/2$. Repeat (a) and (b) until reaching a pre-specified small value, say 0.00001. If there is a non-positive component in weight, remove the support point with the smallest weight. Go back to (a).

Theorem 6. *If $\boldsymbol{M}_{\boldsymbol{\xi}_{S^{(0)}}}$ is nonsingular, the sequence of designs $\{\boldsymbol{\xi}_{\boldsymbol{S}^{(t)}} : \forall t \geq 0\}$ converges to the design $\boldsymbol{\xi}^*$ that minimizes $\Phi_p(\boldsymbol{\xi})$.*

5. Examples

5.1. Michaelis-Menten model

The Michaelis-Menten model is a nonlinear model that is widely used in the biological sciences. The model can be written in the form of (2.5) with

$$f(x_{ij}, \boldsymbol{\theta}) = \frac{\theta_1 x_{ij}}{\theta_2 + x_{ij}}, \ \boldsymbol{\theta} = (\theta_1, \theta_2).$$

So we have

$$\frac{\partial \boldsymbol{F(X_i, \theta)}}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \dfrac{x_{i1}}{\theta_2 + x_{i1}} & \dfrac{x_{i2}}{\theta_2 + x_{i2}} & \cdots & \dfrac{x_{ik}}{\theta_2 + x_{ik}} \\ -\dfrac{\theta_1 x_{i1}}{(\theta_2 + x_{i1})^2} & -\dfrac{\theta_1 x_{i2}}{(\theta_2 + x_{i2})^2} & \cdots & -\dfrac{\theta_1 x_{ik}}{(\theta_2 + x_{ik})^2} \end{pmatrix}^T.$$

Under an approximate design $\boldsymbol{\delta}$ with identical within-group $\{(x_i, w_i), i = 1, \ldots, m\}$, the information matrix can be written as

$$\begin{aligned} \boldsymbol{M(\delta)} =& c_1 \sum_{j=1}^{m} w_j \begin{pmatrix} \dfrac{x_j^2}{(\theta_2 + x_j)^2} & -\dfrac{\theta_1 x_j^2}{(\theta_2 + x_j)^3} \\ -\dfrac{\theta_1 x_j^2}{(\theta_2 + x_j)^3} & \dfrac{\theta_1^2 x_j^2}{(\theta_2 + x_j)^4} \end{pmatrix} \\ & - c_2 \sum_{j=1}^{m} w_j \begin{pmatrix} \dfrac{x_j}{\theta_2 + x_j} \\ -\dfrac{\theta_1 x_j}{(\theta_2 + x_j)^2} \end{pmatrix} \sum_j w_j \left(\frac{x_j}{\theta_2 + x_j}, -\frac{\theta_1 x_j}{(\theta_2 + x_j)^2} \right). \end{aligned}$$

Let $P(\theta) = \begin{pmatrix} 1 & 0 \\ 1 & \theta_2/\theta_1 \end{pmatrix}^{-1}$. Then we have

$$\begin{aligned} P(\boldsymbol{\theta})^{-1} \boldsymbol{M(\delta)} \left\{ P(\boldsymbol{\theta})^T \right\}^{-1} =& c_1 \sum_{j=1}^{m} w_j \begin{pmatrix} C_j^2 & C_j^3 \\ C_j^3 & C_j^4 \end{pmatrix} \\ & - c_2 \left\{ \sum w_j \begin{pmatrix} C_j \\ C_j^2 \end{pmatrix} \right\} \left\{ \sum w_j \begin{pmatrix} C_j \\ C_j^2 \end{pmatrix} \right\}^T, \end{aligned}$$

where $C_j = x_j/(\theta_2 + x_j)$. Let $\phi_1(C) = C$, $\phi_2(C) = C^2$, $\phi_3(C) = C^3$, and $\phi_4(C) = C^4$. Applying Theorem 2 with $n = 4$, we find $\Gamma(C) = 24 > 0$. Thus we can focus on the class of within-group designs with at most three support points, including upper and lower bounds of C_j.

With the design space $\chi = [0, 3]$, Table 1 lists different optimal designs for

Table 1. Optimal designs for Michaelis-Menten model.

Block size $k = 3$						
	D-optimal			A-optimal		
(θ_1, θ_2)	ρ	(x_i, w_i)		ρ	(x_i, w_i)	
(5,6)	0.4	(3,	0.5)	0.5	(3,	0.3456)
		(1.199,	0.5)		(1.1884,	0.6544)
	0.5	(0,	0.1111)	0.6	(0,	0.0664)
		(3,	0.4444)		(3,	0.3242)
		(1.2003,	0.4444)		(1.1998,	0.6094)
(1,2)	0.4	(3,	0.5)	0.5	(3,	0.3411)
		(0.8576,	0.5)		(0.8529,	0.6589)
	0.5	(0,	0.1111)	0.6	(3,	0.3158)
		(3,	0.4444)		(0,	0.0763)
		(0.8576,	0.4444)		(0.857,	0.608)
Block size $k = 10$						
(5,6)	0.1	(3,	0.5)	0.2	(3,	0.3417)
		(1.2009,	0.5)		(1.1624,	0.6583)
	0.2	(0,	0.0667)	0.3	(3,	0.3272)
		(3,	0.4667)		(0,	0.0579)
		(1.199,	0.4667)		(1.1998,	0.6149)

different configurations of the correlation coefficient ρ, pre-specified θ, block size k and the optimality criterion (A or D). All numeric solutions are based on 30,000 grids on $\chi = [0, 3]$, and all values of support points or weights are rounded to multiples of 0.0001. Table 1 reveals a few interesting patterns. First, boundary points may not always be support points. Theorem 2 (c) indicates that at most three support points are necessary, while in Table 1, some optimal designs only require two support points. By Theorem 2, lower and upper bound of C_j should be in the support set when there are three supporting points. When an optimal design has only two support points, it may not necessarily include both upper and lower bound of C_j. From Table 1, we can see only upper bound is included when optimal design has only two support points.

The number of support points tend to increase when ρ or block size k increase. This is not surprising. From (2.4), $\boldsymbol{I_\xi}$ is proportional to

$$\int g(x)g(x)^T \boldsymbol{\xi_i}(dx) - \frac{c_2}{c_1}\left\{\int g(x)\boldsymbol{\xi_i}(dx)\right\}\left\{\int g(x)\boldsymbol{\xi_i}(dx)\right\}^T. \tag{5.1}$$

When there is no correlation within a block, the information matrix is the first part of (5.1) and optimal design are based on two support points (Example 14.6, Biedermann and Biedermann and Yang (2015)). When c_2/c_1 is small, optimal designs mainly depend on the first part. As it increases, the second part becomes more dominant. On the other hand,

$$\frac{c_2}{c_1} = \frac{k\rho}{1 + (k-1)\rho}$$

is a increasing function of ρ and k.

The saturated D-optimal design always has equal weights. This phenomena has been well known in the independent observation case. The numerical results shows it also holds for the correlated data. Now we confirm this by Theorem 7. Cheng (1995) showed a similar result for linear models with the same correlation structure under the setup of exact design. Here we show it is also true for nonlinear model under approximate design.

Theorem 7. *For any model in the form of* (2.5)*, when* $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$*, if* $\boldsymbol{\xi} = \{(x_j, w_j)\}$ *supported on* m *points is* D*-optimal design when* $\rho = 0$*, then* $\boldsymbol{\xi}$ *is also the saturated* D*-optimal design when* $\rho \neq 0$*. Furthermore,* $\boldsymbol{\xi}$ *has equal weights at all support points.*

Proof. In $\boldsymbol{M_\xi} = \sigma^{-2}\boldsymbol{F}^T\boldsymbol{V}^{-1}\boldsymbol{F}$, $\boldsymbol{F}$ is always square matrix when $\boldsymbol{\xi}$ is saturated design, thus we have

$$\begin{aligned} |\boldsymbol{M_\xi}| &= \sigma^{-2}|\boldsymbol{F}^T||\boldsymbol{V}|^{-1}|\boldsymbol{F}| \\ &= \sigma^{-2}|\boldsymbol{V}|^{-1} \times |\boldsymbol{F}^T\boldsymbol{F}| \\ &= \sigma^{-1}\left(1 + \frac{k\rho}{1-\rho}\right)^{-1}\left(\frac{k}{1-\rho}\right)^m \prod w_j \times |\boldsymbol{F^T}\boldsymbol{F}| \end{aligned} \tag{5.2}$$

By (5.2), it is obvious that for any $\{w_j | j = 1, \ldots, m\}$, $|\boldsymbol{M_\xi}|$ is maximized when $|\boldsymbol{F}^T\boldsymbol{F}|$ – the D-optimal function for the uncorrletated model – is maximized, for all $\rho \in [0, 1]$, and for any fixed $\boldsymbol{F}$, $|\boldsymbol{M_\xi}|$ achieves its maximum when $w_1 = \cdots = w_m = 1/m$.

Remark 3. Since the D-optimal design for the Michaelis-Menten model with independent errors is based on two points, the proof of Theorem 7 also shows that a two-points D-optimal design when $\rho \neq 0$ must be the D-optimal design for the independent case. Since the block size k is irrelevant to optimal design when observations are independent, it is not surprising, for $(\theta_1, \theta_2) = (5, 6)$, that the two D-optimal designs when $(\rho, k) = (0.4, 3)$ and $(\rho, k) = (0.1, 10)$ are identical with the understanding slight differences are due to computing errors.

5.2. Exponential model

The model can be written in the form of (2.5) with

Table 2. Optimal designs for exponential model.

Block size $k = 3$				
	D-optimal		A-optimal	
(θ_1, θ_2)	ρ	(x_i, w_i)	ρ	(x_i, w_i)
(5,6)	0.6	(0, 0.5)	0.6	(0, 0.6411)
		(3, 0.5)		(3, 0.3589)
	0.9	(0, 0.5)	0.9	(0, 0.6411)
		(3, 0.5)		(3, 0.3589)
(1,2)	0.5	(3, 0.5)	0.7	(3, 0.2861)
		(0.9982, 0.5)		(1.4878, 0.7139)
	0.9	(0, 0.2634)	0.9	(0, 0.076)
		(3, 0.4391)		(3, 0.2977)
		(1.5775, 0.2975)		(1.8531, 0.6263)
Block size $k = 10$				
(5,6)	0.9	(0, 0.5)	0.9	(3, 0.3906)
		(3, 0.5)		(0.7768, 0.6094)
(1,2)	0.4	(3, 0.5)	0.6	(3, 0.3183)
		(0.9992, 0.5)		(1.7483, 0.6817)
	0.5	(0, 0.1037)	0.7	(0, 0.0423)
		(3, 0.4882)		(3, 0.3122)
		(1.1488, 0.4081)		(1.848, 0.6455)

$$f(x_{ij}, \boldsymbol{\theta}) = \theta_1 \exp\left(\frac{x_{ij}}{\theta_2}\right),$$
$$\boldsymbol{\theta} = (\theta_1, \theta_2).$$

Under the approximate design $\boldsymbol{\delta}$ with identical within-group $\{(x_i, w_i), i = 1, \ldots, m\}$, we have

$$\{P(\boldsymbol{\theta})\}^{-1} \boldsymbol{M(\delta)} \left\{P(\boldsymbol{\theta})^T\right\}^{-1} = c_1 k \sum_j w_j \begin{pmatrix} e^{2C_j} & C_j e^{2C_j} \\ C_j e^{2C_j} & C_j^2 e^{2C_j} \end{pmatrix}$$
$$- c_2 \sum_j w_j \begin{pmatrix} e^{C_j} \\ C_j e^{C_j} \end{pmatrix} \sum_j w_j \begin{pmatrix} e^{C_j} & C_j e^{C_j} \end{pmatrix},$$

where $C_j = x_j/\theta_2$ and $P(\boldsymbol{\theta}) = \begin{pmatrix} 1 & 0 \\ 0 & -(\theta_1/\theta_2) \end{pmatrix}$. Let $\phi_1(C) = e^C$, $\phi_2(C) = Ce^C$, $\phi_3(C) = e^{2C}$, $\phi_4(C) = Ce^{2C}$, and $\phi_5(C) = C^2 e^{2C}$. Applying Theorem 2 with $n = 5$, we find that $\Gamma(C) = 4e^{2C} > 0$, for any C. Thus, we can focus on within-group designs supported by at most three distinct points including the upper bound of C.

Table 2 lists the optimal designs under different configurations of the parameters when the design space $[0, 3]$ has 30,000 grids. It exhibits similar patterns as seen in Table 1.

5.3. Three parameters E_{max} model

Dette, Melas and Wong (2005) studied another version of the E_{max} model, that can be written in the form (2.5) with

$$f(x_{ij}, \boldsymbol{\theta}) = \frac{\theta_0 x_{ij}^{\theta_2}}{\theta_1 + x_{ij}^{\theta_2}},$$
$$\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2),$$

where $\theta_0, \theta_1 > 0$ and $\theta_2 \neq 0$. Under the approximate design $\boldsymbol{\delta}$ with identical within-group $\{(x_i, w_i), i = 1, \ldots, m\}$, we have

$$\begin{aligned}
&\boldsymbol{M}(\boldsymbol{\delta}^*) \\
&= c_1 P(\boldsymbol{\theta}) \sum_j w_j \begin{pmatrix} \frac{1}{(1+C_j)^2} & \frac{1}{(1+C_j)^3} & \frac{C_j \log C_j}{(1+C_j)^3} \\ \frac{1}{(1+C_j)^3} & \frac{1}{(1+C_j)^4} & \frac{C_j \log C_j}{(1+C_j)^4} \\ \frac{C_j \log C_j}{(1+C_j)^3} & \frac{C_j \log C_j}{(1+C_j)^4} & \frac{C_j^2 \log^2 C_j}{(1+C_j)^4} \end{pmatrix} P(\boldsymbol{\theta})^T \\
&- c_2 P(\boldsymbol{\theta}) \sum_j w_j \begin{pmatrix} \frac{1}{1+C_j} \\ \frac{1}{(1+C_j)^2} \\ \frac{C_j \log C_j}{(1+C_j)^2} \end{pmatrix} \sum_j w_j \begin{pmatrix} \frac{1}{1+C_j} & \frac{1}{(1+C_j)^2} & \frac{C_j \log C_j}{(1+C_j)^2} \end{pmatrix} P(\boldsymbol{\theta})^T,
\end{aligned}$$

where $C_j = \theta_1 x_j^{-\theta_2}$ and

$$P(\boldsymbol{\theta}) = \begin{pmatrix} 1 & 0 & 0 \\ \frac{-\theta_0}{\theta_1} & \frac{\theta_0}{\theta_1} & 0 \\ \frac{\theta_0}{\theta_2} \log \theta_1 & -\frac{\theta_0}{\theta_2} \log \theta_1 & \frac{-\theta_0}{\theta_2} \end{pmatrix},$$

Let $\phi_1(C) = 1/(1+C)^4$, $\phi_2(C) = 1/(1+C)^3$, $\phi_3(C) = (C \log C)/(1+C)^4$, $\phi_4(C) = 1/(1+C)^2$, $\phi_5(C) = (C \log C)/(1+C)^3$, $\phi_6(C) = 1/(1+C)$, $\phi_7(C) = (C^2 \log^2 C)/(1+C)^4$, and $\phi_8(C) = (C \log C)/(1+C)^2$. Applying Theorem 2 with $n = 8$, we can verify that $\Gamma(C) = (-12)/\{C^5(1+C)^4\} < 0$, for any $C > 0$, which is satisfied when the design space is $\mathcal{X} = [1, 4]$.

Table 3 lists the optimal designs under different configurations of the param-

Table 3. Optimal designs for E_{max} model.

Block size $k=3$						
	D-optimal			A-optimal		
$(\theta_0,\theta_1,\theta_2)$	ρ	$(x_i,$	$w_i)$	ρ	$(x_i,$	$w_i)$
Block size $k=3$						
(1,2,3)	0.5	(1,	0.3333)	0.5	(1,	0.409)
		(4,	0.3333)		(4,	0.2215)
		(1.757,	0.3333)		(1.759,	0.3695)
	0.9	(1,	0.3333)	0.9	(4,	0.2301)
		(4,	0.3333)		(1.045,	0.3902)
		(1.755,	0.3333)		(1.846,	0.3797)
(1,5,6)	0.8	(1,	0.1241)	0.9	(4,	0.1642)
		(4,	0.325)		(1,	0.0217)
		(1.563,	0.3182)		(1.205,	0.4472)
		(1.143,	0.2327)		(1.629,	0.3669)
Block size $k=10$						
(1,2,3)	0.8	(1,	0.3333)	0.7	(4,	0.2325)
		(4,	0.3333)		(1.1035,	0.3743)
		(1.755,	0.3333)		(1.9145,	0.3932)
	0.9	(1,	0.3112)	0.8	(4,	0.234)
		(4,	0.3256)		(1,	0.0084)
		(1.34,	0.0647)		(1.167,	0.3527)
		(1.819,	0.2984)		(1.987,	0.405)
Block size $k=3$						
(1,5,6)	0.5	(1,	0.1822)	0.6	(4,	0.1748)
		(4,	0.3139)		(1,	0.0067)
		(1.188,	0.2049)		(1.716,	0.3774)
		(1.5845,	0.2989)		(1.2185,	0.4411)

eters when design space $[1,4]$ has 30,000 grids. It exhibits similar patterns as seen in Table 1.

6. Robustness of Locally Optimal Designs

The optimal designs discussed in previous sections are computed based on the pre-specified value of ρ, which is usually unknown in practice. We study design efficiency with a wrongly specified ρ. The D- and A-efficiencies of a given design, say ξ, are defined as

$$\begin{aligned} \text{D-eff}(\xi) &= \frac{\Phi_0(\xi_D^*)}{\Psi_0(\xi)}, \\ \text{A-eff}(\xi) &= \frac{\Phi_1(\xi_A^*)}{\Psi_1(\xi)}, \end{aligned} \tag{6.1}$$

Table 4. D-efficiency and A-efficiency of optimal designs with wrong ρ.

true ρ	D-efficiency	A-efficiency
Design:	(3, 0.5) (1.2,0.5)	(3, 0.3081) (0.96,0.6919)
0	1.0000	1.0000
0.1	1.0000	0.9981
0.4	1.0000	0.9643
0.45	0.9864	0.9526
0.5	0.9492	0.9382
0.6	0.8216	0.8852
0.7	0.6460	0.7735
0.8	0.4424	0.5973
0.9	0.2242	0.3460

All designes have prespecified $\rho = 0$ and truely specified $\boldsymbol{\theta} = (5, 6)$.

Table 5. D-efficiency and A-efficiency of optimal desings with wrong $\boldsymbol{\theta}$.

true $\boldsymbol{\theta}$	D-efficiency	A-efficiency
Design:	(3, 0.4444) (1.2,0.4444) (0, 0.1111)	(3, 0.3458) (1.19,0.6542)
(5,3)	0.9596	0.9292
(5,6)	1.0000	1.0000
(5,16)	0.9731	0.9609
(5,60)	0.9386	0.9154

All designs have pre-specified $\boldsymbol{\theta} = (5, 6)$, and truely specified $\rho = 0.5$.

where ξ_D^* and ξ_A^* are D- and A- optimal design with true values of parameters, respectively.

We generated a design when ρ is specified as 0 and measured its efficiencies when the true value of ρ takes other values as in Table 4. When the true value of ρ, and hence c_2/c_1, is small, the optimal design is identical to that for $\rho = c_2/c_1 = 0$. When the true $\rho \leq 0.5$, the corresponding design efficiencies are close to or higher than 95%. When ρ is further away from the wrongly misspecified value 0, design efficiencies decrease. This is also true when the ρ are wrongly specified to values other than 0.

Locally optimal designs also depend on pre-specified $\boldsymbol{\theta}$. Under the D- optimal criteria, saturated optimal design always have 100% efficiency. Under the D- and A- optimal criteria, design efficiency decreases as $\boldsymbol{\theta}$ diverges from its true value.

Table 5 provides the efficiencies of a design when of $\rho = 0.5$ and $\boldsymbol{\theta} = (5, 6)$, at various true values of $\boldsymbol{\theta}$. Even if $\boldsymbol{\theta}$ is far away from the pre-specified value, design efficiencies are mostly higher than 90%.

Tables 4 and 4 indicate that optimal designs are quite robust with respect to misspecified values of $\boldsymbol{\theta}$ and ρ under Michaelis-Menten model. Simulations with other examples yields similar conclusions. They are omitted here due to space limit.

7. Discussion

Although nonlinear models with correlated responses are not uncommon in practice, little optimality work has been done. The main challenge is that the information matrix does not have the "additive" property, where one applies powerful tools.

For the nonlinear models with random block effects, the variance-covariance matrix for the observations within a block is compound symmetric. We are then able to characterize the format of optimal designs and derive the corresponding general equivalence theorem. Unlike nonlinear models with independent observations, in which optimal designs are often based on saturated designs, for optimal designs under nonlinear models with random block effects, this is no longer the case. The number of support points depends on how strong the correlation ρ is. When ρ is close to 0, it is often equal to minimum number of support points, just as in the independent case. When ρ is close to 1, optimal designs often have one more support points than the saturated designs.

For nonlinear models with other correlation structures, the information matrix becomes more complicated. The method employed here is unlikely to work. Specifically, it is not clear whether the general equivalence theorem still holds. Given the importance of nonlinear models with correlated responses, more research in this direction is certainly needed.

Supplementary Materials

Proofs of Theorem 4, Theorem 5, and 6 can be found in the on-line supplement material.

Acknowledgment

The authors are thankful for detailed comments and suggestions by the Editor, an associate editor, and two referees on an earlier version of the article, which clearly helped to improve the final version. Yang's research was supported by NSF grant DMS-1407518 and DMS-1811291 and Zheng's research is supported by NSF grant DMS-1612978.

References

Atkins, J. E. and Cheng, C.-S. (1999). Optimal regression designs in the presence of random block effects. *Journal of Statistical Planning and Inference* **77**, 321–335.

Atkinson, A. C. (2008). Examples of the use of an equivalence theorem in constructing optimum experimental designs for random-effects nonlinear regression models. *Journal of Statistical Planning and Inference* **138**, 2595–2606.

Biedermann, S. and Yang, M. (2015). Designs for selected non-linear models. *In Handbooks on Modern Statistical Methods, Design of Experiments* (Edited by A. Dean, M. Max, J. Stufken and D. Bingham).

Cheng, C.-S. (1995). Optimal regression designs under random block-effects models. *Statistica Sinica* **5**, 485–497.

Cutler, D. R. (1993). Efficient block designs for comparing test treatments to a control when the errors are correlated. *Journal of Statistical Planning and Inference* **36**, 107–125.

Dette, H., Pepelyshev, A., Holland-Letz, T. and et al. (2010). Optimal designs for random effect models with correlated errors with applications in population pharmacokinetics. *The Annals of Applied Statistics* **4**, 1430–1450.

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

Dette, H. and Kunert, J. (2014). Optimal designs for the michaelis–menten model with correlated observations. *Statistics* **48**, 1254–1267.

Dette, H. and Melas, V. B. (2011). A note on the de la garza phenomenon for locally optimal designs. *The Annals of Statistics* **39**, 1266–1281.

Dette, H. and Schorning, K. (2013). Complete classes of designs for nonlinear regression models and principal representations of moment spaces. *The Annals of Statistics* **41**, 1260–1267.

Fedorov, V. V. (1972). *Theory of Optimal Experiments*. Elsevier.

Fedorov, V. V. and Hackl, P. (1997). *Model-Oriented Design of Experiments*, volume 125. Springer, New York.

Holland-Letz, T., Dette, H. and Renard, D. (2012). Efficient algorithms for optimal designs with correlated observations in pharmacokinetics and dose-finding studies. *Biometrics* **68**, 138–145.

Huang, S.-H. and Cheng, C.-S. (2016). Optimal designs for quadratic regression with random block effects: The case of block size two. *Journal of Statistical Planning and Inference* **175**, 67–77.

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

Kiefer, J. and Wynn, H. (1981). Optimum balanced block and latin square designs for correlated observations. *The Annals of Statistics* **9**, 737–757.

Kunert, J., Martin, R. and Eccleston, J. (2010). A-optimal block designs for the comparison with a control for correlated errors and analysis with the weighted least squares estimate. *Journal of Statistical Planning and Inference* **140**, 2719–2738.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall, London.

McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear and Mixed Models*. John Wiley & Sons, New York.

Müller, W. G. and Pázman, A. (1999). An algorithm for the computation of optimum designs under a given covariance structure. *Computational Statistics* **14**, 197–211.

Pázman, A. (2010). Information contained in design points of experiments with correlated observations. *Kybernetika* **46**, 771–783.

Schmelter, T. (2007). The optimality of single-group designs for certain mixed models. *Metrika* **65**, 183–193.

Stufken, J. and Yang, M. (2012). On locally optimal designs for generalized linear models with group effects. *Statistica Sinica* **22**, 1765–1786.

Ucinski, D. and Atkinson, A. C. (2004). Experimental design for time-dependent models with correlated observations. *Studies in Nonlinear Dynamics & Econometrics* **8**, 13.

Wynn, H. (1970). The sequential generation of d-optimum experimental designs. *The Annals of Mathematical Statistics* **41**, 1655–1664.

Yang, M. (2010). On the de la Garza phenomenon. *The Annals of Statistics* **38**, 2499–2524.

Yang, M., Biedermann, S. and Tang, E. (2013). On optimal designs for nonlinear models: a general and efficient algorithm. *Journal of the American Statistical Association* **108**, 1411–1420.

Yang, M. and Stufken, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *The Annals of Statistics* **37**, 518–541.

Yang, M. and Stufken, J. (2012). Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences. *The Annals of Statistics* **40**, 1665–1681.

Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL 60607, USA.

E-mail: xwang202@uic.edu

Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL 60607, USA.

E-mail: myang2@uic.edu

Department of Business Analytics and Statistics, University of Tennessee, Indianapolis, IN 46202, USA.

E-mail: wzheng9@utk.edu

(Received January 2017; accepted July 2017)