Statistica Sinica **32** (2022), 2047-2069
doi:https://doi.org/10.5705/ss.202020.0278

A MAXIMIN Φ_P-EFFICIENT DESIGN FOR MULTIVARIATE GENERALIZED LINEAR MODELS

Yiou Li, Lulu Kang and Xinwei Deng

DePaul University, Illinois Institute of Technology and Virginia Tech

Abstract: Experimental designs for generalized linear models often depend on the specification of the model, including the link function, predictors, and unknown parameters, such as the regression coefficients. To deal with the uncertainties of these model specifications, it is important to construct optimal designs with high efficiency under such uncertainties. Existing methods, such as Bayesian experimental designs, often use prior distributions of model specifications to incorporate model uncertainties into the design criterion. Alternatively, one can obtain the design by optimizing the worst-case design efficiency with respect to the uncertainties of the model specifications. In this work, we propose a new Maximin Φ_p-Efficient (or Mm-Φ_p for short) design that aims to maximize the minimum Φ_p-efficiency under model uncertainties. Based on the theoretical properties of the proposed criterion, we develop an efficient algorithm with sound convergence properties to construct the Mm-Φ_p design. The performance of the proposed Mm-Φ_p design is assessed using several numerical examples.

Key words and phrases: Φ_p-criterion, dsesign efficiency, efficient algorithm, model uncertainty, optimal design.

1. Introduction

Optimal design for generalized linear models (GLMs) (Khuri et al. (2006); Fedorov and Leonov (2013)) is an important topic in the design of experiments. Here, recent theoretical and algorithmic developments include the works of Woods and Lewis (2011), Yang, Zhang and Huang (2011), Burghaus and Dette (2014), Wu and Stufken (2014), and Wong, Yin and Zhou (2019), among many others. A key challenge of the optimal design for a GLM is that the design criterion often depends on the regression model assumption, including the specification of the link function, the linear predictor, and the values of the unknown regression coefficients. Many existing works focus on locally optimal designs, given a certain model specification, as in Yang and Stufken (2009), Li and Majumdar (2009),

Corresponding author: Xinwei Deng, Department of Statistics, Virginia Tech, Blacksburg, VA 24061-0131, USA. E-mail: xdeng@vt.edu.

Wu and Stufken (2014), and Li and Deng (2020). In contrast to the locally optimal design, one type of global optimal design considers the parameter uncertainty under two directions. One direction is to consider a prior distribution of the unknown parameters, when constructing the so-called Bayesian optimal design (Khuri et al. (2006); Woods et al. (2017)). The design criterion is typically the integral of the local design criterion or efficiency with respect to the prior of the parameters. When such integration is not analytically available, a standard solution is to sample from the prior distribution, and to use the weighted average of local design criteria or efficiencies as the objective function (Atkinson and Woods (2015)). Another direction is to use the minimax/maximin approach to minimize the design criterion or to maximize the efficiency under the "worst-case" scenario. Sitter (1992) introduced a minimax procedure for obtaining a design to deal with parameter uncertainty. King and Wong (2000) proposed an efficient algorithm to construct a maximin design for the logistic regression model under D-optimality. Imhof and Wong (2000) developed an algorithm to maximize the minimum efficiency under two competing optimality criteria using a graphical method. Note that existing studies on maximin/minimax designs often focus on D-optimality and the uncertainty of the unknown parameters. The biggest challenge in maximin/minimax designs is that the design construction can be quite difficult (Atkinson and Woods (2015)).

In addition to the unknown parameters, there could be other uncertainties involved in a GLM, such as the specification of the link function and the linear predictor. However, the literature on GLM designs that deal with such model uncertainty is relatively scarce. Woods et al. (2006) proposed a compromise design that minimizes the weighted average of the criteria, and each criterion is based on a potential model. Later, Dror and Steinberg (2006) proposed using clustered locally optimal designs, and showed the resulting design had comparable performance with the compromise design through numerical examples.

In this work, we propose a new maximin Φ_p-efficient design (denoted as Mm-Φ_p) criterion for GLMs using the Φ_p-efficiency (Kiefer (1974)), and develop an efficient algorithm to construct the design. The proposed design, the Mm-Φ_p design, can accommodate several types of uncertainty, including (i) uncertainty over the unknown parameter values, (ii) uncertainty over the linear predictor, and (ii) uncertainty over the link function. Here, we focus on an *approximate design* (Atkinson (2014)), which describes the design as a probability measure on a group of support points. This provides a framework for us to investigate the theoretical properties of the proposed design criterion, as well as a theoretical foundation from which to develop an efficient algorithm with desirable convergence properties.

The key idea of this work is to adopt a continuous and convex relaxation (i.e., the "log-sum-exp" approximation) as a tight approximation of the worst-case Φ_p-efficiency with respect to the uncertainty of the model specifications. With this relaxation, we arrive at a tractable design criterion that facilitates the theoretical investigation for developing an efficient algorithm to construct the corresponding design. The merits of this idea are not restricted to the Φ_p criterion, even though Φ_p is already a quite general criterion that includes the A-, D-, E-, and I-optimality criteria as special cases. A demonstration of the proposed approach based on the Φ_p-criterion reveals that this convex and smooth relaxation idea can be applied to other maximin designs, as long as the criterion is convex in the design. The proposed framework, including the general equivalence theorem and the design construction algorithm and its convergence, can be extended to other maximin designs as well.

Other main contributions of this work are summarized as follows. First, the proposed Mm-Φ_p design criterion is very general, covering various design criteria, such as D-, A-, E-optimality for estimation accuracy and I- and EI-optimality for prediction accuracy (Li and Deng (2020)). Second, in contrast to the Bayesian optimal design, the proposed Mm-Φ_p design is a maximin design, which avoids having to choose prior distributions on the model specifications. Third, the proposed Mm-Φ_p design can flexibly accommodate the aforementioned three types of model uncertainty in a GLM. Finally, the proposed algorithm has impressive computational efficiency with sound theoretical properties, and can be easily modified to construct compromise designs and Bayesian optimal designs.

The rest of the paper is organized as follows. Section 2 describes the Mm-Φ_p design criterion and investigates its theoretical properties. In Section 3, an efficient algorithm is developed. Numerical examples are conducted in Section 4 to examine the performance of the proposed method. We summarize the work with a discussion in Section 5. All technical proofs are relegated to the Supplementary Material.

2. The Mm-Φ_p Design Criterion and Its Properties

Consider an experiment with d design variables, $\boldsymbol{x} = [x_1, \ldots, x_d]$, and $x_j \in \Omega_j$, where Ω_j is a measurable domain of all possible values for x_j. The experimental region, Ω, is a certain measurable subset of $\Omega_1 \times \cdots \times \Omega_d$. For a GLM, the response $Y(\boldsymbol{x})$ is assumed to follow a distribution in the exponential family. The link function, $h : \mathbb{R} \to \mathbb{R}$, provides the relationship between the linear predictor, $\eta = \boldsymbol{\beta}^\top \boldsymbol{g}(\boldsymbol{x})$, and the mean of the response $Y(\boldsymbol{x})$, $\mu(\boldsymbol{x})$, as $\mu(\boldsymbol{x}) = \mathbb{E}[Y(\boldsymbol{x})] = h^{-1}\left(\boldsymbol{\beta}^\top \boldsymbol{g}(\boldsymbol{x})\right)$, where $\boldsymbol{g} = [g_1, \ldots, g_l]^\top$ are the known basis functions of the design variables, $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_l]^\top$ are the corresponding regression coefficient parameters, and h^{-1} is the inverse function of h. The approximate design ξ is defined as $\xi = \begin{Bmatrix} \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \\ \lambda_1, \ldots, \lambda_n \end{Bmatrix}$, where $\boldsymbol{x}_1, \ldots \boldsymbol{x}_n$ are the support points, and $0 < \lambda_i < 1$ represents the probability mass allocated to the corresponding support point $\boldsymbol{x}_i$. We use $M = (h, \boldsymbol{g}, \boldsymbol{\beta})$ to denote the model specification of a GLM with link function h, the basis functions $\boldsymbol{g}$, and the vector of the regression coefficients $\boldsymbol{\beta}$. The Fisher information matrix of the GLM, M, is

$$\mathsf{I}(\xi; M) = \sum_{i=1}^{n} \lambda_i \boldsymbol{g}(\boldsymbol{x}_i) w(\boldsymbol{x}_i; M) \boldsymbol{g}^\top(\boldsymbol{x}_i), \tag{2.1}$$

where $w(\boldsymbol{x}_i; M) = \left[\mathrm{var}(Y(\boldsymbol{x}_i))[h'(\mu(\boldsymbol{x}_i))]^2\right]^{-1}$. Clearly, $\mathsf{I}(\xi; M)$ depends on all three components of $M = (h, \boldsymbol{g}, \boldsymbol{\beta})$. Various locally optimal design criteria in the literature are based on the Fisher information with a specified M.

2.1. The Mm-Φ_p design criterion

To represent the uncertainties of a GLM, we denote the set of candidate link functions, set of the candidate basis functions, and domain of the regression coefficients as $\mathcal{H}$, $(\mathcal{G}|\mathcal{H})$, and $(\mathcal{B}|\mathcal{H}, \mathcal{G})$, respectively. The notation of conditioning represents the dependence of the basis functions $\boldsymbol{g}$ on the choice of the link function h, and the dependence of the regression coefficients $\boldsymbol{\beta}$ on the choice of both h and $\boldsymbol{g}$. The set $\mathcal{M} = \{M = (h, \boldsymbol{g}, \boldsymbol{\beta}) : h \in \mathcal{H}, \boldsymbol{g} \in (\mathcal{G}|\mathcal{H}), \boldsymbol{\beta} \in (\mathcal{B}|\mathcal{H}, \mathcal{G})\}$ contains all model specifications of interest.

In optimal design theory, *efficiency* is a popular and scale-free performance measurement used to compare designs for a given criterion. Specifically, for a generic design criterion $\Psi(\xi; \mathcal{M})$, which is to be minimized, the efficiency of a design ξ relative to another design ξ' is defined as (Atkinson, Donev and Tobias (2007))

$$\mathrm{eff}_\Psi(\xi, \xi'; \mathcal{M}) = \frac{\Psi(\xi'; \mathcal{M})}{\Psi(\xi; \mathcal{M})}. \tag{2.2}$$

Using this definition of efficiency, the design ξ is more efficient than the design ξ' as long as the efficiency in (2.2) is larger than one. When a single model specification is considered, that is, $\mathcal{M} = \{M\}$, the criterion Ψ becomes a locally optimal design criterion. When multiple specifications are considered, the criterion Ψ corresponds to a global optimal design criterion, such as the Bayesian optimality, compromise design optimality, minimax/maximin optimality, and so on.

Throughout this work, for a specified model M, we use the generalized Φ_p-optimality introduced in Kiefer (1974), which is

$$\Phi_p(\xi; M) = \left(q^{-1} \operatorname{tr} \left[\frac{\partial \boldsymbol{f}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top} \mathsf{I}(\xi; M)^{-1} \left(\frac{\partial \boldsymbol{f}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top} \right)^\top \right]^p \right)^{1/p}, \quad 0 < p < \infty, \tag{2.3}$$

where $\boldsymbol{f}(\boldsymbol{\beta}) = [f_1(\boldsymbol{\beta}), \ldots, f_q(\boldsymbol{\beta})]^\top$ are some functions of $\boldsymbol{\beta}$. Common examples are linear contrasts of the coefficients, such as β_k and $\beta_j - \beta_{j'}$. Note that the Φ_p-optimality is essentially D-optimality as $p \to 0$ and E-optimality as $p \to \infty$. We denote ξ_M^{opt} to be the locally optimal design that minimizes the Φ_p-criterion for the model M. According to (2.2), the Φ_p-efficiency of any design ξ relative to a locally optimal design ξ_M^{opt}, given a specific $M = (h, \boldsymbol{g}, \boldsymbol{\beta})$, is

$$\operatorname{eff}_{\Phi_p}(\xi, \xi_M^{\text{opt}}; M) = \frac{\Phi_p(\xi_M^{\text{opt}}; M)}{\Phi_p(\xi; M)}. \tag{2.4}$$

It is obvious that $0 \leq \operatorname{eff}_{\Phi_p}(\xi, \xi_M^{\text{opt}}; M) \leq 1$ for any ξ, and a larger Φ_p-efficiency represents a more efficient design ξ. Under the idea of a global maximin design, we consider the maximin Φ_p-efficient design, which maximizes the smallest possible $\operatorname{eff}_{\Phi_p}(\xi, \xi_M^{\text{opt}}; M)$ over all $M \in \mathcal{M}$. That is, we consider the following maximin design:

$$\xi^* = \operatorname{argmax}_\xi \inf_{M \in \mathcal{M}} \operatorname{eff}_{\Phi_p}(\xi, \xi_M^{\text{opt}}; M). \tag{2.5}$$

In the optimization problem (2.5), the infimum is used instead of the minimum, because it is not certain whether the minimum is attainable. To simplify the problem, we take a closer look at the model set $\mathcal{M}$. In practice, $\mathcal{H}$ usually contains several candidate link functions. For example, the link function of a GLM for binary data could be the logistic function $h(\mu(\boldsymbol{x})) = \ln(\mu(\boldsymbol{x})/(1 - \mu(\boldsymbol{x})))$, probit function $h(\mu(\boldsymbol{x})) = \Phi^{-1}(\mu(\boldsymbol{x}))$, or complementary log-log function $h(\mu(\boldsymbol{x})) = \ln(-\ln(1 - \mu(\boldsymbol{x})))$. The link function of a GLM for counting data could be the log function $h(\mu(\boldsymbol{x})) = \ln(\mu(\boldsymbol{x}))$ or the power function $h(\mu(\boldsymbol{x})) = (\mu(\boldsymbol{x}))^\alpha$, with a proper choice of α. The set of candidate basis functions $(\mathcal{G}|\mathcal{H})$ is often also finite. The typical basis functions used in GLMs are linear and/or higher-order polynomials of $\boldsymbol{x}$. Note that $(\mathcal{B}|\mathcal{H}, \mathcal{G})$, the domain of $\boldsymbol{\beta}$, is often uncountable when $\boldsymbol{\beta}$ is considered to be continuous. Consequently, the set $\mathcal{M}$ is an uncountable set, which may not ensure an attainable minimum. A common remedy (Dror and Steinberg (2006); Woods et al. (2006); Atkinson and Woods (2015); Woods et al.

(2017)) is to discretize $(\mathcal{B}|\mathcal{H},\mathcal{G})$ and create a finite subset $(\mathcal{B}'|\mathcal{H},\mathcal{G})$. The corresponding surrogate set $\mathcal{M}' = \{M = (h, \boldsymbol{g}, \boldsymbol{\beta}) : h \in \mathcal{H}, \boldsymbol{g} \in (\mathcal{G}|\mathcal{H}), \boldsymbol{\beta} \in (\mathcal{B}'|\mathcal{H},\mathcal{G})\}$ is also a subset of the original $\mathcal{M}$. Replacing $\mathcal{M}$ by $\mathcal{M}'$ in (2.5), the solution of

$$\xi^* = \mathrm{argmax}_\xi \min_{M\in\mathcal{M}'} \left[\mathrm{eff}_{\Phi_p}(\xi, \xi_M^{\mathrm{opt}}; M)\right] \tag{2.6}$$

is a sub-optimal solution of (2.5). When the discretization is adequate to form a close approximation of $\mathcal{M}$, the sub-optimal solution is expected to be close to the original optimal solution.

The design criterion in (2.6) is still a challenging optimization, owing to the nonsmooth objective function $\min_{M\in\mathcal{M}'}[\mathrm{eff}_{\Phi_p}(\xi, \xi_M^{\mathrm{opt}}; M)]$ (Wong (1992); King and Wong (1998); Atkinson and Woods (2015)). We use "Log-Sum-Exp" as a tight and smooth approximation to the minimum function, which is widely used in machine learning (Calafiore and El Ghaoui (2014)). With "Log-Sum-Exp," one has

$$\begin{aligned}\left[\ln\left(\sum_{j=1}^m \exp\left(\frac{1}{\mathrm{eff}_{\Phi_p}(\xi, \xi_{M_j}^{\mathrm{opt}}; M_j)}\right)\right)\right]^{-1} &\le \min_{M\in\mathcal{M}'} \mathrm{eff}_{\Phi_p}(\xi, \xi_M^{\mathrm{opt}}; M) \\ \le \left[\ln\left(\sum_{j=1}^m \exp\left(\frac{1}{\mathrm{eff}_{\Phi_p}(\xi, \xi_{M_j}^{\mathrm{opt}}; M_j)}\right)\right) - \ln(m)\right]^{-1}&, \end{aligned} \tag{2.7}$$

where m is the cardinality of $\mathcal{M}'$, that is, the number of potential model specifications in $\mathcal{M}'$. The equality in the first inequality is obtained when $m = 1$, and the equality in the second inequality holds when $\mathrm{eff}_{\Phi_p}(\xi, \xi_{M_j}^{\mathrm{opt}}; M_j)$ remains the same for all $M_j \in \mathcal{M}'$. Thus, maximizing $[\ln(\sum_{j=1}^m \exp(1/\mathrm{eff}_{\Phi_p}(\xi, \xi_{M_j}^{\mathrm{opt}}; M_j)))]^{-1}$ leads to maximizing both the lower and the upper bound of the worst (or the smallest) Φ_p-efficiency. Therefore, instead of solving (2.6), which involves an inner minimization of Φ_p-efficiency, we propose using the "Log-Sum-Exp" approximation of the worst-case Φ_p-efficiency as the design criterion, thus minimizing

$$\mathrm{LEA}(\xi; \mathcal{M}') \triangleq \ln\left(\sum_{j=1}^m \exp\left(\frac{1}{\mathrm{eff}_{\Phi_p}(\xi, \xi_{M_j}^{\mathrm{opt}}; M_j)}\right)\right). \tag{2.8}$$

Minimizing $\mathrm{LEA}(\xi; \mathcal{M}')$ is the same as maximizing $[\ln(\sum_{j=1}^m \exp(1/\,\mathrm{eff}_{\Phi_p}(\xi, \xi_{M_j}^{\mathrm{opt}}; M_j)))]^{-1}$ because $\ln(\sum_{j=1}^m \exp(1/\mathrm{eff}_{\Phi_p}(\xi, \xi_{M_j}^{\mathrm{opt}}; M_j))) > 0$. We call $\mathrm{LEA}(\xi; \mathcal{M}')$, which aims at maximizing the minimal Φ_p-efficiency, the Mm-Φ_p criterion. The design that minimizes $\mathrm{LEA}(\xi; \mathcal{M}')$ is called the Mm-Φ_p design for the surrogate model set $\mathcal{M}'$, denoted by $\xi^{\mathrm{Mm}}_{\mathcal{M}'}$.

It is obvious that minimizing $\mathrm{LEA}(\xi;\mathcal{M}')$ is equivalent to minimizing

$$\mathrm{EA}(\xi;\mathcal{M}') \triangleq \sum_{j=1}^{m} \exp\left(\frac{1}{\mathrm{eff}_{\Phi_p}(\xi,\xi^{\mathrm{opt}}_{M_j};M_j)}\right) = \sum_{j=1}^{m} \exp\left(\frac{\Phi_p(\xi;M_j)}{\Phi_p(\xi^{\mathrm{opt}}_{M_j};M_j)}\right). \quad (2.9)$$

That is, $\xi^{\mathrm{Mm}}_{\mathcal{M}'} = \mathrm{argmin}_{\xi}\,\mathrm{EA}(\xi;\mathcal{M}') = \mathrm{argmin}_{\xi}\,\mathrm{LEA}(\xi;\mathcal{M}')$.

In Section 2.2, we show the convexity of $\mathrm{EA}(\xi;\mathcal{M}')$ with respect to ξ, as well as the necessary and sufficient conditions of the Mm-Φ_p design $\xi^{\mathrm{Mm}}_{\mathcal{M}'}$.

2.2. General equivalence theorem

To develop an efficient algorithm to construct the Mm-Φ_p design, we study the convexity of the objective function $\mathrm{EA}(\xi;\mathcal{M}')$ with respect to ξ, and summarize the necessary and sufficient conditions of the Mm-Φ_p design $\xi^{\mathrm{Mm}}_{\mathcal{M}'}$ in a General Equivalence Theorem. To keep this section concise, we present the major results here and place the lemmas and proofs in the Supplementary Material S1.

For a model specification $M_j \in \mathcal{M}'$, we simplify the notation of the information matrix $\mathsf{I}(\xi;M_j)$ to $\mathsf{I}_j(\xi)$, the weight function $w(\boldsymbol{x};M_j)$ in (2.1) to $w_j(\boldsymbol{x})$, the Φ_p-criterion value of a design $\Phi_p(\xi;M_j)$ to $\Phi_p^j(\xi)$, and the Φ_p-criterion value $\Phi_p(\xi^{\mathrm{opt}}_{M_j};M_j)$ of the locally optimal design to $\Phi_p^{\mathrm{opt}_j}$. Then, we can rewrite $\mathrm{EA}(\xi;\mathcal{M}')$ as $\mathrm{EA}(\xi;\mathcal{M}') = \sum_{j=1}^{m}\exp(\Phi_p^j(\xi)/\Phi_p^{\mathrm{opt}_j})$. Lemma 1 in the Supplementary Material proves the convexity of $\mathrm{EA}(\cdot;\mathcal{M}')$ with respect to ξ. Given two designs ξ and ξ', the directional derivative of $\mathrm{EA}(\xi;\mathcal{M}')$ in the direction of ξ' is defined as follows:

$$\begin{aligned}\nabla_{\xi'}\,\mathrm{EA}(\xi;\mathcal{M}') &:= \phi(\xi',\xi)\\ &= \lim_{\alpha\to 0^+}\frac{\mathrm{EA}((1-\alpha)\xi+\alpha\xi';\mathcal{M}') - \mathrm{EA}(\xi;\mathcal{M}')}{\alpha},\ \alpha\in[0,1]. \quad (2.10)\end{aligned}$$

Lemma 2 in the Supplementary Material derives the specific formula for $\phi(\xi',\xi)$. If ξ' contains only a single support point $\boldsymbol{x}$ with corresponding weight $\lambda = 1$, the directional derivative of $\mathrm{EA}(\xi;\mathcal{M}')$ in the direction of ξ' is a special case of Lemma 2. We denote this directional derivative as $\phi(\boldsymbol{x},\xi)$, and give its formula in Lemma 3 in the Supplementary Material. Following Lemma 3, we also provide specific formulae for $\phi(\boldsymbol{x},\xi)$ for D-, A-, and EI-optimality. Using these results, we obtain General Equivalence Theorem 1 for the Mm-Φ_p design that minimizes $\mathrm{LEA}(\xi;\mathcal{M}')$ or, equivalently, minimizes $\mathrm{EA}(\xi;\mathcal{M}')$.

Theorem 1 (General Equivalence Theorem)**.** *The following two conditions of a* *design* $\xi^{\text{Mm}}_{\mathcal{M}'}$ *are equivalent:*

1. *The design* $\xi^{\text{Mm}}_{\mathcal{M}'}$ *minimizes* $\text{LEA}(\xi;\mathcal{M}')$ *and* $\text{EA}(\xi;\mathcal{M}')$.

2. $\phi(\boldsymbol{x}, \xi^{\text{Mm}}_{\mathcal{M}'}) \geq 0$ *holds for any* $\boldsymbol{x} \in \Omega$, *and the inequality becomes an equality if* $\boldsymbol{x}$ *is a support point of the design* $\xi^{\text{Mm}}_{\mathcal{M}'}$.

The General Equivalence Theorem 1 for the LEA criterion in (2.8) provides important guidelines on how the support points of the Mm-Φ_p design should be added sequentially. The proposed algorithm for the Mm-Φ_p design (detailed in Section 3) iterates between adding the support point and updating the weights λ_i, which can be considered a Fedorov–Wynn-type of algorithm (Dean et al. (2015)). In each iteration, to achieve the maximum reduction of $\text{EA}(\xi;\mathcal{M}')$, the design point $\boldsymbol{x}^* = \operatorname{argmin}_{\boldsymbol{x}} \phi(\boldsymbol{x}, \xi) < 0$ is added into the current design.

After the design point $\boldsymbol{x}^*$ is added, the weights of all design points in the current design are optimized. Thus, it is important to investigate the property of the optimal weights when the design points are given. Given design points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, the weight vector $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_n]^\top$ is the only variable for the design. We emphasize this by adding a superscript $\boldsymbol{\lambda}$ in the notation of the design, and denote it as $\xi^{\boldsymbol{\lambda}} = \left\{ \begin{matrix} \boldsymbol{x}_1, & \ldots, & \boldsymbol{x}_n \\ \lambda_1, & \ldots, & \lambda_n \end{matrix} \right\}$. Consider $\text{EA}(\xi^{\boldsymbol{\lambda}};\mathcal{M}')$ as a function of $\boldsymbol{\lambda}$; that is,

$$\text{EA}(\cdot;\mathcal{M}') : \{\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) : \lambda_i > 0, \sum \lambda_i = 1\} \mapsto \sum_{j=1}^{m} \exp\left(\frac{\Phi_p(\xi^{\boldsymbol{\lambda}}; M_j)}{\Phi_p^{\text{opt}_j}} \right). \tag{2.11}$$

The optimal weight vector $\boldsymbol{\lambda}^*$ should be the one that minimizes $\text{EA}(\xi^{\boldsymbol{\lambda}};\mathcal{M}')$ with the given support points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. Lemma 4 in the Supplementary Material proves the convexity of $\text{EA}(\xi^{\boldsymbol{\lambda}};\mathcal{M}')$ with respect to $\boldsymbol{\lambda}$. Corollary 1 provides a sufficient and necessary condition on the optimal weights for a design with fixed support points. A special case of Theorem 1 is when the experimental region is restricted to the set $\Omega = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$.

Corollary 1 (Conditions of Optimal Weights). *Given a set of design points* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, *the following two conditions on the weight vector* $\boldsymbol{\lambda}^* = [\lambda_1^*, \ldots, \lambda_n^*]^\top$ *are equivalent:*

1. *The weight vector* $\boldsymbol{\lambda}^*$ *minimizes* $\text{LEA}(\xi^{\boldsymbol{\lambda}};\mathcal{M}')$ *and* $\text{EA}(\xi^{\boldsymbol{\lambda}};\mathcal{M}')$.

2. *For all* $\boldsymbol{x}_i$, *with* $\lambda_i^* > 0$, $\phi(\boldsymbol{x}_i, \xi^{\boldsymbol{\lambda}^*}) = 0$; *for all* $\boldsymbol{x}_i$, *with* $\lambda_i^* = 0$, $\phi(\boldsymbol{x}_i, \xi^{\boldsymbol{\lambda}^*}) \geq 0$.

3. Efficient Algorithm of Constructing Mm-Φ_p Design

This section details the proposed sequential algorithm, the **Mm-Φ_p Algorithm**, used to construct the Mm-Φ_p design $\xi_{\mathcal{M}'}^{\mathrm{Mm}}$. The proposed algorithm has a sound theoretical rationale and an efficient computation. Following a similar spirit to the sequential Wynn–Fedorov-type algorithm, in each sequential iteration, a new design point $\boldsymbol{x}^*$ with the smallest negative value of directional derivative $\boldsymbol{x}^* = \mathrm{argmin}_{\boldsymbol{x}}\, \phi(\boldsymbol{x}, \xi) < 0$ is added to the current design. Then, the Optimal-Weight Procedure (detailed in Section 3.2) is used to optimize the weights of the current design points. Theoretically, the algorithm will terminate when the directional derivatives of all candidate design points in the experimental region are nonnegative. However, this stopping rule is not very practical, because it may require many iterations to make all the directional derivative values strictly positive (numerically, it is unlikely to have exactly zero cases). A common practice is to terminate the algorithm when the directional derivative $\min_{\boldsymbol{x}\in\Omega} \phi(\boldsymbol{x}, \xi) > \epsilon$, with a small negative ϵ. Alternatively, one can use the design efficiency as the stopping rule, which terminates the algorithm when the design efficiency is large enough, say close to one. In this work, we adopt the latter rule, because the design efficiency directly reflects the quality of the constructed design.

Following the general definition of design efficiency in (2.2), we denote the efficiency of a design ξ relative to the Mm-Φ_p design $\xi_{\mathcal{M}'}^{\mathrm{Mm}}$ that minimizes the Mm-Φ_p criterion LEA as

$$\mathrm{Eff}_{\mathrm{LEA}}(\xi, \xi_{\mathcal{M}'}^{\mathrm{Mm}}; \mathcal{M}') = \frac{\mathrm{LEA}(\xi_{\mathcal{M}'}^{\mathrm{Mm}}; \mathcal{M}')}{\mathrm{LEA}(\xi; \mathcal{M}')}. \tag{3.1}$$

Because $\mathrm{Eff}_{\mathrm{LEA}}(\xi, \xi_{\mathcal{M}'}^{\mathrm{Mm}}; \mathcal{M}')$ involves $\xi_{\mathcal{M}'}^{\mathrm{Mm}}$, which is unknown, we derive a lower bound for it in Theorem 2. Instead of using $\mathrm{Eff}_{\mathrm{LEA}}(\xi, \xi_{\mathcal{M}'}^{\mathrm{Mm}}; \mathcal{M}')$ as the stopping rule, we can use this lower bound.

Theorem 2 (A Lower Bound of LEA-Efficiency)**.** *Design $\xi_{\mathcal{M}'}^{\mathrm{Mm}}$ is the Mm-Φ_p design that minimizes the* LEA *criterion in* (2.8)*. The* LEA*-efficiency defined in* (3.1) *of any design ξ relative to $\xi_{\mathcal{M}'}^{\mathrm{Mm}}$ is bounded below by*

$$\mathrm{Eff}_{\mathrm{LEA}}(\xi, \xi_{\mathcal{M}'}^{\mathrm{Mm}}; \mathcal{M}') \geq 1 + 2\frac{\min_{\boldsymbol{x}\in\Omega} \phi(\boldsymbol{x}, \xi)}{\mathrm{EA}(\xi; \mathcal{M}')}.$$

Using the lower bound of LEA-efficiency in Theorem 2 as the stopping criterion, the proposed algorithm terminates when the lower bound $1+2(\min_{\boldsymbol{x}\in\Omega} \phi(\boldsymbol{x}, \xi) /\mathrm{EA}(\xi; \mathcal{M}'))$ exceeds a user-specified value, Tol_{eff}. Here, Tol_{eff} should be set close to one, say $Tol_{\mathrm{eff}} = 0.99$, or equivalently $\min_{\boldsymbol{x}\in\Omega} \phi(\boldsymbol{x}, \xi)/\mathrm{EA}(\xi; \mathcal{M}') \geq -0.005$.

Note that Tol_{eff} is chosen to be 0.99 in all the numerical examples presented here. With this stopping rule, the sequential algorithm used to construct the Mm-Φ_p design is described in Algorithm 1. Here, $MaxIter_1$ is the maximum number of iterations allowed of adding design points, and we set it to be 200. To avoid including design points with almost zero weights in the constructed design, in each iteration of Algorithm 1, one can exclude these design points (say, weight $< 10^{-15}$), and then obtain the optimal weights for the updated set of design points. The candidate pool $\mathcal{C}$ could be evenly spaced grid points or some space-filling designs. Because the directional derivatives of all points in $\mathcal{C}$ are evaluated in each iteration of Algorithm 1, the computational time relies heavily on the size of $\mathcal{C}$. When the grid points are adopted, the size of the candidate pool $\mathcal{C}$ increases dramatically as the dimension d of the design variables $\boldsymbol{x}$ increases. Thus, we suggest using the grid points when the dimension d of the design variables is small, and choosing the Sobol sequence (Sobol (1967)) as the candidate pool when the dimension d is large. The Sobol sequence is a space-filling design (Santner et al. (2003)) that covers the experimental region Ω well, and can be generated efficiently when the dimension d is large.

In Section 3.1, we provide some theoretical properties on the convergence of the Mm-Φ_p Algorithm. Note that the **Mm-Φ_p Algorithm** requires optimizing the weights $\boldsymbol{\lambda}^{(r)}$ of the current design points in each sequential iteration. Section 3.2 describes the procedure on how to optimize the weights given the design points.

3.1. Convergence of the Mm-Φ_p algorithm

The sequential nature of the proposed **Mm-Φ_p Algorithm** (i.e., Algorithm 1) makes it computationally efficient because it adds one design point in each iteration. Moreover, we can establish the theoretical convergence of Algorithm 1, which is stated as follows.

Theorem 3 (Convergence of Algorithm 1(Mm-Φ_p Algorithm)). *Assume the candidate pool $\mathcal{C}$ contains all the support points of the Mm-Φ_p design $\xi^{\text{Mm}}_{\mathcal{M}'}$. The design constructed by Algorithm 1 converges to $\xi^{\text{Mm}}_{\mathcal{M}'}$ that minimizes $\text{LEA}(\xi;\mathcal{M}')$; that is,*

$$\lim_{r\to\infty} \text{LEA}(\xi^{(r)};\mathcal{M}') = \text{LEA}(\xi^{\text{Mm}}_{\mathcal{M}'};\mathcal{M}').$$

In addition to its theoretically guaranteed convergence property, Algorithm 1 also converges fast, within about 50 iterations in all numerical examples, although the maximal number of iterations is set to be 200. More details about the speed of convergence and computational time are reported in Section 4.

Algorithm 1 (**Mm-Φ_p Algorithm**) The Sequential Algorithm for Mm-Φ_p Design.

For each model specification $M_j \in \mathcal{M}'$, construct the locally optimal design and calculate the corresponding optimality criterion value $\Phi_p^{\mathrm{opt}_j}$.
Generate a candidate pool $\mathcal{C}$ of N points using a grid or Sobol sequence from experimental region Ω.
Choose an initial design points set $\mathcal{X}^{(0)} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{l+1}\}$ containing $l+1$ points.
Obtain optimal weights $\boldsymbol{\lambda}^{(0)}$ of initial design points set $\mathcal{X}^{(0)}$ using Algorithm 2 (**Optimal-Weight Procedure**) and form the initial design $\xi^{(0)} = \begin{Bmatrix} \mathcal{X}^{(0)} \\ \boldsymbol{\lambda}^{(0)} \end{Bmatrix}$.
Calculate the lower bound of LEA-efficiency of $\xi^{(0)}$:

$$\text{eff.low} = 1 + 2\frac{\min_{\boldsymbol{x}\in\mathcal{C}} \phi(\boldsymbol{x}, \xi^{(0)})}{\mathrm{EA}(\xi^{(0)}; \mathcal{M}')}.$$

Set $r = 1$.
while eff.low $< Tol_{\mathrm{eff}}$ and $r < MaxIter_1$ **do**
    Add the point $\boldsymbol{x}_r^* = \mathrm{argmin}_{\boldsymbol{x}\in\mathcal{C}}\, \phi(\boldsymbol{x}, \xi^{(r-1)})$ to the current design points set, i.e., $\mathcal{X}^{(r)} = \mathcal{X}^{(r-1)} \cup \{\boldsymbol{x}_r^*\}$, where $\phi(\boldsymbol{x}, \xi^{(r)})$ is given in Lemma 3.
    Obtain optimal weights $\boldsymbol{\lambda}^{(r)}$ of the current design points set $\mathcal{X}^{(r)}$ using Algorithm 2 (**Optimal-Weight Procedure**) and form the current design $\xi^{(r)} = \begin{Bmatrix} \mathcal{X}^{(r)} \\ \boldsymbol{\lambda}^{(r)} \end{Bmatrix}$.
    Calculate the lower bound of LEA-efficiency of $\xi^{(r)}$,

$$\text{eff.low} = 1 + 2\frac{\min_{\boldsymbol{x}\in\mathcal{C}} \phi(\boldsymbol{x}, \xi^{(r)})}{\mathrm{EA}(\xi^{(r)}; \mathcal{M}')}.$$

    $r = r + 1$.
end while

Note that, at the beginning of Algorithm 1, the locally optimal design and the corresponding optimality criterion value $\Phi_p^{\mathrm{opt}_j}$ need to be calculated for each model specification $M_j \in \mathcal{M}'$. This is because they are involved in $\mathrm{EA}(\xi; \mathcal{M}')$ and all its derivatives. However, we only need to compute them once. Using the algorithm proposed by Li and Deng (2020), we can construct local Φ_p-optimal designs for GLMs efficiently with guaranteed convergence.

3.2. An optimal-weight procedure given design points

Based on Corollary 1, with a given set of design points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, a sufficient condition that $\boldsymbol{\lambda}^*$ minimizes $\mathrm{EA}(\xi^{\boldsymbol{\lambda}}; \mathcal{M}')$ is $\phi(\boldsymbol{x}_i, \xi^{\boldsymbol{\lambda}^*}) = 0$, for $i = 1, \ldots, n$ or, equivalently (based on Lemma 3),

$$\begin{cases} q\sum_{j=1}^{m}\tilde{\Phi}_0^j(\xi^{\boldsymbol{\lambda}^*}) \\ =\sum_{j=1}^{m}\tilde{\Phi}_0^j(\xi^{\boldsymbol{\lambda}^*})w_j(\boldsymbol{x}_i)\boldsymbol{g}_j^\top(\boldsymbol{x}_i)\mathsf{M}_j(\xi^{\boldsymbol{\lambda}^*})\boldsymbol{g}_j(\boldsymbol{x}_i)\end{cases}, \quad p=0; \\ \begin{cases} q^{1/p}\sum_{j=1}^{m}\tilde{\Phi}_p^j(\xi^{\boldsymbol{\lambda}^*})\Phi_p^j(\xi^{\boldsymbol{\lambda}^*}) \\ =\sum_{j=1}^{m}\tilde{\Phi}_p^j(\xi^{\boldsymbol{\lambda}^*})w_j(\boldsymbol{x}_i)\left(\mathrm{tr}\left(\mathsf{F}_j(\xi^{\boldsymbol{\lambda}^*})\right)^p\right)^{1/p-1}\boldsymbol{g}_j^\top(\boldsymbol{x}_i)\mathsf{M}_j(\xi^{\boldsymbol{\lambda}^*})\boldsymbol{g}_j(\boldsymbol{x}_i)\end{cases}, \quad p>0; \tag{3.2}$$

where $\tilde{\Phi}_p^j(\xi)=[\Phi_p^{\mathrm{opt}_j}]^{-1}\exp(\Phi_p^j(\xi)/\Phi_p^{\mathrm{opt}_j})$ and $\mathsf{M}_j(\xi)=\mathsf{I}_j(\xi)^{-1}\mathsf{B}_j^\top\mathsf{F}_j(\xi)^{p-1}\mathsf{B}_j\mathsf{I}_j(\xi)^{-1}$ with $\mathsf{B}_j=(\partial\boldsymbol{f}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^\top)|_{\boldsymbol{\beta}=\boldsymbol{\beta}_j}$ and $\mathsf{F}_j(\xi)=\mathsf{B}_j\mathsf{I}_j(\xi)^{-1}\mathsf{B}_j^\top$. For convenience, we denote the right side of (3.2) as $d_p(\boldsymbol{x}_i,\xi^{\boldsymbol{\lambda}^*})$. For *any* weight vector $\boldsymbol{\lambda}=[\lambda_1,\ldots,\lambda_n]^\top$, simple linear algebra yields

$$\begin{cases} q\sum_{j=1}^{m}\tilde{\Phi}_0^j(\xi^{\boldsymbol{\lambda}})=\sum_{i=1}^{n}\lambda_i d_0(\boldsymbol{x}_i,\xi^{\boldsymbol{\lambda}}), & p=0; \\ q^{1/p}\sum_{j=1}^{m}\tilde{\Phi}_p^j(\xi^{\boldsymbol{\lambda}})\Phi_p^j(\xi^{\boldsymbol{\lambda}})=\sum_{i=1}^{n}\lambda_i d_p(\boldsymbol{x}_i,\xi^{\boldsymbol{\lambda}}), & p>0.\end{cases} \tag{3.3}$$

Combining (3.2) and (3.3), the sufficient condition of the optimal weights is equivalent to

$$\sum_{s=1}^{n}\lambda_s^* d_p(\boldsymbol{x}_s,\xi^{\boldsymbol{\lambda}^*})=d_p(\boldsymbol{x}_i,\xi^{\boldsymbol{\lambda}^*}),\quad p\geq 0, \tag{3.4}$$

for all design points $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$. To obtain the optimal weight $\boldsymbol{\lambda}^*$ that minimizes $\mathrm{EA}(\xi^{\boldsymbol{\lambda}};\mathcal{M}')$, the current weights of the design points are adjusted according to the two sides of (3.4). For a design point $\boldsymbol{x}_i$, if $d_p(\boldsymbol{x}_i,\xi^{\boldsymbol{\lambda}})>\sum_{s=1}^{n}\lambda_s d_p(\boldsymbol{x}_s,\xi^{\boldsymbol{\lambda}})$, then the weight of point $\boldsymbol{x}_i$ is increased based on (3.4). However, if $d_p(\boldsymbol{x}_i,\xi^{\boldsymbol{\lambda}})<\sum_{s=1}^{n}\lambda_s d_p(\boldsymbol{x}_s,\xi^{\boldsymbol{\lambda}})$, the weight of point $\boldsymbol{x}_i$ is decreased based on (3.4). Thus, following the similar idea in classic multiplicative algorithms (Silvey, Titterington and Torsney (1978); Yu (2010)), the ratio $\left(d_p(\boldsymbol{x}_i,\xi^{\boldsymbol{\lambda}})/\sum_{s=1}^{n}\lambda_s d_p(\boldsymbol{x}_s,\xi^{\boldsymbol{\lambda}})\right)^\delta$ is a good adjustment for the weight of design point $\boldsymbol{x}_i$. Because this weight updating scheme is inspired by the classic multiplicative algorithm, we call it the modified multiplicative procedure, and describe it in Algorithm 2 in Supplementary Material S2.

Note that Yu (2010) proved the convergence of the classical multiplicative algorithm (Silvey, Titterington and Torsney (1978)) used to construct a locally optimal design for a class of optimality $\mathrm{tr}(\mathsf{I}(\xi^{\boldsymbol{\lambda}};M)^p), p<0$, and Li and Deng (2020) extended the results to a more general class of Φ_p-optimality. However, the proof in Yu (2010) cannot be extended easily to prove the convergence of Algorithm 2, because the derivative of $\mathrm{EA}(\xi^{\boldsymbol{\lambda}};\mathcal{M}')$ to λ_i cannot be reformulated into the general form in Equation (2) in Yu (2010), where only one model is involved. Nevertheless, Lemma 4 shows that the optimization problem solved by Algorithm 2 is a convex optimization,

$$\begin{aligned} &\min_{\boldsymbol{\lambda}} \mathrm{EA}(\xi^{\boldsymbol{\lambda}}; \mathcal{M}') = \textstyle\sum_{j=1}^{m} \exp\left(\frac{\Phi_p^j(\xi^{\boldsymbol{\lambda}})}{\Phi_p^{\mathrm{opt}_j}}\right), \\ &\text{s.t. } \mathbf{1}^\top \boldsymbol{\lambda} = 1,\ \boldsymbol{\lambda} \geq \mathbf{0} \end{aligned} \tag{3.5}$$

with linear constraints. Some existing optimization tools are available to solve such an optimization. Based on our empirical study, Algorithm 2 converges to a solution as good as those of the commonly used optimization tools, but with a much faster computational speed. Example 1 is relegated to the Supplementary Material S2 owing to space limitations.

4. Numerical Examples

In this section, we conduct several numerical examples to evaluate the performance of the proposed Mm-Φ_p design under different types of model uncertainty. Woods et al. (2006) proposed a compromise design that optimizes the weighted average of certain criteria, where each criterion is based on a potential model from some prior distribution $p(M)$. There are two ways to define a compromise design. The first way aims at maximizing a weighted average of the local Φ_p-efficiencies. That is, $\xi_{\mathcal{M}'}^{\text{eff-com}} = \mathrm{argmax}_\xi \sum_{j=1}^{m} p(M_j)\, \mathrm{eff}_{\Phi_p}(\xi, \xi_{M_j}^{\mathrm{opt}}; M_j)$, which is henceforth called the eff-compromise design. Another type of compromise design minimizes the weighted average of a local Φ_p-criterion. That is, $\xi_{\mathcal{M}'}^{\Phi_p\text{-com}} = \mathrm{argmin}_\xi \sum_{j=1}^{m} p(M_j)\Phi_p(\xi; M_j)$, which is henceforth called the Φ_p-compromise design. The Φ_p-compromise design coincides with the Bayesian optimal design when considering only the uncertainty from unknown regression coefficients. The performance of the proposed Mm-Φ_p design is compared to that of the eff-compromise and Φ_p-compromise designs. A detailed discussion about the connection between the proposed Mm-Φ_p design and the compromise designs is relegated to the Supplementary Material S3.

For all the designs in the examples, the candidate pool $\mathcal{C}$ is constructed using grid points, and each dimension of $\boldsymbol{x}$ has 51 equally spaced grid points. We use the default uniform prior distribution on the model specification for the compromise designs. All algorithms are programmed using MATLAB and run on a MacBook Pro with a 2.4 GHz Intel Core i5 processor. For $\boldsymbol{f}(\boldsymbol{\beta}) = [f_1(\boldsymbol{\beta}), \ldots, f_q(\boldsymbol{\beta})]^\top$ in $\Phi_p(\xi, M)$ in (2.3), we set $f_j(\boldsymbol{\beta}) = \beta_j$.

4.1. Model uncertainty

In Example 2, we investigate the performance of the Mm-Φ_p design and algorithm when there is uncertainty in both the link functions and the basis functions in the model space $\mathcal{M}$.

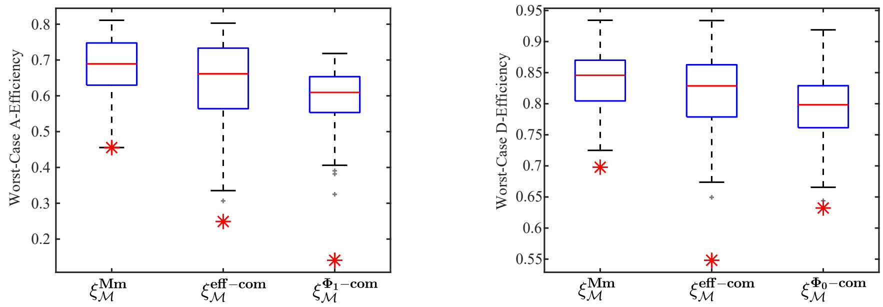
Table 1. Minimum and Median of the Worst-Case A- and D-Efficiency across 100 Randomly Generated Model Spaces for Comparison of Designs.

	Worst-Case A-Efficiency		Worst-Case D-Efficiency	
	min	median	min	median
Mm-Φ_p Design	0.46	0.69	0.70	0.85
Eff-Compromise Design	0.25	0.66	0.55	0.83
Φ_p-Compromise Design	0.14	0.61	0.63	0.80

Example 2. For an experiment with $d = 2$ input variables and one binary response, consider three link functions: logit link (h_1), probit link (h_2), and complementary log-log link (h_3). Consider possible polynomial basis functions up to degree two, that is, $\mathcal{G} = \{\boldsymbol{g}_1 = (1, x_1, x_2)^\top, \boldsymbol{g}_2 = (1, x_1, x_2, x_1x_2)^\top, \boldsymbol{g}_3 = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)^\top\}$. For the basis $\boldsymbol{g}_3$, the regression coefficients $\boldsymbol{\beta}_3 = [\beta_{3,1}, \ldots, \beta_{3,6}]^\top$ are drawn randomly from a standard multivariate normal distribution. For the basis $\boldsymbol{g}_2$, the regression coefficients $\boldsymbol{\beta}_2 = [\beta_{2,1}, \ldots, \beta_{2,4}]^\top$ are drawn independently with $\beta_{2,j} \sim \mathrm{N}(\beta_{3,j}, (0.5\beta_{3,j})^2)$, for $j = 1, 2, 3, 4$. The variance $(0.5\beta_{3,j})^2$ that depends on the regression coefficient $\beta_{3,j}$ allows a larger perturbation for $\beta_{2,j}$ when the corresponding $\beta_{3,j}$ is large. This accommodates the situation in which the values of the regression coefficients could change when the quadratic terms are not included in the model. For the basis $\boldsymbol{g}_1$, the regression coefficients $\boldsymbol{\beta}_1 = [\beta_{1,1}, \beta_{1,2}, \beta_{1,3}]^\top$ are drawn independently with $\beta_{1,i} \sim \mathrm{N}(\beta_{3,i}, (0.5\beta_{3,i})^2)$, for $i = 1, 2, 3$. Thus, the model space $\mathcal{M}$ consists of nine models: $\mathcal{M} = \left\{M = (h_i, \boldsymbol{g}_j, \boldsymbol{\beta}_j)_{i=1,j=1}^{3,3}\right\}$. We generate 100 parameter sets $\mathcal{B} = \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3\}$ to form 100 model spaces. For each generated model space, we construct the Mm-Φ_p design, eff-compromise design, and Φ_p-compromise design.

To compare the designs, we use the Φ_p-efficiency defined in (2.4) as a larger-the-better performance measure. In particular, we consider $\Phi_0(\xi; M)$ (i.e., $\lim_{p\to 0} \Phi_p(\xi; M)$) and $\Phi_1(\xi; M)$, which are the D- and A-optimality, respectively. For each model space, we compute the Φ_p-efficiency in (2.4) of all three designs relative to the corresponding locally optimal design. The locally optimal design ξ_M^{opt} is obtained using the algorithm of Li and Deng (2020). For each model space, we calculate the worst-case efficiency as $\min_{M_i \in \mathcal{M}} \mathrm{eff}_{\Phi_p}(\xi, \xi_{M_i}^{\mathrm{opt}}; M_i)$.

Figure 1 shows the box plots of the worst-case A- and D-efficiency of the Mm-Φ_p design, eff-compromise design, and Φ_p-compromise design across 100 different model spaces. The red asterisks "$*$" in the box plots denote the minimum worst-case A- and D-efficiency, and the larger the minimum, the better the design. Table 1 summarizes the minimum and median of the worst-case A- and D-efficiency of the three designs. The results show that the Mm-Φ_p design gives



(a) A-optimality (b) D-optimality

Figure 1. Box Plots of Worst-Case A- and D-Efficiency of Mm-Φ_p Design, Eff-Compromise Design, and Φ_p-Compromise Design across 100 Randomly Generated Model Spaces.

the largest values on the minimum and median of the worst-case efficiency. In terms of the worst-case A- and D-efficiency, the Mm-Φ_p design outperforms the eff-compromise design for 98% and 94% of the 100 model spaces, respectively. Note too that the eff-compromise design often gives the highest mean efficiency for a given model space, which is expected because it is designated to achieve the maximum mean efficiency. However, the mean A- and D-efficiency of all three designs are comparable, on average, over the 100 model sets. The computational times of Algorithm 1 to construct the Mm-Φ_p design, eff-compromise design, and Φ_p-compromise design are about 8.57 seconds, 8.59 seconds, and 6.48 seconds, respectively, for A-optimality, and 6.03 seconds, 7.98 seconds, and 3.30 seconds, respectively, for D-optimality.

4.2. Uncertain regression coefficients

In Example 3, we further illustrate the advantages of the Mm-Φ_p design by considering uncertain regression coefficients with the specified link function h and basis functions $\boldsymbol{g}$. Note that when the regression coefficient space $\mathcal{B}$ is continuous, a discretization is needed. In Example 3, we investigate the performance of the proposed design and algorithm over the unsampled values of the regression coefficient $\boldsymbol{\beta}$.

Example 3. For a univariate logistic regression model with experimental domain $\Omega = [-1, 1]$ and a quadratic basis, that is, $\boldsymbol{g}(x) = [1, x, x^2]^\top$, consider a regression coefficient space $\mathcal{B} = \{\beta_1 \in [0, 6], \beta_2 \in [-6, 0], \beta_3 \in [5, 11]\}$. Because $\mathcal{B}$ is continuous, we choose a Sobol sample of size 26 and the centroid $\boldsymbol{\beta}_c = [3, -3, 8]^\top$ of $\mathcal{B}$, that is, $m = 27$, to form the surrogate coefficient set $\mathcal{B}'$. The Sobol sample is a low discrepancy sequence that converges to a uniform distribution on a bounded set, and is widely used in Monte Carlo methods (Sobol (1967)). The surrogate model set is $\mathcal{M}' = \{M = (h, \boldsymbol{g}, \boldsymbol{\beta}) : h, \boldsymbol{g}, \boldsymbol{\beta} \in \mathcal{B}'\}$, where h is the link function of the logistic regression. Four designs are considered: (1) an Mm-Φ_p design $\xi^{\mathrm{Mm}}_{\mathcal{M}'}$; (2) an eff-compromise design $\xi^{\text{eff-com}}_{\mathcal{M}'}$; (3) a locally optimal design ξ^{center} of the centroid of $\mathcal{B}$, that is, $\boldsymbol{\beta}_c = [3, -3, 8]^\top$, which can be viewed as either an Mm-Φ_p or a compromise design with $m = 1$; and (4) a Bayesian optimal design $\xi^{\text{Bayesian}}_{\mathcal{M}'}$ with a uniform prior, which is also the Φ_p-compromise design. The constructed designs under D- and A-optimality are shown in Figure 1 in the Supplementary Material S4.

To compare the four designs, we use the Φ_p-efficiency defined in (2.4) as a performance measure. Specifically, we generate a Sobol sample of size 10,000 from the original continuous region $\mathcal{B}$. For each element of the sample, we compute the Φ_p-efficiency in (2.4) of all four designs relative to the corresponding locally optimal design, and the locally optimal design ξ^{opt}_M is obtained in the same way as in Example 1. Figure 2 shows box plots of the A- and D-efficiency of $\xi^{\mathrm{Mm}}_{\mathcal{M}'}$, $\xi^{\text{eff-com}}_{\mathcal{M}'}$, ξ^{center}, and $\xi^{\text{Bayesian}}_{\mathcal{M}'}$ over 10,000 randomly sampled $\boldsymbol{\beta}$ values. The red asterisks "$*$" in the box plots denote the worst-case A- and D-efficiency, where a larger value is better. Table 2 summarizes the minimum and median A- and D-efficiency of the four designs.

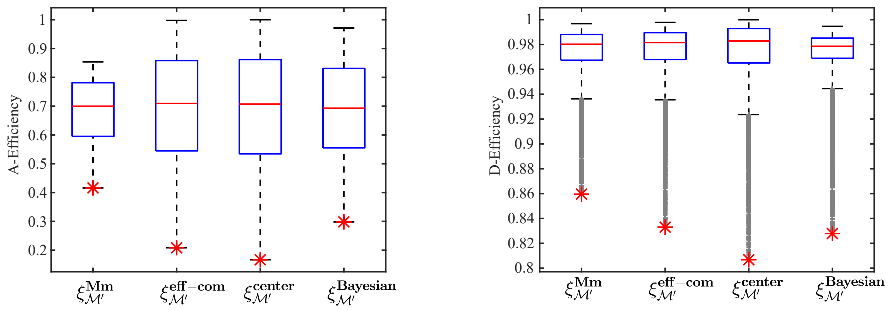
The results show that the Mm-Φ_p design $\xi^{\mathrm{Mm}}_{\mathcal{M}'}$ outperforms the other three designs in terms of the worst-case design efficiency, especially for A-optimality. Specifically, the worst-case A-efficiency of the Mm-Φ_p design is 0.41, and is much larger than those of the other three designs. The worst-case D-efficiency of the Mm-Φ_p design is 0.86, and is only slightly larger than those of other designs. As shown in Figure 2, the Mm-Φ_p design has the lowest maximum efficiency among the four designs over 10,000 values of $\boldsymbol{\beta}$. This is because the Mm-Φ_p design criterion, as a smooth approximation of the reciprocal of the minimum Φ_p-efficiency, aims at regulating the minimum efficiency rather than the maximum efficiency.

To illustrate the computational efficiency of the proposed Algorithm 1, Figure 2 in the Supplementary Material S4 shows how the Mm-Φ_p design criterion $\mathrm{LEA}(\xi^{(r)}; \mathcal{M}')$ decreases with respect to the number of iterations. The algorithm converges in 12 and 6 iterations to construct $\xi^{\mathrm{Mm}}_{\mathcal{M}'}$ for A- and D-optimality, respectively. The computation times are 1.68 seconds and 1.47 seconds for A- and D-optimality, respectively.

When the regression coefficient space $\mathcal{B}$ is continuous and uncountable, a

Table 2. Minimum and Median of A- and D- Efficiency across 10,000 Sampled $\boldsymbol{\beta}$ for Comparison of Four Designs.

	A-Efficiency		D-Efficiency	
	min	median	min	median
Mm-Φ_p Design	0.42	0.70	0.86	0.98
Eff-Compromise Design	0.21	0.71	0.83	0.98
Centroid Optimal Design	0.15	0.71	0.81	0.98
Bayesian Optimal Design	0.30	0.69	0.83	0.98



(a) A-optimality (b) D-optimality

Figure 2. Box Plots of A- and D-Efficiency of Four Designs at 10,000 Sampled $\boldsymbol{\beta}$.

discretization is needed to form the finite surrogate coefficient space $\mathcal{B}'$. Here, we investigate how the discretization size affects the performance of the constructed Mm-Φ_p design. Figure 3 shows a box plot of the A-efficiency of the Mm-Φ_p design constructed with $m = 8, 27, 64$, and 125 sampled $\boldsymbol{\beta}$ from $\mathcal{B}$. When the discretization size $m = 8$ is too small, not enough information about $\mathcal{B}$ is included in the surrogate coefficient space $\mathcal{B}'$. As a result, the constructed Mm-Φ_p design achieves a smaller minimum A-efficiency. For the other three discretization sizes, $m = 27, 64$, and 125, the performance of the constructed Mm-Φ_p design is relatively similar in terms of the minimum and median A-efficiency, because the surrogate coefficient space $\mathcal{B}'$ captures enough information about the continuous coefficient space $\mathcal{B}$.

4.3. Potato packing example

We consider a real-world example, the potato packing example in Woods et al. (2006), to further evaluate the proposed Mm-Φ_p design. The experiment contains $d = 3$ quantitative variables: vitamin concentration in the prepackaging dip, and the amounts of two kinds of gas in the packing atmosphere. The response is binary, representing the presence or absence of liquid in the pack after seven days.

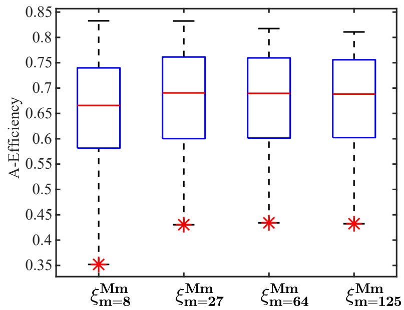


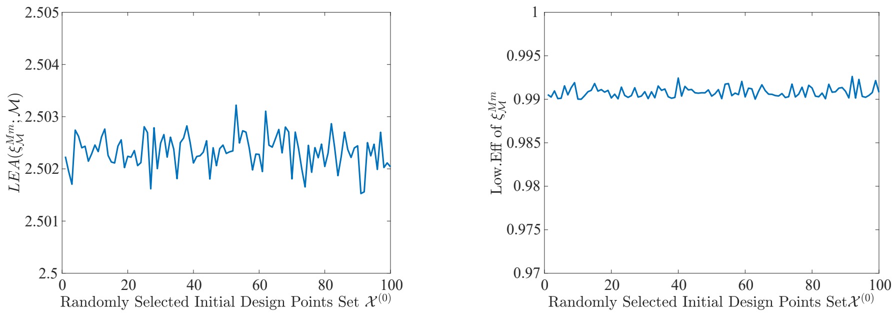
Figure 3. Box Plot of A-Efficiency of Mm-Φ_p Design with Various Sample Sizes m from Regression Coefficient Space $\mathcal{B}$.

Table 3. I-Efficiency of Mm-Φ_p Design, Eff-Compromise Design, and I-Compromise Design.

	M_1	M_2	M_3
Mm-Φ_p Design	0.64	0.71	0.82
Eff-Compromise Design	0.52	0.78	0.92
I-Compromise Design	0.49	0.80	0.92

The basis functions of the logistic regression model always include the linear and quadratic terms of the input variables. However, one set of the basis functions contains the interaction terms, and the other one does not. The estimates of the regression coefficients from the preliminary study in Woods et al. (2006) are given in Table 2 in the Supplementary Material S4. Because enhancing the prediction accuracy is a major goal of the experiment, we use the prediction-oriented I-optimality (Atkinson (2014)) to evaluate the design efficiency. Note that I-optimality shares the same mathematical structure as Φ_1-optimality. The design points of the designs are shown in Figure 3 in the Supplementary Material S4. Table 3 summarizes the I-efficiency of the Mm-Φ_p design, eff-compromise design, and I-compromise design for the three potential model specifications. In terms of worst-case efficiency (i.e., smallest value of I-efficiency among M_1, M_2, and M_3), the proposed Mm-Φ_p design outperforms the other two designs by a large margin.

We explore how the initial design points $\mathcal{X}^{(0)}$ in Algorithm 1 affect the performance of the algorithm. We generate 100 sets of $\mathcal{X}^{(0)}$, each of which consists of eight randomly chosen points from the candidate pool $\mathcal{C}$. The computational time, LEA-criterion value, and lower bound of the efficiency ("eff.low") of the constructed designs are presented in Figure 4. For 100 different initial design point sets $\mathcal{X}^{(0)}$, the computational time of Algorithm 1 ranges from 59 seconds



(a) LEA-Criterion (b) Lower Bound of the Efficiency

Figure 4. Performance of Algorithm 1 for 100 Randomly Generated $\mathcal{X}^{(0)}$.

to 93 seconds, with a median of 78 seconds. Although the convergence of Algorithm 1 is theoretically guaranteed regardless of $\mathcal{X}^{(0)}$, the computational time of the algorithm could vary to some extent with different $\mathcal{X}^{(0)}$. Figure 4 shows that the algorithm converges within $MaxIter1 = 200$ iterations for all 100 generated $\mathcal{X}^{(0)}$, and the LEA-criterion values fall between 2.501 and 2.503. This indicates that the quality of the constructed Mm-Φ_p design is quite robust against the initial design points set $\mathcal{X}^{(0)}$.

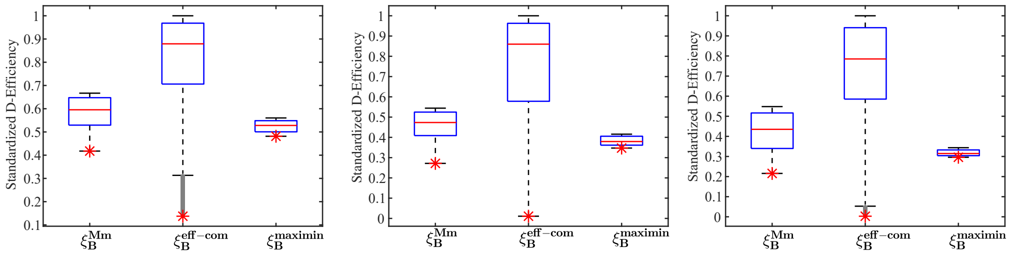
4.4. Comparison with maximin design

The proposed Mm-Φ_p design is based on a smooth approximation (2.7) of the minimum Φ_p-efficiency. Although the corresponding Mm-Φ_p design maximizes both the lower and the upper bounds of the minimum Φ_p-efficiency, there could be a gap between the Mm-Φ_p design and the true maximin design that maximizes the minimum Φ_p-efficiency. In this section, using the nonlinear model example in Braess and Dette (2007), we compare the performance of the Mm-Φ_p design (ξ^{Mm}) with that of the true maximin design (ξ^{maximin}).

Consider Example 3.2 in Braess and Dette (2007) of a univariate exponential growth model

$$y = e^{-\beta x} + \epsilon, \quad \beta \in [1, B], \quad x \in [0, 1],$$

where B is the upper bound of the regression coefficient β, and $\epsilon \sim N(0, \sigma^2)$ is the normally distributed homoscedastic error. The locally D-optimal design is then a single-point design at $x = 1/\beta$, and the corresponding D-optimality criterion value is $\Phi_0(\xi_\beta^{\mathrm{opt}}; \beta) = \ln[(e\beta)^2]$. For various values of B, Braess and Dette (2007) provided standardized maximin D-optimal designs, which are those that maximize the minimum D-efficiency without taking the natural log of the deter-



(a) $B = 10$ (b) $B = 40$ (c) $B = 100$

Figure 5. Box Plots of Standardized D-Efficiency of Mm-Φ_p Designs and Standardized Maximin D-Optimal Design at 10,000 Sampled β.

Table 4. Minimum and Median Standardized D-Efficiency across 10,000 Sampled β for Comparison of Mm-Φ_p Design, Eff-Compromise Design, and Standardized Maximin D-Optimal Design.

	$B = 10$		$B = 40$		$B = 100$	
	min	median	min	median	min	median
Mm-Φ_p Design	0.42	0.60	0.27	0.47	0.22	0.44
Eff-Compromise Design	0.14	0.88	0.01	0.86	0.003	0.79
Maximin Design	0.48	0.53	0.35	0.38	0.30	0.31

minants; that is,

$$\xi_B^{\text{maximin}} = \text{argmax}_{\xi} \inf_{\beta \in [1,B]} (e\beta)^2 |\mathbf{I}(\xi;\beta)|, \tag{4.1}$$

where $(e\beta)^2|\mathbf{I}(\xi;\beta)|$ is the standardized D-efficiency of design ξ relative to the locally standardized D-optimal design. The standardized maximin D-optimal designs ξ_B^{maximin} for different values of B in Braess and Dette (2007) are presented in Table 3 in the Supplementary Material S4.

To construct the Mm-Φ_p and eff-compromise designs, for each value of $B = 10, 40$, and 100, evenly spaced grid samples of size $m = B/2$ for $\beta \in [1, B]$ are used. The standardized D-efficiency of the Mm-Φ_p design ξ_B^{Mm}, eff-compromise design $\xi_B^{\text{eff-com}}$, and standardized maximin D-optimal design ξ_B^{maximin} are calculated over a 10,000 grid sample from $[1, B]$. Figure 5 shows box plots of the standardized D-efficiency of the Mm-Φ_p design, eff-compromise design $\xi_B^{\text{eff-com}}$, and standardized maximin D-optimal design at 10,000 sampled β, and Table 4 summarizes the minimum and median standardized D-efficiency of these designs.

The results show that, although the Mm-Φ_p design has a smaller worst-case D-efficiency than that of the true maximin D-optimal design for all scenarios, it achieves a much larger median and maximum D-efficiency. The eff-compromise design has the largest median and maximum D-efficiency, but the corresponding minimum D-efficiency may be very close to zero.

5. Discussion

There are several directions for further research to enhance the proposed Mm-Φ_p design and algorithm. First, to construct the Mm-Φ_p design, one needs to form a set of possible model specifications. An interesting direction is how to extend the proposed design when such information is limited or unavailable. Second, it would be interesting to rigorously establish the convergence property of the optimal-weight procedure (Algorithm 2), which requires developing some other mathematical results. Third, the numerical study shows that the lower bound of the LEA-efficiency, $1 + 2\min_{\boldsymbol{x}\in\Omega} \phi(\boldsymbol{x}, \xi^{(r)})/\mathrm{EA}(\xi^{(r)}; \mathcal{M}')$, may decrease for some iterations. Thus, it would be of interest to further investigate this lower bound theoretically. Lastly, the use of log-sum-exp approximation can be applied to other maximin designs with a convex design criterion, and the theoretical and algorithmic developments can be adapted similarly. We plan to extend the framework to a more general setting for other maximin designs with convex criteria.

Supplementary Material

The online Supplementary Material contains a proof of the convexity of $\mathrm{EA}(\xi^{\boldsymbol{\lambda}}; \mathcal{M}')$, a derivation of the directional derivatives $\phi(\xi', \xi)$ and $\phi(\boldsymbol{x}, \xi)$, proofs of Theorems 1,2, and 3, a detailed description of the modified multiplicative algorithm (Algorithm 2), a discussion on the connection and comparison between the proposed Mm-Φ_p design and the compromise design, and additional tables and figures for the numerical examples in Section 4.

Acknowledgments

We are grateful to the associate editor and two referees for their constructive comments. Deng's research was supported by the National Science Foundation CISE Expedition Grant CCF-1918770.

References

Atkinson, A. C. (2014). Optimal design. *Wiley StatsRef: Statistics Reference Online*, 1–17.

Atkinson, A. C., Donev, A. N. and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press. New York.

Atkinson, A. C. and Woods, D. C. (2015). Designs for generalized linear models. *Handbook of Design and Analysis of Experiments*, 471–514.

Braess, D. and Dette, H. (2007). On the number of support points of maximin and bayesian optimal designs. *The Annals of Statistics* **35**, 772–792.

Burghaus, I. and Dette, H. (2014). Optimal designs for nonlinear regression models with respect to non-informative priors. *Journal of Statistical Planning and Inference* **154**, 12–25.

Calafiore, G. C. and El Ghaoui, L. (2014). *Optimization Models*. Cambridge University Press, Cambridge.

Dean, A., Morris, M., Stufken, J. and Bingham, D. (2015). *Handbook of Design and Analysis of Experiments*. CRC Press, Boca Raton.

Dror, H. A. and Steinberg, D. M. (2006). Robust experimental design for multivariate generalized linear models. *Technometrics* **48**, 520–529.

Fedorov, V. V. and Leonov, S. L. (2013). *Optimal Design for Nonlinear Response Models*. CRC Press, New York.

Imhof, L. and Wong, W. K. (2000). A graphical method for finding maximin efficiency designs. *Biometrics* **56**, 113–117.

Khuri, A. I., Mukherjee, B., Sinha, B. K. and Ghosh, M. (2006). Design issues for generalized linear models: A review. *Statistical Science* **21**, 376–399.

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

King, J. and Wong, W. K. (1998). Optimal minimax designs for prediction in heteroscedastic models. *Journal of Statistical Planning and Inference* **69**, 371–383.

King, J. and Wong, W.-K. (2000). Minimax d-optimal designs for the logistic model. *Biometrics* **56**, 1263–1267.

Li, G. and Majumdar, D. (2009). Some results on D-optimal designs for nonlinear models with applications. *Biometrika* **96**, 487–493.

Li, Y. and Deng, X. (2020). An efficient algorithm for Elastic I-optimal design of generalized linear models. *Canadian Journal of Statistics* **49**, 438–470.

Santner, T. J., Williams, B. J., Notz, W. I. and Williams, B. J. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York.

Silvey, S. D., Titterington, D. H. and Torsney, B. (1978). An algorithm for optimal designs on a finite design space. *Communications in Statistics - Theory and Methods* **7**, 1379–1389.

Sitter, R. R. (1992). Robust designs for binary data. *Biometrics* **48**, 1145–1155.

Sobol, I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics* **7**, 86–112.

Wong, W.-K. (1992). A unified approach to the construction of minimax designs. *Biometrika* **79**, 611–619.

Wong, W. K., Yin, Y. and Zhou, J. (2019). Optimal designs for multi-response nonlinear regression models with several factors via semidefinite programming. *Journal of Computational and Graphical Statistics* **28**, 61–73.

Woods, D. C. and Lewis, S. M. (2011). Continuous optimal designs for generalized linear models under model uncertainty. *Journal of Statistical Theory and Practice* **5**, 137–145.

Woods, D. C., Lewis, S. M., Eccleston, J. A. and Russell, K. G. (2006). Designs for generalized linear models with several variables and model uncertainty. *Technometrics* **48**, 284–292.

Woods, D. C., Overstall, A. M., Adamou, M. and Waite, T. W. (2017). Bayesian design of experiments for generalized linear models and dimensional analysis with industrial and scientific application. *Quality Engineering* **29**, 91–103.

Wu, H.-P. and Stufken, J. (2014). Locally ϕ_p-optimal designs for generalized linear models with a single-variable quadratic polynomial predictor. *Biometrika* **101**, 365–375.

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., Zhang, B. and Huang, S. (2011). Optimal designs for generalized linear models with multiple design variables. *Statistica Sinica* **21**, 1415–1430.

Yu, Y. (2010). Monotonic convergence of a general algorithm for computing optimal designs. *The Annals of Statistics* **38**, 1593–1606.

Yiou Li

Department of Mathematical Sciences, DePaul University, Chicago, IL 60614, USA.

E-mail: yli139@depaul.edu

Lulu Kang

Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616, USA.

E-mail: lkang2@iit.edu

Xinwei Deng

Department of Statistics, Virginia Tech, Blacksburg, VA 24061-0131, USA.

E-mail: xdeng@vt.edu

(Received July 2020; accepted March 2021)